Children's reasoning with peers and parents about moral dilemmas

OPEN ACCESS

Children's reasoning with peers and parents about moral dilemmas

Maria Mammen[1], Bahar Köymen[2], & Michael Tomasello[3-4]

[1]Ludwig-Maximilians-Universität München, Munich, Germany

[2]University of Manchester, Manchester, UK

[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[4]Duke University, Durham, NC

Article published in *Developmental Psychology*

Correspondence concerning this article should be addressed to Maria Mammen, Ludwig-Maximilians-Universität München, Department of Psychology, Leopoldstr. 13, D- 80802, Munich, Germany. E-mail: maria.mammen@lmu.de

Abstract

Children encounter moral norms in several different social contexts. Often it is in hierarchically structured interactions with parents or other adults, but sometimes it is in more symmetrically structured interactions with peers. Our question was whether children's discussions of moral norms differ in these two contexts. Consequently, we had 4- and 6-year-olds ($N = 72$) reason about moral dilemmas with their mothers or peers. Both age groups opposed their partner's views and explicitly justified their own views more often with peers than with mothers. Mothers adapted their discussions to the cognitive levels of their children (e.g., focused more on the abstract moral norms with 6-year-olds than with 4-year-olds), but almost always with a pedagogical intent. Our results suggest that with mothers moral judgments are experienced mostly as non-negotiable dictums, but with co-equal peers they are experienced more as personal beliefs that can be actively negotiated.

Keywords: Moral development, reasoning, mother–child interactions, peer interactions, moral dilemmas, justifications

Children's moral reasoning with peers and parents

To navigate their social world, children have to learn the norms that govern their social groups. Piaget (1932) proposed two fundamentally different, but equally important contexts in which moral development takes place: children's interactions with their adult caregivers and with their same-age peers. In adult–child interactions, due to the hierarchical structure of the relationship, the child often acts as the recipient of knowledge: the adult provides the child with knowledge of social and moral norms and the child often accepts this without questioning. In peer interactions, however, both participants are often equally knowledgeable and have equally valid perspectives; so which perspective is better has to be negotiated. Thus, in their interactions with adults, children usually tend to conform, whereas in peer interactions there is more joint decision-making guided by cooperative reasoning (Tomasello, 2019).

Research has shown that starting around 14-months, children usually recognise adults as more reliable information sources than children (Rakoczy, Hamann, Warneken, & Tomasello, 2010; Zmyj, Daum, Prinz, Nielsen, & Aschersleben, 2012). Children often learn about the moral norms of their social groups from their parents, who respond to moral transgressions by their children through explaining the reasons for rules and the consequences of their transgressions (Chapman & Zahn-Waxler, 1982; see also Smetana, 1999 for a review). Walker and Taylor (1991) found a correlation between the complexity of parents' explanations and their children's level of moral understanding; for example, the parents whose children had low scores in moral reasoning tasks, provided simpler explanations than the parents whose children had high scores in these tasks.

On the other hand, peer interactions, as argued by Piaget, offer children a "safe haven" to explore the moral norms of their social groups. To resolve their peer conflicts, children refer to norms of property entitlements or possessions and fairness (turn-taking), highlighting the "equal footing" between the participants (Ingram & Bering, 2010; Shantz, 1987). Recent studies have shown that preschoolers create and enforce social norms in their peer interactions (Göckeritz, Schmidt, & Tomasello, 2014; Köymen, Lieven, et al., 2014; Köymen, Schmidt, Rost, Lieven, & Tomasello, 2015). A recent study by Mammen, Köymen, and Tomasello (2018) also showed that preschoolers can reason with their peers about third-party moral transgressions and adjust the informativeness of their explanations for their moral judgments based on their mutual knowledge about social norms. For instance, when justifying their judgment of punishing a person who violated an unfamiliar social norm, they produced informative justifications, explaining the *general norm* (e.g., "because in Mia's kindergarten, you must not put yellow cars in the green box, but only in the yellow one", p. 256). However, when justifying their judgment for punishing a person who violated a familiar moral norm, they simply referred to the *fact* or the act itself (e.g., "because she stole"), seeing no need to state the general norm involved (e.g., "One must not steal") as moral norms are known to almost everyone. Moreover, some intervention studies have shown that children's individual scores on moral reasoning tasks significantly improved after having a discussion with peers (Damon & Killen, 1982; Kruger, 1992).

Despite the importance of these two contexts, there is no systematic comparison of moral reasoning of children with their peers and their caregivers, with one exception. Kruger and Tomasello (1986) found that in their discussions with mothers, both 7- and 11-year-old girls followed the mother's lead and mostly explained their views after being asked; whereas in their discussions with peers, children more often used transforming statements ("transacts") which

were defined as commenting on their partners' statements, challenging their partners' views and asking for feedback. However, their analyses did not necessarily focus on the content of children's justifications or how they actually talked about social norms (e.g., general norms vs. facts), how the moral dilemmas were treated and resolved, or whether mothers acted differently depending on the age of their children. Analyzing these would provide a better picture of children's experience in the two contexts of mother–child and peer interactions. Moreover, Kruger and Tomasello (1986) analyzed the discussions by school-aged children. Yet, studies show that already at preschool age, children are able to reason about third-party moral transgressions with adults (Mulvey, 2016) and peers (Mammen et al., 2018). Finally, Kruger and Tomasello (1986) predicted that the hierarchy between the children and the mothers would be stronger with younger children, but they did not find an age difference. This lack of age difference could be explained by the ceiling effect (the children were too old).

In the current study, therefore, we investigated how younger children, 4- and 6-year-olds, reasoned about moral dilemmas with their peers and their mothers. We selected these age groups because children are observed to justify their moral judgments about third-party moral transgressions in their spontaneous peer conversations around age 3-4 (Ingram & Bering, 2010; Mammen et al., 2018). We presented the [mother–child or peer] dyads with moral dilemmas, in which two characters could help a child at the expense of breaking a promise. One of the two characters kept the promise and thus could not help the child. The other character helped the child and thus could not keep the promise. The dyads were asked to decide which character did the right thing by distributing five gemstones between the two characters. We picked these dilemmas because findings suggest that preschoolers are aware of the moral norms of keeping a promise (e.g., Kanngiesser, Köymen, & Tomasello, 2017) and of helping or cooperating (e.g.,

McAuliffe, Jordan, & Warneken, 2015). We had two sets of research questions:

1. Do 4- and 6-year-old children reason differently with their peers and with their mothers?

We predicted that children would produce justifications spontaneously and justify their disagreements with the partner's view more often with their peers than with their mothers, as observed with older children. We also predicted 6-year-olds to be more active with their mothers than the 4-year-olds, because the hierarchy between the mothers and the children might get weaker as children get older (Kruger & Tomasello, 1986).

