

Young Children's Ability to Produce Valid and Relevant Counter-Arguments

Bahar Köymen 
University of Manchester

Cathal O'Madagain
École Normale Supérieure

Andreas Domberg 
Max Planck Institute for Human Development

Michael Tomasello
Duke University and Max Planck Institute for Evolutionary
Anthropology

In collaborative problem solving, children produce and evaluate arguments for proposals. We investigated whether 3- and 5-year-olds ($N = 192$) can produce and evaluate arguments against those arguments (i.e., counter-arguments). In Study 1, each child within a peer dyad was privately given a reason to prefer one over another solution to a task. One child, however, was given further information that would refute the reasoning of their partner. Five-year-olds, but not 3-year-olds, identified and produced valid and relevant counter-arguments. In Study 2, 3-year-olds were given discourse training (discourse that contrasted valid and invalid counter-arguments) and then given the same problem-solving tasks. After training, 3-year-olds could also identify and produce valid and relevant counter-arguments. Thus, participating in discourse about reasons facilitates children's counter-argumentation.

Recent accounts of reasoning have emphasized its social dimension: individuals give and evaluate reasons for beliefs primarily in discourse (Mercier & Sperber, 2011). Reasoning may even have an ontogenetic origin in discourse—children come to differentiate good reasoning from bad reasoning by engaging in discourse in which interlocutors express different perspectives on a common topic (O'Madagain & Tomasello, 2019; Tomasello, 2019).

Studies with young children have looked mostly at what might be called passive reasoning. When evaluating others' reasons, 2-year-olds trust speakers who support their claims with information more than those who repeat themselves (Castelain, Bernard, & Mercier, 2018); 3-year-olds find claims based on eye-witness testimony more convincing than claims based on wishes (Koenig, 2012). Reasoning, however, requires more than trusting one reason more than another: it also involves weighing reasons against one another and rejecting bad reasons with valid objections.

One way to investigate reasoning is to explore children's ability to identify and produce counter-arguments, which has only been documented with adolescents (Kuhn & Udell, 2003). While a recent study showed that 7-year-olds can argue against taking some course of action, they did not necessarily do this as an objection to their peer's argument (Domberg, Köymen, & Tomasello, 2018). Use of counter-arguments has been little explored in preschoolers, despite findings showing how competent they are in justifying their proposals to convince their partners (Köymen, Mammen, & Tomasello, 2016).

In two studies, we investigated 3- and 5-year-olds' ability to identify and produce counter-arguments. In Study 1, peer dyads were presented with a problem to solve (e.g., "A girl must go to school in the rain. One of these two boxes contains the thing she needs. Which box is it?"). In the experimental trial, the target child was given extra information about the content of the nontarget child's (NT) box that would refute the NT's reasoning. For example, if the NT proposed her box "because it contains an umbrella," the target child would be given the information in advance that the umbrella was broken. She could then raise this valid objection "it is broken." In

We thank Katahrina Haberl, Susanne Mauritz, Gesa Volland, Hannah Streck, Milagros Galve, Sina Lehne, and Anja Dreves for their help in recruiting the children and collecting data; Anneli Arnold-Hofbauer and Kathrin Pellinger for their help in coding; and 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. Electronic mail may be sent to bahar.koymen@manchester.ac.uk.

© 2019 Society for Research in Child Development
All rights reserved. 0009-3920/2020/9103-0001
DOI: 10.1111/cdev.13338

the control trial, the extra information was an invalid or irrelevant objection to the peer's argument (e.g., "the boots have polka dots"). We expected that children would use the extra information as a counter-argument in the experimental trial, but not in the control trial. We expected this of 5-year-olds, but not 3-year-olds, because of younger preschoolers' difficulty in understanding others' beliefs and reasons for these beliefs before age 4–5 (Wellman, Cross, & Watson, 2001).

Since we found that 3-year-olds produced few counter-arguments, the question arose: how do children learn to produce counter-arguments between ages 3 and 5? Previous studies have shown that being exposed to discourse featuring conflicting attitudes to beliefs facilitates 3-year-olds' false belief understanding (Lohmann & Tomasello, 2003). Evaluating arguments as valid or invalid, might similarly improve children's counter-argumentation. Therefore, in Study 2, we provided 3-year-olds with two kinds of training that contrasted valid and invalid arguments. The *evaluation group* heard similar problems as in Study 1 and evaluated the validity of the counter-arguments. The *exposure group* heard the same counter-arguments but did not evaluate them. After training, 3-year-olds solved the same problems as in Study 1. We predicted that the 3-year-olds in both groups would perform better than the 3-year-olds in Study 1, with those in the evaluation group performing best.