We further predicted that in their discussions of moral dilemmas, 4-year-olds would focus on the facts in the stories (e.g., "she helped"; "she kept her promise") whereas 6-year-olds might discuss general moral norms (e.g., by stating the norm "one should always help"). Moreover, older preschoolers might display higher levels of moral reasoning than younger preschoolers due to their more advanced socio-cognitive and linguistic skills (Wellman, Cross, & Watson, 2001). Children might focus more on abstract general norms with mothers than with their peers, due to the mothers reminding them of the social norms.

2. Do mothers reason differently with their 4- and 6-year-old children?

We predicted that mothers would adopt their reasoning to their children's knowledge of moral norms (Walker & Taylor, 1991), simplify the stories and focus on facts, such as who did the right thing (i.e., "She broke her promise but she helped") with the 4-year-olds, whereas they would discuss the moral norms generally and treat the stories as dilemmas with 6-year-olds (i.e., "Helping is important but keeping promises is important, too").

Methods

Participants

The participants were 36 4-year-olds ($M = 4;08$, *Range* = 4;06 - 4;11; 18 girls); 36 6-year-olds ($M = 6;08$, *Range* = 6;06 - 6;11; 18 girls) and 24 mothers ($M = 36;04$, *Range* = 22;07 - 48;02). We used a 2x2 between subjects design: age (4- vs. 6-year-olds) and condition (peer vs. mother–child). There were 12 same-sex peer dyads and 12 mother-child dyads per age group, resulting in a total of 24 peer dyads and 24 mother-child dyads. The peer dyads were formed based on the teachers' recommendation of frequent play partners. We only tested mothers to avoid confounds due to gender match of the parent and the child. The sample size was determined prior to data collection and was based on previous studies with similar designs (Köymen et al., 2014, 2015; Kruger & Tomasello, 1986). Moreover, to increase the statistical power, each dyad went through three trials, so we had three observations from each dyad. We did not run a power analysis prior to data collection. We decided against running a post-hoc power analysis based on our observed effects, as this would potentially overestimate our power (see Yuan & Maxwell, 2005). Instead, we calculated power by simulating data based on our experimental design (sample size and number of observations per participant) and assuming a medium effect size (*Cohen's d* = .50). For the analyses on children's justifications, repeated simulations with a logistic regression with random intercepts for participants and stories revealed a power of 0.71 and 0.73 for condition and age groups respectively. For the analyses on mothers' justifications, the simulations revealed a power of 0.87 for age groups. Although our sample is slightly under-powered, 48 dyads was the biggest feasible sample size due to practical constraints. The peer dyads were tested in ten nurseries; the mother–child dyads in six nurseries; all in different neighbourhoods of a mid-size city in Germany. The nurseries were selected

randomly. All participants were fluent German speakers and had different socio-economic backgrounds.

Materials

In the first warm-up, the dyads were presented with a toy cat and three envelopes containing two pictures each: a ball of wool and a bouquet of flowers as gifts for the cat; fish and a lollipop as food items for the cat; a dog and a mouse as friends for the cat.

In the second warm-up, the dyads were presented with three sets of hand-drawn pictures of two children and had to distribute five gemstones between the two characters.

In the three experimental trials, we used three hand-drawn 5-page picture books, each of which depicted a moral dilemma. In the filler trial, a fourth picture book was used which depicted a moral transgression (not a dilemma). There were two versions of each book; one with girls and one with boys, to match the gender of the children (see Appendix A for the pictures).

Procedure

The procedure of this project was approved by the Research Ethics Committee of X (project code: ReasAP, project title: "The two normative worlds"). All sessions were videotaped. For familiarization, all children first played a jigsaw with the experimenter (E); whilst the mothers filled out the consent forms in the mother-child condition.

There were two sets of warm-up trials. All warm-up sessions were conducted as interactive discussions between E and the dyads to show the dyads how the game should be played. In all warm-up trials, E facilitated the decision-making and made sure all dyads solved the task the same way. The first warm-up was introduced to encourage the dyads to provide justifications for their choices. E said that they will plan a birthday party for the toy cat Mimi. Then, E presented the three envelopes, saying: "For her birthday party, we choose a gift [from the envelope with

wool and flowers], something to eat [from the envelope with fish and a lollipop], and somebody to invite to the party [from the envelope with a dog and a mouse]. In each of these envelopes, there are two items. You two will decide together, which of these two items Mimi likes." E elicited reasons from them by asking why. E always stated the reasons: "I say, she gets the ball of wool, because cats like to play with it and cats can't really use flowers"; "I say, she gets fish, because cats like fish and lollypops are for humans"; "I say, we invite the dog, because dogs and cats can get along sometimes, but the mouse might be eaten by the cat." After hearing these reasons, all dyads agreed with E and made the same right decisions. E then praised the dyads for choosing the right items.

The second warm-up consisted of three trials that reminded dyads about the moral norms of helping and keeping promises and involved distributing five gemstones. In the first warm-up trial, E introduced the drawings of two children: one who helped cleaning up their toys, the other who did not help. E then asked, "Who did the right thing?"; "How many gemstones should they get? Why?" E always concluded, "I say, she does not deserve any gemstones, because she did not help, but he helped, so we can give him all five", to which all dyads agreed. In the second warm-up trial, E introduced the drawings of two children who promised to share their crayons with a third child. One of them kept the promise; the other did not. Again, E led the decision towards a 0/5 distribution favoring the promise keeper (e.g., "Because she kept his promise and he did not"), to which all dyads agreed. Unlike other warm-up trials in which one character did the right thing and the other did the wrong thing, in the third warm-up trial, the dyads were introduced to a dilemma, i.e. that a character can do something right and something wrong at the same time and how the distribution of the gemstones can reflect this dilemma. To avoid priming children in how to resolve moral dilemmas, we chose conventional transgressions. E presented

two drawings of children: "Else washed her hands and then she ate this sweet cake right before lunch"; "Moritz did not wash his hands and then ate this healthy sandwich for lunch". This time, E led the decision towards a 2/3 distribution, always ending with "Moritz did not wash his hands, so he cannot have all the gemstones. But he ate a healthy sandwich, so we can give him some gemstones. Else, she washed her hands, so she can also have some gemstones, but she ate sweets right before lunch, so we cannot give her all the gemstones." It was up to the dyads to decide which character should get three gemstones. In this last warm-up trial, children were only praised for settling on the 2/3 distribution and not about whom they favored.

E ensured that the dyads solved all warm-up tasks in the same correct way. In the first five warm-up trials, there was a clearly better option. Not correcting some dyads making an incorrect decision might have primed these dyads to think that whatever they decide would be fine later on. Eventually, these dyads might have been less likely to talk or justify their decisions.

In the three experimental trials, E presented the three stories with moral dilemmas (see-saw story, lego house story, and building blocks story; see Appendix A for the exact narration for each story). For example, in the see-saw story, on page 1, two characters were introduced. On page 2, the two characters made a promise (e.g., "They promise each other to stay on the see-saw until lunch."). On page 3, a third, a needy character was introduced, turning the situation into a dilemma for the two other characters (e.g., "This is Max, he likes the swings. But he is not good at it, so he asks the others to give him a push."). On page 4, one character [the promise-keeper] decides to keep his promise instead of helping the needy character, who is then sad. On page 5, the other character [the helper] decides to help the needy character instead of keeping his promise. As a result, the promise-keeper is sad. E always presented the promise-keeper first because if the character that broke the promise was presented first, it would not be logical for the

other character to keep a promise that is already broken. We chose these moral dilemmas for four

reasons. 1) We wanted to contrast two moral transgressions with comparable seriousness so that

dyads treat the stories as moral dilemmas. 2) We contrasted keeping promises and helping

because they are both positive moral duties ("one should help", as opposed to negative duties as,

"one should *not* steal") 3) We wanted to avoid participants justifying or excusing serious moral

transgressions (e.g., "It might be ok to steal if …"). 4) As stated earlier, the literature suggest that

preschoolers are aware of the moral norms of keeping promises (e.g., Kanngiesser et al., 2017)

and helping others (e.g., McAuliffe et al., 2015).

After the dilemmas, E first presented a drawing of the promise-keeper and then the

drawing of the helper and repeated what each character did (e.g., "He kept his promise, but he

did not help. He did not keep his promise but he helped."). Then, E gave five gemstones to the

dyads to distribute between the two characters. We asked the dyads to distribute an odd, rather

than an even number of gemstones so that they favor one character over the other. Since both

characters were portrayed as equally good and bad, it was difficult to decide who did the right

thing so the dyads would engage in longer discussions. E said, "Now you two can decide

together who did the right thing and explain to each other why. You can distribute the gemstones

as you like." To reduce the memory load, the picture books remained on the table whilst the

dyads made their decision. E then left the room and only entered when the dyads made their

decisions or got distracted. The dyads did not receive any feedback on their decisions at the end

of the trials. This procedure was repeated for two more experimental trials with similar dilemmas

about helping and keeping a promise (see Appendix A). After the first experimental trial, there

was a filler trial, which followed the same procedure as the experimental trials, except that the

story was not a dilemma but a simple transgression. A neutral character lent a toy to each of two

characters. One character returned the toy and the other did not. This filler trial was introduced to

prevent mothers from predicting that the next story might be a similar dilemma. The order of the

stories in the experimental trials and the presentation order of the actions were counterbalanced.

That is, all children heard about the promise keeper first in the story, but at the end of the story

when children were reminded of what each character did, half of the children heard "He/she

didn't help, but he/she kept his promise"; the other half heard "He/she kept his promise, but

he/she didn't help" to avoid any biases towards the first or the last clause.

Coding

The dyads' discussions during the experimental trials were transcribed. We coded for two

features of joint decisions: 1) how dyads justified their decisions/proposals; 2) how dyads

distributed the gemstones.

After we identified the justifications, we modified the coding scheme of Kruger and

Tomasello (1986), who used a general term "transact", and coded each justification for its

1) spontaneity, 2) disagreement, and 3) content. All three aspects were coded independently from

each other (i.e., all justifications were coded for spontaneity and disagreement regardless of their

content). For spontaneity, we identified whether a justification was *elicited* by a question from

the partner or produced *spontaneously*. For disagreement, we identified *challenges:* all

justifications that were produced in disagreement with the partner's previous proposals or

justifications about the distribution of the gemstones (e.g., after the partner proposed to give the

promise-keeper more gemstones, saying "But he [the promise-keeper] didn't help"). For the

content of the justifications, we used a coding scheme similar to that of Mammen et al. (2018), as

this study also investigated how children justified their moral judgments. This coding scheme

differentiated between three types:

General norms – Justifications referring to the norms (e.g., "Helping is important.") or

evaluating the action (e.g., "He helped, that is nicer.").

Facts – Justifications describing the action of the characters (e.g., "She did not help/keep

her promise.")

Irrelevant: justifications unrelated to the story (e.g., "Because I want to.").

For each trial, we coded how the dyads resolved the dilemma in terms of how they

distributed the gemstones. Further, we coded if the dyads came up with alternative endings to the

story. Although the task was to decide on who did the right thing, some dyads suggested that a

character could technically have done both (e.g., "She could come back and keep her promise

after helping her."), which went out of the script of the story. We termed this "alternative

ending". This category was not part of our initial coding. We observed this in mothers'

utterances, and added this to our coding scheme.

A second coder, who was blind to conditions and hypotheses, coded 25% of the transcripts,

three dyads from each age group and condition, for identifying the justifications, justification

spontaneity, challenges, justification content and trials with alternative endings. The agreements

were $\kappa = .80$; $\kappa = .80$; $\kappa = .82$; $\kappa = .87$ and $\kappa = .87$ respectively.

Results

The length of the discussions varied across the conditions: Peer discussions were shorter

than the discussions between children and their mothers. In the peer condition, 4-year-olds'

discussions across the three experimental trials contained 7 to 54 on-task utterances ($M = 24.2$;

$SD = 13.8$), 6-year-olds' discussions contained 3 to 31 on-task utterances ($M = 19.7$; $SD = 8.9$).

In the mother–child condition, discussions with 4-year-olds contained 31 to 153 on-task

utterances ($M = 84.3$; $SD = 37.4$), discussions with 6-year-olds contained 26 to 146 on-task

utterances ($M = 79.8$; $SD = 42.1$).