Study 1

Method

Participants

Sixty-four 3-year-olds ($M = 3;9$, range = 3;6–3;11, 32 girls) and sixty-four 5-year-olds ($M = 5;9$, range = 5;7–5;11, 32 girls) in 64 same-age and same-sex dyads participated in the study. Each dyad participated in both the experimental trial and the control trial. An additional nine dyads (eight 3-year-old dyads, one 5-year-old dyad) were excluded from the analyses because they responded to memory questions incorrectly (see below). Children were native speakers of German and had varied socioeconomic backgrounds.

Materials

There were five puzzles with five toys (three dolls, a dog, a bear), 10 boxes in different colors, and pictures of box contents.

Procedure

After playing a game with two experimenters (E1, E2) for familiarization, one child within a dyad was randomly assigned to be the target child (T), the other to be the NT. For each puzzle, henceforth trial, E1 said, "Here are two locked boxes. [E2] will tell you what is in them." Children were asked to sit back-to-back. Each child learned the content of one box individually, whilst the other listened to music over headphones.

In warm-up Trial 1, E1 said, "This girl is going swimming. Only one box contains the thing she needs. You will decide which box this is." E2 told T, "The gray box is empty" and told NT, "The golden box contains her bathing suit [showing the picture of the bathing suit]." With headphones removed, E1 asked NT, "Which box? Why?" The NTs often said, "the golden one because it has a bathing suit." Then E1 asked T, "What is in the other box?" Ts often said, "it is empty." E1 summarized, "This box has the bathing suit; this is empty. Which is correct?" Children chose the box with a bathing suit.

Warm-up Trial 2 added a counter-argument against one box. E1 said, "This puppy is thirsty." E2 told NT, "The black box has its milk." and told T, "We have water and milk. But the milk has gone bad [counter-argument]. The red box has its water." E1 asked NT, "Which is correct? Why?" Children often said, "The black box because it has milk." Then E1 asked T "Which box? Why?" Children often said, "The red box, because it has water." Then E1 summarized, "Milk and water, both sound good. Which one is correct?" If T did not provide the counter-argument, E1 said, "I heard the milk has gone bad, so which box?" Children chose the box with water.

Then E1 told children that they would solve three more puzzles by themselves. The trials were in the following order: critical Trial 1, filler trial, critical Trial 2. If the first critical trial was the experimental trial, the second was the control trial and vice versa. The filler trial ensured that NT's proposal was correct in one trial. In critical Trial 1, the puzzle was "It's raining. This girl must go to school." T heard the same thing in both trials: "She has an umbrella, a pair of boots, and a raincoat. Her umbrella is broken [valid counter-argument]. Her boots have polka dots [invalid counter-argument]. The orange box contains her raincoat." The only difference between the experimental trial and the control trial was what NT heard:

1. In the experimental trial, NT heard, "The yellow box contains her umbrella" so the valid counter-argument was relevant. If NT proposed the yellow box, T could refute this by saying, "Her umbrella is broken" (omitting the invalid counter-argument).
2. In the control trial, NT heard, "The yellow box contains her boots" so the valid counter-argument was not relevant. If NT proposed the yellow box, T could not use the valid counter-argument ("Her umbrella is broken") because it was irrelevant (there was no umbrella) or the relevant counter-argument about the boots ("Her boots have polka-dots") because it was not a valid objection (see Appendix for a diagram).

In critical Trial 2, the puzzle was "The bear needs to eat." T heard, "It has honey, apples, and meat. The honey is gone bad [valid counter-argument]. The apples are sliced [invalid counter-argument]. The white box contains the meat." In the experimental trial, NT heard, "The brown box contains honey," making the valid counter-argument relevant. In the control trial, NT heard, "The brown box contains apples," making the invalid argument relevant. In the filler trial, the puzzle was "The boy needs a present for his friend." T heard, "The blue box is empty"; NT heard, "The green box contains a bike."

Ts, who received the extra information, always received it after their partner. The order of the arguments (valid, invalid), trials (experimental, control), stories (girl, bear) and arrangement of boxes (top, bottom) were counterbalanced.