1. *Preliminary analyses: Gemstone distribution*

Mother–child dyads with 4-year-olds distributed the gemstones reflecting a dilemma (i.e.,

2/3 distribution) in 18 out of 36 trials (50%), mother–child dyads with 6-year-olds in 28 out of 36

trials (78%) so mothers treated the stories as dilemmas more often with 6-year-olds than with 4-

year-olds. That is, mothers with 6-year-olds rarely favored one character by choosing a 4/1 or 5/0

distribution. Peer dyads of both ages preferred each option equally. Four-year-old peer dyads

chose the 2/3 distribution in 22 trials (61%), 6-year-old peers in 19 trials (53%; see Figure 1).

Mother–child dyads with 6-year-olds favored the helper and gave her/him 3 or more

gemstones in 27 of 36 trials (75%). Mother–child dyads with 4-year-olds favored the helper in 20

of the 36 trials (56%). In the peer condition, each age group favored the helper in 19 of the 36

trials (53%). We ran a 2x2 between-subjects ANOVA to see whether dyads' preference for one

character over the other changed across the two conditions and the two age groups. The response

variable was the number of trials (out of 3) in which dyads favored the helper or awarded the

helper 3 or more gemstones. There was no significant difference between the two age groups

($F(1, 44) = 0.89$; $p = .35$, $\eta^2 = .02$) or between the two conditions ($F(1, 44) = 1.46$; $p = .23$, $\eta^2 =$

.03) on the number of trials in which dyads favored the helper. The interaction between age and

condition was not significant either ($F(1, 44) = .89$; $p = .35$, $\eta^2 = .02$; see Figure 2). The findings

that there was no significant age or condition difference in dyads' preference for the characters

and that the dyads favored each character around 50% of the time suggest that the dyads treated

the stories as dilemmas.

In the filler trial, in each age group 11 of the 12 mother–child dyads chose the 5/0 distribution in the filler trial and the remaining two dyads chose the 4/1 distribution, punishing the transgressor. The majority of the 4-year-old peer dyads also punished the transgressor in the filler trials: five dyads (42%) chose the 5/0 distribution; three dyads (25%) the 4/1 distribution and four dyads (33%) the 3/2 distribution. Similarly, of the 6-year-old peer dyads, nine dyads (75%) chose the 5/0 distribution and three dyads (25%) the 4/1 distribution.

2. *Do 4- and 6-year-old children reason differently with their peers and with their mothers?*

Descriptive data

Four-year-olds produced more justifications in the peer condition than in the mother–child condition (65 vs. 46). Six-year-olds showed the reverse pattern and produced fewer justifications in the peer condition than in the mother–child condition (43 vs. 80). Both age groups produced more spontaneous justifications (4-year-olds: 98% vs. 24%; 6-year-olds: 88% vs. 51%) and more challenges (4-year-olds: 48% vs. 7%; 6-year-olds: 33% vs. 4%) in the peer condition than in the mother–child condition (see Table 1).

Of the 65 justifications 4-year-olds produced in the peer condition, 5 (8%) referred to the general norm, 55 (85%) to facts and 5 (8%) were irrelevant. In the mother–child condition, of the 46 justifications, 6 (13%) referred to the general norm, 37 (80%) to facts and 3 (7%) were irrelevant. Of the 43 justifications 6-year-olds produced in the peer condition, 9 (21%) referred to the general norm and 34 (79%) to facts. In the mother–child condition, of the 80 justifications 5 (6%) referred to the general norm, 72 (90%) to facts and 3 (4%) were irrelevant (see also Table 2). The peers never referred to alternative endings.

Statistical analyses

To investigate the partner-effect, we ran three analyses using Generalized Linear Mixed Models (GLMM) with binomial error distribution on a subset of the data, which only included the children's justifications. We investigated whether there was a condition or age difference in children's spontaneous justifications (analysis 1), challenges (analysis 2) and justification content (analysis 3). We chose GLMM for four reasons. 1) In the peer condition, both children could produce justifications; whereas in the mother–child condition, we focused only on the child. GLMM allowed us to individually include each child regardless of the condition they were in. Random factors in GLMM allowed us to incorporate the information that the observations of the two children who interacted with one another in the peer condition were not independent. 2) Unlike ANOVAs that require more or less equal observations in each condition, GLMMs can handle datasets in which, for instance, there are twice as many observations in one condition (our peer condition) than the other (our mother–child condition). 3) GLMM allowed us to include the speaker as a random factor such that several justifications by each child could be accounted for. 4) All dyads were observed in three different experimental trials. GLMM allowed us to include story (see-saw, lego, blocks) and trial order (1, 2, 3) not only as fixed factors but also as random slopes to account for random variation due to stories and order across dyads.

In analysis 1, we investigated whether *children's use of spontaneous justifications* changed depending on their partner and age. The unit of analysis was each justification. The response variable was whether a justification was spontaneous or not. To test the significance of the full model, we compared its fit with a null model (see Appendix B for the predictors in the models and the results; see also the supplemental material for the datasets and the R-script used in the statistical analyses). The full model improved the fit ($\chi^2 = 57.30$; $df = 3$; $p < .001$). The full model revealed a significant interaction of age group and condition ($\chi^2 = 11.43$; $df = 1$; $p < .001$).

To better understand the interaction, we conducted pairwise comparisons. There was a significant condition difference for each age group (4-year-olds: $\chi^2 = 33.20$; $df = 1$; $p < .001$; 6-year-olds: $\chi^2 = 15.59$; $df = 1$; $p < .001$): both age groups produced significantly more spontaneous justifications in the peer condition than in the mother–child condition. There was also a significant age difference in the mother–child condition ($\chi^2 = 7.15$; $df = 1$; $p = .008$): with their mothers, 6-year-olds produced spontaneous justifications more than 4-year-olds (see Figure 3). There was no age difference in the peer condition (due to the lack of variance in the data we could not conduct a model comparison).

In analysis 2, we investigated whether *children's use of challenges* changed depending on their partner and age. The unit of analysis was each justification (spontaneous and elicited). The response variable was whether a justification was a challenge or not. The full model improved the fit as compared to the null model ($\chi^2 = 7.37$; $df = 3$; $p = .061$; see Appendix B for the predictors in the models and the results). The interaction of age group and condition was not significant ($\chi^2 = 0.03$; $df = 1$; $p = .875$) so we dropped the interaction term to get interpretable results for the main effects. The reduced model without the interaction term revealed a significant main effect of condition ($\chi^2 = 7.16$; $df = 1$; $p = .008$): both age groups produced challenges more in the peer condition than in the mother–child condition (see Figure 4).

In analysis 3, we investigated whether the *content of children's justifications* changed depending on their partner and age. The unit of analysis was each justification (spontaneous and elicited). The response variable was whether the justification was a general norm or a fact (we excluded the irrelevant justifications). The full model did not improve the fit ($\chi^2 = 5.09$; $df = 3$; $p = .165$; see Appendix B for the predictors in the models and the results). Thus, there was no age or condition difference in the content of the children's justifications.

3. *Do mothers reason differently with their 4- and 6-year-old children?*

Descriptive data

With 4-year-olds, all 12 mothers produced a total of 273 justifications. Of these, 225 (82%) referred to facts, 45 (17%) to general norms, 3 (1%) were irrelevant. With 6-year-olds, all 12 mothers produced a total of 177 justifications. Of these, 120 (68%) referred to facts and 55 (31%) to general norms, 2 (1%) were irrelevant (see also Table 2). Mothers referred to alternative endings with 4-year-olds in 9 out of 36 trials (25%), with 6-year-olds in 20 out of 36 trials (56%).

Statistical analyses

We ran two analyses, using again GLMMs with binomial error distribution on a subset of the data that contained only the mother's utterances. In analysis 4, we investigated whether the content of the *mothers' justifications* changed depending on the age of the child. The unit of analysis was each justification produced by mothers. The response variable was whether the justification referred to a general norm or a fact (we excluded the irrelevant justifications). When compared with a null model, which only lacked the age of the child, the full model improved the fit (χ^2 = 5.21; df = 1; p = .022; see Appendix C for a list of all predictors in the models and the results) and revealed a significant main effect of age group: mothers referred to the general norms more often with 6-year-olds than with 4-year-olds in their justifications (see Figure 5).

In analysis 5, we investigated whether *mothers' references to alternative endings* in their justifications change depending on the age of the child. The unit of analysis was the dyad. The response variable was a count of the trials in which mothers referred to an alternative ending vs. not. We used the "cbind" function in R with a binomial error structure. The cbind function allows to analyze the number of successes (mentioning alternative endings) and failures (not

mentioning alternative endings), as opposed to poisson models, where only the number of successes is counted (Baayen, 2008). When compared with a null model, which only lacked the age of the child, the full model improved the fit ($\chi^2 = 4.77$; $df = 1$; $p = .029$; see Appendix C for the predictors in the models and the results) and revealed a significant main effect of age group: mothers referred to alternative endings more often with 6-year-olds than with 4-year-olds (see Figure 6).

Summary of the results

In analysis 1, we found that children of both age groups produced more spontaneous justifications in the peer condition than in the mother–child condition, with the condition difference being greater for the 4-year-olds than for the 6-year-olds. In analysis 2, we found that both age groups produced significantly more challenges in the peer condition than in the mother–child condition. In analysis 3, we found no age or condition difference in the content of the justifications of children. Both age groups used predominantly facts in both conditions. In analysis 4, we found that mothers referred to the general norms more often with 6-year-olds than with 4-year-olds. In analysis 5, we found that mothers referred to an alternative ending more often with 6-year-olds than with 4-year-olds.

Discussion

Our results suggest that children of both age groups engage differently in collaborative reasoning in their two normative worlds. Overall, children produced more spontaneous justifications with their peers than with their mothers. Moreover, children of both age groups produced more justifications opposing their partners view (challenges) with their peers than with

their mothers. Thus, children were more active reasoners with their peers than with their mothers. Unlike in mother–child interactions in which mothers are always more knowledgeable than children, peer interactions, due to their equal footing, provided children a context in which peers could contrast and challenge one another's different proposals and arguments for resolving a moral dilemma. These results are in line with the findings of Kruger and Tomasello (1986), but the current study is the first to demonstrate this difference at a much younger age, namely in preschool years. It has been assumed that children learn about social norms through interacting with the other members of their culture (Piaget, 1932; Tomasello, 2016). Our study is the first to zoom into the interactions in which moral development takes place by systematically comparing how preschoolers reason about the moral norms with different social partners (mothers vs. peers) and how they respond to the moral judgments of these different partners.

Moreover, 6-year-olds were more active in their reasoning with their mothers than 4-year-olds, so the effect of the partner (a mother vs. a peer) was more pronounced with 4-year-olds. Younger children might have perceived a greater power asymmetry with their mothers. With the children's increasing reasoning skills, the asymmetry in mother–child interactions dissolves and children contribute to the interactions more actively. Mothers' perceptions of their children's competence in moral reasoning at different ages might have contributed to this as well (see Walker & Taylor, 1991). In fact, our findings suggest that mothers structured the discussions differently with 6-year-olds and 4-year-olds. During the discussions, mothers' justifications referred more often to the abstract general norms, like, "Keeping promises is very important", with the older than with the younger children. With the 4-year-olds, mothers' justifications referred mostly to facts from the stories, such as "He kept his promise". This difference in the mothers' reason-giving reflects their tendency to focus more on the here-and-now with younger

children. Moreover, through repeating the facts, mothers guided young children's attention and helped them to focus on what is relevant to solve the task.

Mothers were also more likely to propose an alternative ending, like "He [the helper] could come back to the see-saw after giving Max [the boy on the swings] a push" with 6-year-olds than with 4-year-olds. These alternative endings required counterfactual thinking as they go beyond the dilemma presented in the story and resolve it in a different way. With 6-year-olds, mothers used these rather complex arguments and went beyond the concrete events of the story by proposing alternative outcomes, which indicates again that they adjust the structure of their justifications to their children's level of understanding. In addition, because the alternative endings mostly suggested that one character could first help the needy child and then return to the friend to keep the promise, entertaining these alternative endings led mother–child dyads with 6-year-olds to favor the helper in majority of the trials (Figure 2). However, the dyads' overall preferences for the helper did not differ significantly across the age groups or conditions.

The content of children's justifications was similar in both conditions and in both age groups. They mostly referred to facts in their justifications and focused on the here-and-now (Köymen, Mammen, & Tomasello, 2016; Mammen et al., 2018). In the peer condition, children did not talk about alternative endings. Perhaps children here were too young to talk about counterfactuals on their own, unless they were with an adult (Harris, German, & Mills, 1996). Overall, it was when children were paired with their mothers that the age differences in children's reasoning become apparent. Thus, the power asymmetry influenced both partners' behavior: mothers drove the interaction differently with younger than with older children and offered more help based on what they believed was necessary for their child to understand the

task. At the same time, younger children became more passive, and let their mothers guide the decision-making process.