After the three trials, E1 asked memory questions about the last trial. For the story with the girl, E1 asked NT, "Did your box have an umbrella or pajamas?" in the experimental trial; "Did your box have boots or pajamas?" in the control trial. E1 asked T:

1. "Did your box have a raincoat or a potato?"
2. "Did the boots have polka dots or stripes?"
3. "Was the umbrella broken or in good shape?"

For the bear story, E1 asked NT, "Did your box have honey or pajamas?" in the experimental trial; "Did your box have apples or a pajamas?" in the control trial. E1 asked T:

1. "Did your box have meat or a potato?"
2. "Were the apples sliced or round?"
3. "Was the honey gone bad or empty?"

If children answered any of these questions incorrectly, they were excluded from the analyses. We used recognition questions, as preschoolers often refuse to answer open-ended questions (Köymen & Tomasello, 2018).

Coding

For each trial, we coded whether T reproduced any of the arguments. Two 5-year-olds produced the relevant and valid argument in the experimental trial before NT proposed a box. They were treated as not having produced counter-arguments.

Another coder (blind to the hypotheses) re-coded 60 dyads. For each dyad, coders focused on five questions, concerning whether in the experimental trial:

1. NT produced the initial argument "This box contains the umbrella/honey";
2. T produced the valid counter-argument;
3. T produced the invalid counter-argument;

and whether in the control trial:

4. T produced the valid counter-argument;
5. T produced the invalid counter-argument.

The agreement was $\kappa = .89$.

Results

In the experimental trial, 72% (23/32) of the 5-year-old NTs and 50% (16/32) of the 3-year-old NTs produced the initial argument ("This box contains the umbrella/honey"), but there was no age difference in how often NTs produced this argument ($\chi^2(1, N = 64) = 2.36, p = .124$). The first two sets of analyses used the subset of dyads in which NTs produced the initial argument ($N = 39$ dyads) so Ts were given a chance to produce counter-arguments.

Analysis 1 tested whether Ts produced the valid counter-argument depending on its relevance. We ran a generalized linear mixed model (GLMM) with binomial error distribution because it allowed us to compare the likelihood of producing a counter-argument in each trial, while taking into account the noise created by factors such as stories etc. The full model included age (3, 5), trial (experimental, control), their interaction, trial order, gender, story (girl, bear), and the random factor of dyad. The null model included trial order, gender, story, and the random factor of dyad. The full model improved

the fit as compared to the null model ($\chi^2 = 27.49$, $df = 3$, $p < .001$). The interaction between age and trial was not significant ($\chi^2 = 0.80$, $df = 1$, $p = .371$). The reduced model without the interaction term revealed significant main effects of age ($\chi^2 = 9.82$, $df = 1$, $p = .002$) and trial ($\chi^2 = 18.31$, $df = 1$, $p < .001$). Five-year-olds produced the valid counter-arguments more often than did 3-year-olds. Both age groups produced the valid counter-arguments more in the experimental trial, when it was relevant, than in the control trial, when it was irrelevant (Figure 1).

Analysis 2 tested whether Ts produced relevant arguments, depending on their validity, using the same models as above. However, the models did not converge, due to very few arguments in the control trial. Chi-square analyses revealed that in the experimental trial, 5-year-olds produced the valid and relevant counter-arguments significantly more often than did 3-year-olds ($\chi^2(1, N = 39) = 8.83$, $p = .003$); whereas in the control trial, both age groups produced the relevant but invalid counter-arguments equally infrequently ($\chi^2(1, N = 39) = 0.22$, $p = .636$, Figure 1).

Analysis 3 tested whether raising the valid relevant counter-arguments predicted the correct decision in the experimental trial with all dyads ($N = 64$), using generalized linear model (GLM; instead of GLMM because we had one observation

from each dyad). The response variable was whether the correct box was chosen. The full model included age, mention of the counter-argument, their interaction, trial order, gender, and story. The null model included trial order, gender, and story. The full model improved the fit as compared to the null model ($\chi^2 = 24.13$, $df = 3$, $p < .001$). The interaction between age and mention of the counter-argument was not significant ($\chi^2 = 0.04$, $df = 1$, $p = .842$). The reduced model without this interaction revealed significant main effects of age ($\chi^2 = 9.21$, $df = 1$, $p = .002$) and mention of the counter-argument ($\chi^2 = 4.49$, $df = 1$, $p = .034$). Five-year-olds chose the correct box more often than did 3-year-olds (Figure 2). The dyads that produced the valid relevant counter-arguments were more likely to choose the correct box than those dyads that did not.