Our study had some limitations. First, in our warm-up trials, E ensured that all dyads solved the task the same "correct" way and praised them for solving the tasks correctly (except for the last warm-up trial). This might have primed the dyads to think that there is a correct way of resolving the moral dilemmas in the subsequent experimental trials, even though there was no correct answer. This might have discouraged children from providing justifications for their decisions because they would be worried to give the "wrong" answer. We opted for this approach because in the first five warm-up tasks, one option was clearly better than the other. Accepting any decision and giving neutral feedback would create a more ambiguous situation: Neutral feedback can be interpreted differently depending on the kind of decision it follows. If some dyads had knowingly made the wrong decision (e.g., choosing a lollipop to feed a cat, instead of fish), then neutral feedback might have been interpreted as "whatever we decide is fine". If some dyads had made the right decision, then the neutral feedback might have been interpreted as "we need to find the right choice". As a result, dyads would have been primed to do different things in the subsequent experimental trials. Instead, the warm-up trials were completed the same way across all four groups so we believe that the kind of warm-up activities that we used is unlikely to have a significant impact in our results.

A second concern could be the complexity of the stories, especially for the 4-year-olds. Yet, previous studies with similar stimuli have shown that 3-year-old children comprehend stories about moral transgressions (e.g., Mammen et al., 2018). In our study, the dyads could consult the picture books while they made their decisions, so the memory demands were reduced. Moreover, 4-year-olds produced justifications during peer conversations as often as 6-

year-olds did. Thus, we believe that the difficulty to recall the stories was unlikely to have an effect on our results. Finally, one could argue that the complexity of the stories varied such that some stories required better perspective-taking skills or Theory of Mind competence than others. However, we did not find any effect of story on any of our measures. Nonetheless, having an additional measure of theory of mind competence (e.g., false belief tasks) could have explained how theory of mind understanding affects moral reasoning.

In the present study, we observed children in two contexts: mother-child interactions with power asymmetry and peer interactions with no power asymmetry. The power dynamics within these two contexts, however, do change as children get older (Smetana, 1999). Mother-child interactions become more symmetrical (Grusec & Goodnow, 1994) and peer interactions might become more asymmetrical (Rubin, Bukowski, & Parker, 2008). Thus, although our study is a first and important step towards understanding the role of parents and peers in preschoolers' moral reasoning, future research should investigate how mothers and peers impact children's moral development in different ages, focusing more on the power dynamics in each context.

To conclude, as early as age 4, children reason differently in their two normative worlds. We have shown that the partner they are reasoning with significantly influences the way children navigate their social interactions. They are more active with their peers, recognizing the similarities in their knowledge, authority, and experience; whereas they are more passive with their mothers, recognizing the differences in knowledge, authority, and experience. Thus, peer interactions offer children a unique environment to practice their moral, cognitive, and reasoning skills.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Chapman, M., & Zahn-Waxler, C. (1982). Young children's compliance and noncompliance to parental discipline in a natural setting. *International Journal of Behavioral Development*, *5*(1), 81–94. https://doi.org/10.1177/016502548200500104

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly (1982-)*, *28*(3), 347–367.

Göckeritz, S., Schmidt, M. F. H., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development*, *30*, 81–95. https://doi.org/10.1016/j.cogdev.2014.01.003

Grusec, J. E., & Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology*, *30*(1), 4.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233–259. https://doi.org/10.1016/S0010-0277(96)00715-9

Ingram, G. P. D., & Bering, J. M. (2010). Children's tattling: The reporting of everyday norm violations in preschool settings. *Child Development*, *81*(3), 945–957. https://doi.org/10.1111/j.1467-8624.2010.01444.x

Kanngiesser, P., Köymen, B., & Tomasello, M. (2017). Young children mostly keep, and expect others to keep, their promises. *Journal of Experimental Child Psychology*, *159*, 140–158. https://doi.org/10.1016/j.jecp.2017.02.004

Köymen, B., Lieven, E., Engemann, D. A., Rakoczy, H., Warneken, F., & Tomasello, M. (2014). Children's norm enforcement in their interactions with peers. *Child Development*, *85*(3), 1108–1122. https://doi.org/10.1111/cdev.12178

Köymen, B., Mammen, M., & 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., Schmidt, M. F. H., Rost, L., Lieven, E., & Tomasello, M. (2015). Teaching versus enforcing game rules in preschoolers' peer interactions. *Journal of Experimental Child Psychology*, *135*, 93–101. https://doi.org/10.1016/j.jecp.2015.02.005

Kruger, A. C. (1992). The effect of peer and adult-child transactive discussions on moral reasoning. *Merrill-Palmer Quarterly (1982-)*, *38*(2), 191–211.

Kruger, A. C., & Tomasello, M. (1986). Transactive discussions with peers and adults. *Developmental Psychology*, *22*(5), 681–685. http://dx.doi.org/10.1037/0012-1649.22.5.681

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

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*, 1–10. https://doi.org/10.1016/j.cognition.2014.08.013

Mulvey, K. L. (2016). Evaluations of moral and conventional intergroup transgressions. *British Journal of Developmental Psychology*, *34*(4), 489–501.

Piaget, J. (1932). *The moral judgment of the child*. New York: Free Press.

Rakoczy, Hannes., Hamann, Katharina., Warneken, Felix., & Tomasello, Michael. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than

from peers. *British Journal of Developmental Psychology*, *28*(4), 785–798.

https://doi.org/10.1348/026151009X479178

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2008). Peer interactions, relationships, and

groups. In W. Damon & R. M. Lerner, *Child and adolescent development: An advanced

course* (pp. 141–180). Hoboken, N.J.: Wiley.

Shantz, C. U. (1987). Conflicts between children. *Child Development*, *58*(2), 283.

https://doi.org/10.2307/1130507

Smetana, J. G. (1999). The role of parents in moral development: A social domain analysis.

Journal of Moral Education, *28*(3), 311–321. https://doi.org/10.1080/030572499103106

Tomasello, M. (2016). *A natural history of human morality*. Cambridge, MA: Harvard

University Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge, MA: The Belknap

Press of Harvard University Press.

Walker, L. J., & Taylor, J. H. (1991). Family interactions and the development of moral

reasoning. *Child Development*, *62*(2), 264–283. https://doi.org/10.2307/1131002

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development:

the truth about false belief. *Child Development*, *72*(3), 655–684.

https://doi.org/10.1111/1467-8624.00304

Yuan, K.-H., & Maxwell, S. (2005). On the post hoc power in testing mean differences. *Journal

of Educational and Behavioral Statistics*, *30*(2), 141–167.

https://doi.org/10.3102/10769986030002141

Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-month-olds' imitation of differently aged models: Imitation of differently aged models. *Infant and Child Development*, *21*(3), 250–266. https://doi.org/10.1002/icd.750

Tables & Figures

Table 1. Children's justifications across age groups and conditions

age	condition	speakers	justifications	spontaneous justifications		challenges	
4	Mother–child	9	46	11	23.9%	3	6.5%
	Peer	18	65	64	98.5%	31	47.7%
6	Mother–child	12	80	41	51.3%	2	2.5%
	Peer	16	43	38	88.4%	14	32.6%

Table 2. Justification content across age groups and conditions

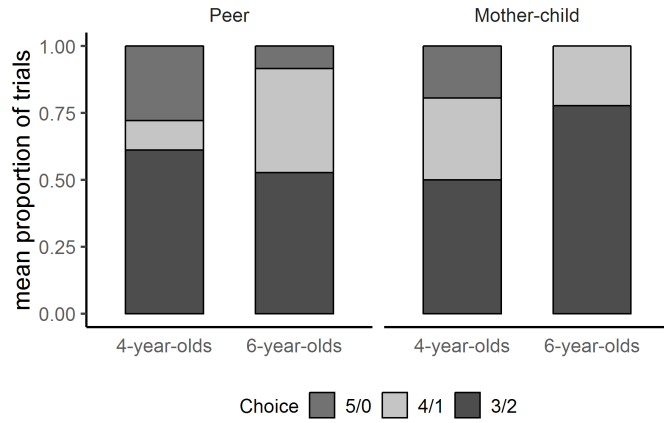| age	condition	role	speakers	justifications	general norm		facts		irrelevant justifications	
4	Mother–child	child	9	46	6	13.0%	37	80.4%	3	6.5%
		mother	12	273	45	16.5%	225	82.4%	3	1.1%
	Peer	-	18	65	5	7.7%	55	84.6%	5	7.7%
6	Mother–child	child	12	80	5	6.3%	72	90.0%	3	3.8%
		mother	12	177	55	31.1%	120	67.8%	2	1.1%
	Peer	-	16	43	9	20.9%	34	79.1%	0	0.0%

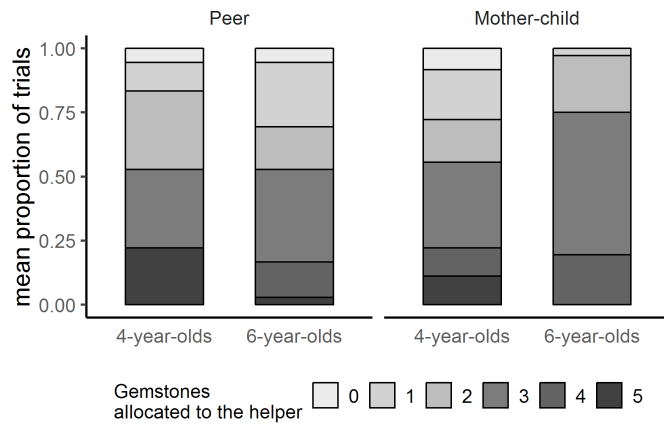*Figure 1:* Gemstone distribution in the three experimental trials

Figure 2: Proportions of trials with respective numbers of gemstones allocated to the helper by the dyads
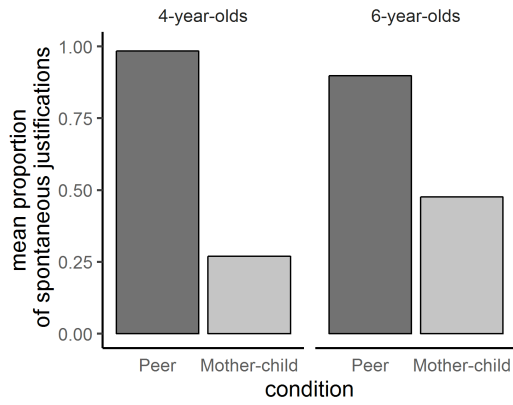
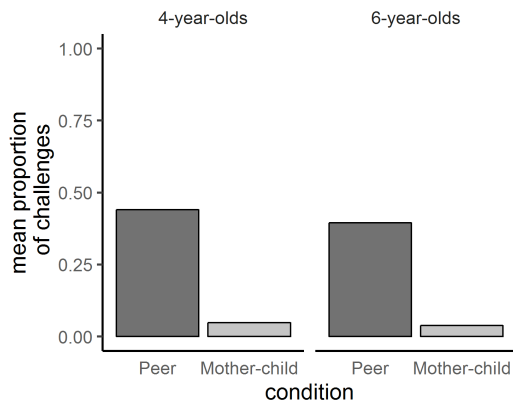*Figure 3:* Mean proportion of spontaneous justifications (children's talk)

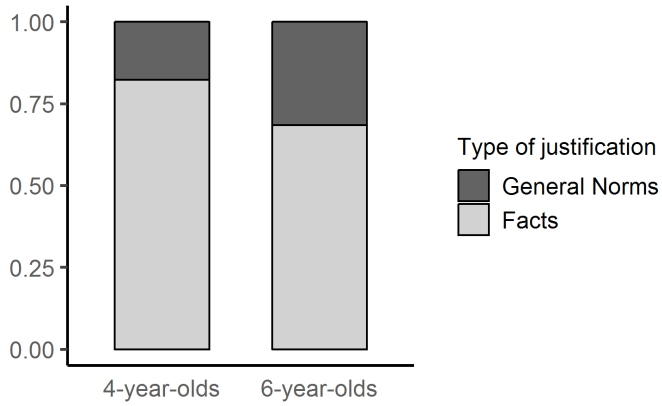*Figure 4:* Mean proportion of challenges (children's talk)

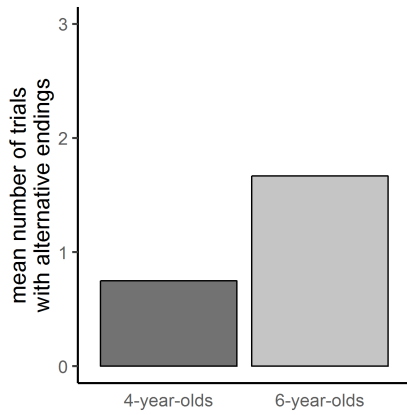*Figure 5:* Mean proportion of justification types (mother's talk)
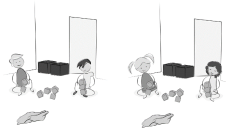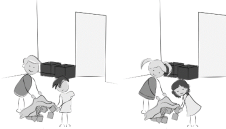

Figure 6: Mean number of trials with alternative endings (mother's talk)

Appendix A
Picture books and their narration in English

Experimental trials

1.1. This is [Child A] and this is [Child B]. They are 4/6 years old. Here they are playing with building blocks together. Those are their favorite blocks.