Discussion

Our results suggest that 5-year-olds, but not 3-year-olds, identified and produced counter-arguments to critique their peers' reasoning. Moreover, using valid objections facilitated reaching correct decisions.

Overall, 3-year-olds produced valid and relevant counter-arguments infrequently in the experimental trial and eventually could not identify the correct

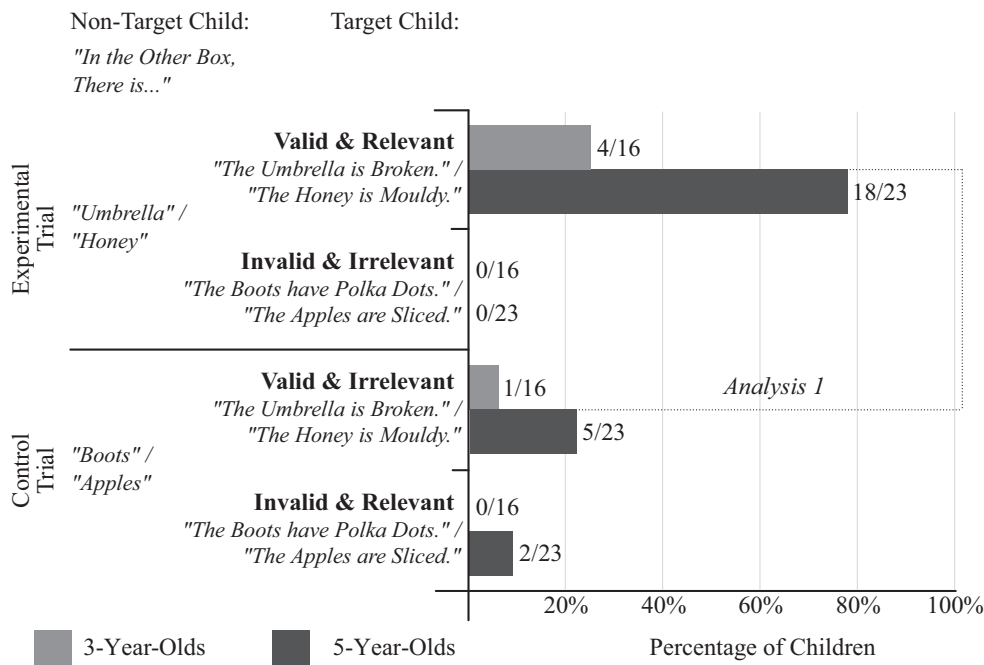


Figure 1. Percentage of children who produced counter-arguments out of 39 dyads (in which nontarget child produced the initial argument) in Study 1.

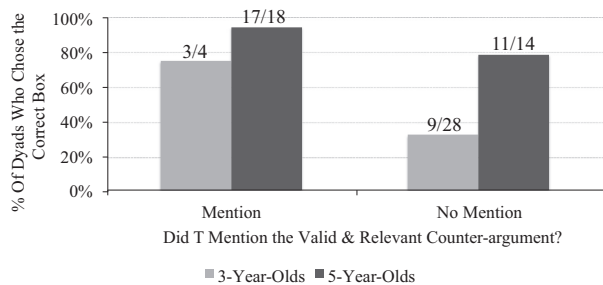


Figure 2. Percentage of dyads that chose the correct box in the experimental trial in Study 1. The numbers above the bars show the raw frequencies.

box, even though half of the 3-year-old NTs produced the initial argument (“This box contains the umbrella/honey”) giving Ts the opportunity to critique them. 3-year-olds’ infrequent use of counter-arguments might be due to the difficulty of evaluating someone’s reasons for a belief. This developmental pattern is in line with literature showing that children’s thinking about beliefs, including false beliefs, improves at around age 4 (Rakoczy, 2017; Wellman, et al., 2001). However, Lohmann and Tomasello (2003) showed that when 3-year-olds received training in which they engaged in discourse with two individuals who expressed conflicting attitudes to a belief (“perspective shifting,” e.g., “first you thought this was X; now you know this is Y”), their performance on false belief tasks improved. Similarly, perhaps being exposed to discourse, which contrasts valid and invalid arguments, and evaluating arguments might facilitate 3-year-olds’ use of counter-arguments.