1.2. They want to continue playing with them later, so they hide the blocks. They promise each other not to tell anyone where the blocks are hidden. [Child A] says: "I won't tell anybody where the blocks are hidden, I promise." [Child B] says, "I won't tell anybody where the blocks are hidden, either, I promise."

1.3. This is [Child C]. She/He is building a block tower. [Child C] has played by herself/himself all day. Now she/he wants the tower to go higher. [Child C] is looking for someone who has a few blocks and wants to play with her/him.

1.4. [Child A] does not tell [Child C] where the blocks are. [Child A] does not tell the secret, like she promised. Now [Child C] is sad.

1.5. [Child B] tells [Child C] where the blocks are hidden. [Child B] breaks the promise and has now told the secret. Now [Child A] is sad.

2.1. These are [Child A] and [Child B] they are both 4/6 years old. They both like playing on the see-saw.

2.2. [Child A] and [Child B] promise each other to stay on the see-saw until lunch. [Child A] says, "I'll play on the see-saw with you until lunch, I promise." [Child B] says, "I'll play with you on the see-saw until lunch, too, I promise."

2.3. This is [Child C], she/he likes the swings. She/He is not good at it, so she/he asks the others to give her/him a push.

2.4. [Child A] cannot push her. She/He keeps the promise to play on the see-saw with [Child B]. Now [Child C] is sad.

2.5. [Child B] pushes [Child C]. Now she/he cannot play with [Child A] on the see-saw, even though she promised. Now [Child A] is sad.

3.1. These are [Child A] and [Child B], they are 4/6 years old. They are playing with lego blocks and build a house. Only the roof is missing.

3.2. It is snack-time now. They promise each other to finish the roof afterwards. [Child A] says, "I'll come back to finish the roof with you after the snacks, I promise." [Child B] says "I'll come back to finish the house with you after the snacks, too, I promise."

3.3. This is [Child C]. She/He found several boxes in the yard. She wants to bring them upstairs to play with them. The boxes are quite heavy.

3.4. [Child A] cannot help [Child C]. [Child A] has promised to finish the roof with [Child B]. Now [Child C] is sad.
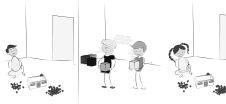
3.5. [Child B] helps [Child C] carrying the boxes. Now she/he cannot finish the roof with [Child A], even though she/he has promised to do so. Now [Child A] is sad.
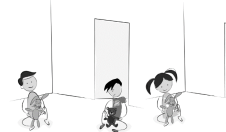
Filler Trial

1. This is [Child C] he has two stuffed toys.

2. This is [Child A], and this is [Child B]. They are both 4/6 years old [Child C] gives each of them one of her/his toys to play with for a little while.

3. They both really like [Child C]'s stuffed toys. [Child A] says, "I really like the stuffed toys." [Child B] says "I really like the stuffed toys, too."

4. Now they finished playing, [Child A] returns the stuffed toy to [Child C], because it belongs to [Child C].

5. [Child B] does not give back the stuffed toy. She/He wants to keep it, even though she/he knows that it belongs to [Child C]. Now [Child C] is sad.

Appendix B

Set 1 Analyses: Do 4- and 6-year-old children reason differently with their peers and with their mothers?

Table B1
Generalized Linear Mixed Models: Binomial Error Distribution

Analysis	Response variable	Unit of analysis	Predictors	χ^2	df	p	
1	Spontaneous justification vs. not	Justification	Full-null	57.300	3	< .001	***
			Age group X condition	11.430	1	< .001	***
			Gender	3.767	1	= .052	
			Trial order	5.691	2	= .058	
			Story	2.078	2	= .354	
2	Challenge vs. not	Justification	Full-null	7.371	3	= .061	
			Age group X condition	0.025	1	= .875	
			Age group[a]	0.193	1	= .660	
			Condition[a]	7.159	1	= .008	**
			Gender[a]	0.003	1	= .954	
			Trial order[a]	0.214	2	= .898	
			Story[a]	0.998	2	= .607	
3	General norms vs. facts	Justification	Full-null	5.095	3	= .165	

Notes. *** indicates $p < .001$; ** indicates $p < .01$.
Full model: response variable ~ age group*condition + gender + trial order + story + (1 + trial order.2 + trial order.3 + story.2 + story.3 ||dyad) + (1|speaker).
Null model: response variable ~ gender + trial order + story + (1 + trial order.2 + trial order.3 + story.2 + story.3 ||dyad) + (1|speaker)

Levels of the factors: age group: 4-year-olds, 6-year-olds; condition: peer condition, mother–child condition; gender: male, female; trial order: experimental trial 1, 2, 3; story: see saw, lego, blocks
Factors trial order & story were dummy coded and centered for the random slopes.
[a]The values come from the reduced model that lacks the interaction term age group X condition

Appendix C

Set 2 analyses: Do mothers reason differently with their 4- and 6-year-old children?

Table C1
Generalized Linear Mixed Models: Binomial Error Distribution

Analysis	Response variable	Unit of analysis	Predictors	χ^2	df	p	
4	General norms vs. facts	Justification	Full-null/age group	5.207	1	= .022	*
			Gender	0.067	1	= .795	
			Trial order	2.056	2	= .358	
			Story	1.723	2	= .423	
5	cbind (number of trials with alternative ending, number of trials without alternative ending)	Dyad	Full-null/age group	4.773	1	= .029	*
			Gender	0.024	1	= .878	

Notes. * indicates $p < .05$.
Analysis 4: Full model: response variable ~ age group + gender + trial order + story + (1 + trial order.2 + trial order.3 + story.2 + story.3 ||speaker)
Null model: response variable ~ gender + trial order + story + (1 + trial order.2 + trial order.3 + story.2 + story.3 ||speaker)

Analysis 5: Full model: age group + gender + (1|speaker.ID)
Null model: gender + (1|speaker.ID)

Levels of the factors: age group: 4-year-olds, 6-year-olds; gender: male, female; trial order: experimental trial 1, 2, 3; story: see saw, lego, blocks
Factors trial order & story were dummy coded and centered for the random slopes.