Study 2

In study 2, we provided 3-year-olds with two kinds of discourse training on reasoning. The *evaluation group* heard a story in which a character needed one of two items (e.g., for drawing). A puppet produced a counter-argument against each option. One counter-argument was valid (e.g., “This pencil has no tip”); the other invalid (e.g., “This marker has a white cap”). Then the child and the experimenter evaluated the validity of each counter-argument (e.g., “Can she still draw with it?”). The *exposure group* was exposed to the same story and heard the same counter-arguments without evaluating their validity. These children then completed the two critical trials from Study 1 with an adult partner. We predicted that both groups would perform better than the 3-year-olds in Study 1 and produce the

counter-arguments in the experimental trial but not the control trial; with the evaluation group performing best.

Method

Participants

Sixty-four 3-year-olds ($M = 3;9$, range = 3;6–3;11, 32 girls) who did not participate in Study 1 participated in Study 2. Each child was randomly assigned to a training group (between-subjects factor) and participated in both trials (within-subjects factor). An additional 17 children were excluded from the analyses: 11 because of their incorrect responses to memory questions, four because of experimenter errors (e.g., forgetting memory questions), one because she refused to complete the trials, and one because he spontaneously evaluated the counter-arguments in the exposure group. Children were native speakers of German and had various socioeconomic backgrounds.

Materials

The materials were the same as in Study 1, except we replaced the two warm-up puzzles and the filler puzzle with three new puzzles for training. The training puzzles were different from the puzzles in the critical trials so the training was on argument validity; not on improving children’s knowledge about, for example, rain.

Procedure

With the evaluation group, Experimenter 1 (E1) explained that the child would solve puzzles with Experimenter 2 (E2). E1 introduced Puzzle 1, “Lisa wants to draw. One box contains a pencil; the other a marker. You (the child and E2) need to find the right box. Perhaps Maja (a puppet enacted by E1) can help you.” Maja gave a counter-argument for each option, “The marker is not good because the marker has a white cap [invalid]; the pencil is not good because the pencil has no tip” [valid]. In each training trial, the puppet used a different connector: *weil* ‘because’, *darum* ‘therefore’, and *nämlich* ‘namely’.

E2 evaluated the counter-arguments with the child, “The marker has a white cap. Can she still draw with it?”, until the child said “yes.” Then E2 said, “The pencil has no tip. Can she still draw with it?”, until the child said “no.” Next, E2 proposed the marker with a white cap and the child agreed.

The procedure was repeated for the other two puzzles. In Puzzle 2, a girl wanted to eat. One box had an ice cream, but "it had a strawberry flavor" [invalid counter-argument]; the other had a popsicle, but "it melted" [valid counter-argument]. In Puzzle 3, a girl wanted to play. One box had a rubber ball, but "it was colorful" [invalid counter-argument]; the other had a football, but "it was deflated" [valid counter-argument].

After training, children went through the same two critical trials from Study 1, except that E2 replaced the role of the NT. When solving the puzzles, E2 proposed a box "because it contains . . ." We coded whether the child produced a counter-argument against this. E1 asked the child the same memory questions. If children produced a counter-argument during a critical trial but answered the memory question about this incorrectly, they were not excluded from analyses.

With the exposure group, the procedure was the same except that instead of evaluating the counter-arguments, they talked about the task (e.g., "Do you like to draw?"). In each training trial, they made the correct decision.

The two teams who collected data, the order of the training puzzles, arguments (invalid, valid), trials (experimental, control), and stories were counterbalanced. The order of the connectors used in training was randomized.

Coding

Coding was the same as in Study 1 for T. A second coder, who was blind to the group and the hypotheses, re-coded 25% of the children. The agreement was $\kappa = 1.00$.

Results

We could not compare children's production of the counter-arguments in the experimental and control trials depending on the training they received, using GLMM, as we did in Study 1 because the models did not converge due to the low frequency of counter-arguments in the control trials. Instead, we first analyzed whether 3-year-olds' production of valid and relevant counter-arguments in the experimental trial varied depending on the training group. We included the sixteen 3-year-olds in Study 1 whose partners produced the initial argument to see whether any training improved 3-year-olds' production of counter-arguments using GLM with binomial error distribution. The response variable was whether children produced the valid relevant counter-argument. The

full model included the predictors training group (evaluation, exposure, 3-year-olds in Study 1), trial order, gender, and story. The null model only lacked the training group. The full model improved the fit, revealing a significant main effect of training group ($\chi^2 = 8.61$, $df = 2$, $p = .014$, Figure 3). Both training groups produced the valid relevant counter-arguments more often than the 3-year-olds in Study 1. However, this difference was significant for the evaluation group ($z = 2.74$, $p = .006$) but marginally significant for the exposure group ($z = 1.81$, $p = .071$). There was no significant difference between the two training groups ($z = 1.31$, $p = .189$).

Next, we analyzed whether there were group differences in 3-year-olds' production of invalid or irrelevant arguments in the control trial. The response variable was whether children produced any counter-arguments. The full and null models were the same as the previous analysis. The full model did not improve the fit suggesting that the three groups did not differ in the control trial and produced counter-arguments equally infrequently ($\chi^2 = 4.37$, $df = 2$, $p = .113$).

Finally, we analyzed whether children raising the valid and relevant counter-arguments predicted correct choice in the experimental trial. The response variable was whether children chose the correct box. The full model included training group, mention of the counter-argument, their interaction, trial order, gender, and story. The null model included trial order, gender, and story. The full model improved fit ($\chi^2 = 61.44$, $df = 5$, $p < .001$). The interaction was significant ($\chi^2 = 11.83$, $df = 2$, $p = .003$, Figure 4): Children who raised the valid relevant counter-argument were more likely to choose the correct box than those who did not; this difference was the greatest in the evaluation group.

Discussion

The results of Study 2 suggest that being exposed to discourse with counter-arguments facilitated 3-year-olds' ability to refute their partner's reasoning with valid and relevant counter-arguments. Although the two training groups did not differ significantly from each other, it was only the evaluation group that showed a significant improvement, when compared with the 3-year-olds in Study 1.

General Discussion

Our results suggest that by age 5, children spontaneously critique their peers' reasoning through

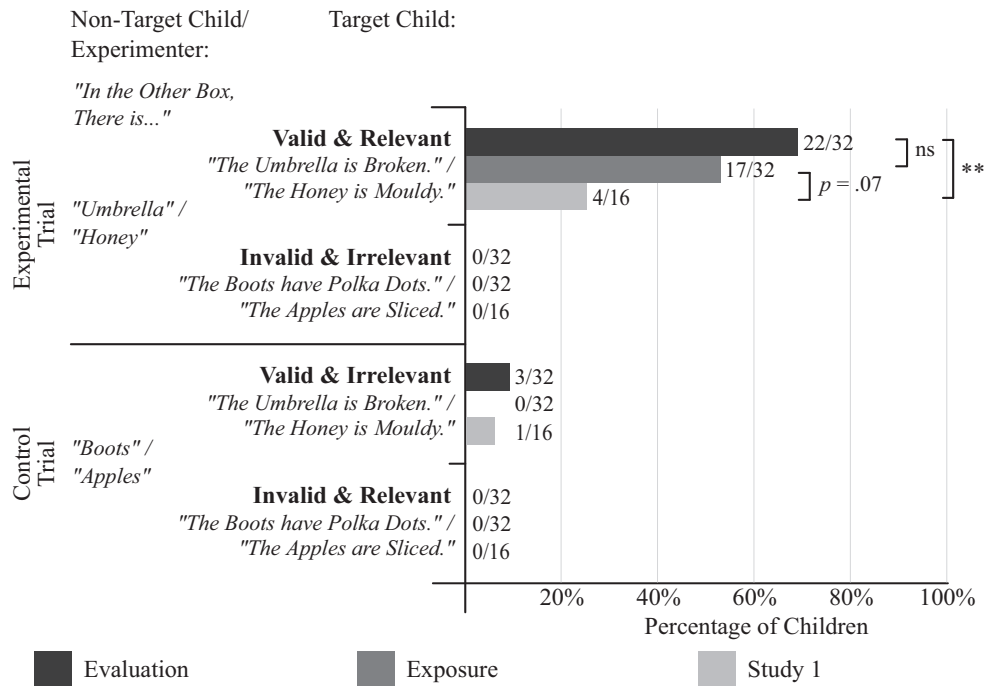


Figure 3. Percentage of 3-year-olds who produced counter-arguments in Study 1 and Study 2. "ns" corresponds to $p > .05$, "***" corresponds to $p < .01$.

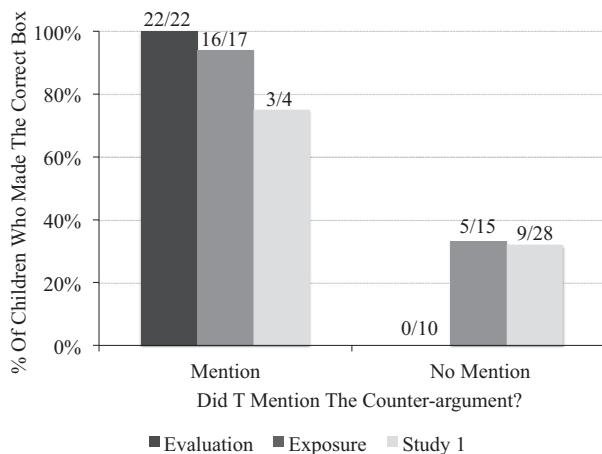


Figure 4. Percentage of children that chose the correct box in the experimental trial in Study 1 and Study 2. The numbers above the bars show the raw frequencies.

counter-arguments. By age 3, after training, they are able to produce counter-arguments. Literature on the development of reasoning has generally focused on what kinds of arguments children will find convincing (Castelain et al., 2018), and under what circumstances they produce arguments to convince others (Köymen et al., 2016). Our findings show that children, as young as age 3 can not only evaluate the validity arguments, but can also, with

training, produce them to critique their partner's reasoning.

The finding that discourse training increased 3-year-olds' ability to recognize information that would refute a partner's argument provides the first evidence to date that such discourse—especially if it involves *evaluation* of reasons—plays an important role in the development of young children's reasoning. Training studies such as ours establish that certain kinds of experiences are sufficient for developmental progress, suggesting the possibility that these experiences may play an important role in development outside the experimental context.

It can be argued that we made counter-arguments too easy for children, and they could simply repeat what they heard. However, children did not just repeat *any* additional information. If this were the case, they would have repeated the counter-arguments in the control trial as well. Our control trial also rules out explanations based on negativity bias, according to which children might produce the valid counter-arguments because they better remember negative things such as something broken (see Vaish, Grossmann, & Woodward, 2008). However, the children in our study did not reproduce the "negative" counter-arguments ("Her umbrella is broken") if they were irrelevant (when there was no umbrella).

Limitations and Directions for Future Research

One limitation for our study is that the better performance of the 3-year-olds in Study 2 could be due to having an adult partner, rather than a peer partner. We cannot rule out this possibility completely. However, giving children an adult partner in Study 2 was necessary because the partner had to provide the initial argument ("Let's pick this box, because it contains honey"). Perhaps adult partners might have made a difference by, for example, keeping the child more focused and facilitating the recall of counter-arguments. However, children have been shown to reason in more sophisticated ways when interacting with a peer than with an adult (Kruger & Tomasello, 1986; Mammen, Köymen, & Tomasello, 2019), making this interpretation less likely. Also if having an adult partner were having this effect, we should have had arguably fewer drop-outs due to incorrect responses to memory questions in Study 2. Yet, the drop-out rates were similar for 3-year-olds in both studies. Thus, we believe that simply having an adult partner, by itself, is unlikely to explain the better performance of 3-year-olds in Study 2. Moreover, children's familiarity with the task cannot explain the results, as children completed five trials in both studies. Future research could investigate whether discourse training on reasons would facilitate collaborative problem solving in peer interactions.

Conclusion

To conclude, by age 5, and with training by age 3, children can identify and use counter-arguments, which is a key element of reasoning and collaborative problem solving. Our results also support the view that participating in discourse that contrasts valid and invalid reasons may play a significant role in the development of children's reasoning.

References

- Castelain, T., Bernard, S., & Mercier, H. (2018). Evidence that two-year-old children are sensitive to information presented in arguments. *Infancy*, *23*, 124–135. <https://doi.org/10.1111/infa.12202>
- Domberg, A., Köymen, B., & Tomasello, M. (2018). Children's reasoning with peers in cooperative and competitive contexts. *British Journal of Developmental Psychology*, *36*, 64–77. <https://doi.org/10.1111/bjdp.12213>
- Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development*, *83*, 1051–1063. <https://doi.org/10.1111/j.1467-8624.2012.01742.x>
- Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, *52*, 423–429. <https://doi.org/10.1037/dev0000089>
- Köymen, B., & Tomasello, M. (2018). Children's meta-talk in their collaborative decision-making with peers. *Journal of Experimental Child Psychology*, *166*, 549–566. <https://doi.org/10.1016/j.jecp.2017.09.018>
- Kruger, A. C., & Tomasello, M. (1986). Transactive discussions with peers and adults. *Developmental Psychology*, *22*, 681–685. <https://doi.org/10.1037/0012-1649.22.5.681>
- Kuhn, D., & Udell, W. (2003). Development of argument skills. *Child Development*, *44*, 1245–1260. Retrieved from <http://www.jstor.org/stable/3696176>
- 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>
- Mammen, M., Köymen, B., & Tomasello, M. (2019). Children's reasoning with peers and parents about moral dilemmas. *Developmental Psychology*, *55*, 2324–2335. <https://doi.org/10.1037/dev0000807>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74. <https://doi.org/10.1017/S0140525X10000968>
- O'Madagain, C., & Tomasello, M. (2019). Joint attention to mental content and the social origin of reasoning. *Synthese*, *1–22*. <https://doi.org/10.1007/s11229-019-02327-1>
- Rakoczy, H. (2017). In defense of a developmental dogma: children acquire propositional attitude folk psychology around age 4. *Synthese*, *194*, 689–707. <https://doi.org/10.1007/s11229-015-0860-8>
- Tomasello, M. (2019). *Becoming human: A theory of ontology*. Cambridge, MA: Harvard University Press.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. <https://doi.org/10.1037/0033-2909.134.3.383>
- 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>

Appendix

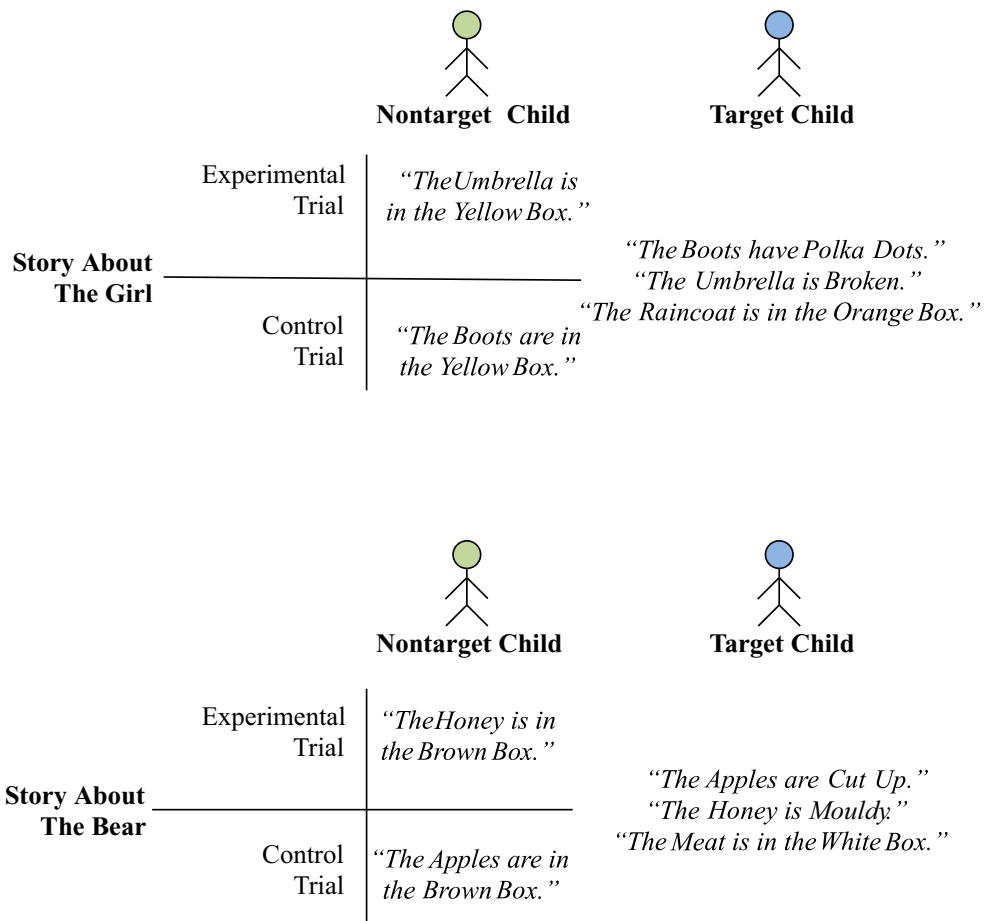


Figure A1. The diagram of the procedure of the critical trials. [Color figure can be viewed at wileyonlinelibrary.com]