Young children give transgressors the benefit of the doubt in the absence of intention information

Marina Proft [a], Owen Waddington [b,*] , Bahar Köymen [b]

[a] *Georg-Elias-Müller-Institute for Psychology, Georg-August-University Göttingen, Göttingen, Germany*
[b] *Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, Manchester, UK*

ABSTRACT

Young children consider transgressors' intentions in their normative judgments. But how do children evaluate moral transgressions in the absence of information about a transgressor's intent? Across three studies, 5-year-old German-speaking children ($N = 216$, 108 girls, 108 boys) observed negative moral outcomes in which the transgressor was either smiling (happy condition), shocked (surprised condition) or was without expression, in that their face was left entirely blank (no-expression condition). Children then reasoned in pairs (Study 1 and 2) or independently (Study 3) about the intentional structure of each transgression. In Study 1, dyads concluded the transgressions were intentional in the happy condition, accidental in the surprised condition, and were at-chance in the no-expression condition. In Studies 2 and 3, methodological changes meant children concluded the transgressions were accidental in the no-expression condition and were at-chance in the happy condition. When intention information was thus unavailable, 5-year-olds preferred to ascribe positive intentions to transgressors devoid of all expression and give them the benefit of the doubt.

Giving the benefit of the doubt rests on the foundation of trust between individuals within a social group and promotes social harmony. Not only does the concept of *in dubio pro reo* (Latin: innocent until proven guilty) constitute the basis of legal systems around the world, it is also seen as a fundamental human right (Art. 6, Para. 2, European Convention of Human Rights). Investigating its psychological underpinnings together with its developmental origins is key to understanding how collective moral judgments are made.

Developmental studies show young children initially make moral judgments based on the severity of a harm (i.e., the outcome), but later focus increasingly on the transgressor's underlying intentions (also known as the "outcome-to-intent-shift", see Cushman et al., 2013; Killen et al., 2011; Li & Tomasello, 2018; Nobes et al., 2009; Vaish et al., 2010). By age 5, children reliably differentiate between well-intentioned and ill-intentioned transgressors (Proft & Rakoczy, 2019; Waddington et al., 2023), though sensitivity to others' sociomoral mental states may be implicitly present from early infancy (e.g., Hamlin et al., 2007; but see Lucca et al., 2025).

An influential source of information on children's intent-based normative judgments is a perpetrator's emotional response to their transgression. Research shows that preschool-aged children associate different types of sociomoral events (e.g., helping, harming) with distinct emotional outcomes (Arsenio, 1988). But when inferring others' emotions, young children focus mostly on behavioural outcomes rather than on intentions. Given a character who wanted to strike another child with a ball, 3-year-olds judged the character as feeling sad despite fulfilling her desired end (Yuill, 1984; Yuill et al., 1996). By age 5, however, children begin to more closely

---

associate others' emotional cues with their mental states (see Hadwin & Perner, 1991). Actors who deliberately misbehave, for example, are ascribed positive emotions at this age given that this was their intention (Nummer-Winkler & Sodian, 1988; Yuill et al., 1996). And when observing transgressors with different facial expressions, 5-year-olds associated surprised expressions with accidental harm and happy expressions with intentional harm (Waddington et al., 2023).

To prevent transgressors' emotions from biasing children's judgements, other studies have depicted or instructed transgressors to express neutral emotion to make their intentions appear perceptually ambiguous (Oostenbroek & Vaish, 2019; Vaish et al., 2010, 2011; Waddington et al., 2022). How "neutral" expressions are interpreted, however, is highly context dependent. Neutral faces, for example, tend to be evaluated as sadder or happier depending on the expression they proceed (Russell & Fehr, 1987; see also Albohn & Adams, 2021) and are more closely associated with negative emotions than positive emotions (Lee et al., 2008). Thus, "neutral" expressions may not be perceived by children as being entirely neutral, leading to a potentially confounding source of variability in their intention ascriptions. At present, it is therefore difficult to ascertain precisely whether children give transgressors the benefit of the doubt under truly ambiguous conditions, leaving the development of this important sociomoral behaviour unclear. A truer test of this capacity may be to remove the transgressor's facial expression altogether, such that there are no visual cues to sway children's intentionality judgements. This raises the question: when children observe a transgression and all intent-based information is withheld, do they rather assume the harm occurred accidentally or deliberately?

Though this question is understudied, theories predict opposite patterns. On the one hand, several theories predict a tendency to lend others the benefit of the doubt. According to the moral core concept (Strohminger et al., 2017), people assume that within each individual exists a true self that defines one's character, is fundamentally good, and which calls us to behave in morally virtuous ways (Newman et al., 2014; Lefebvre & Krettenauer, 2020; Umscheid et al., 2023). Teleology accounts (Roessler & Perner, 2013) argue that people aim to bring about a more attractive state of the world than it is presently in and thus act in ways that submit to and are best explained by normative forces (Perner et al., 2018). Malle and Bennett (2002) emphasize the social function of intention ascriptions: ascribing intentions serves as a character judgment, with a negative intention equaling a negative character. Thus, falsely assigning bad intentions to someone can be costly for both the assigner (e.g., loss of reputation) and the assigned (e.g., unjust punishment), and should therefore be avoided in cases of uncertainty.

On the other hand, other theories predict an aversion to giving the benefit of the doubt. The fundamental attribution error posits that individuals emphasise dispositional factors when evaluating others' behaviours (Heider, 1958; Ross, 1977). That is, rather than give the benefit of the doubt and assume a situational cause for a harm (e.g., being unaware of a norm), observers assume it reflects a more stable personality trait (e.g., this is a deviant person). Moreover, the hostile attribution bias argues that people are universally predisposed in early life to ascribe intentional motives to ambiguous events, and only through socialization do the majority learn that not all provocation is necessarily hostile (Dodge, 2006).

Ontogenetically, children's tendency to offer others the benefit of the doubt likely varies across different developmental stages. Young children, for example, show a robust negativity bias in their processing of information and in their social evaluations (Hamlin et al., 2010; see Vaish et al., 2008 for a review). But as children age, their outlook towards others becomes increasingly optimistic. From age 3, children show reluctance to attribute negative traits to actors without repeated behavioral observations (Boseovski & Lee, 2006, 2008). Shortly before school entry, however, is when children's positivity bias begins to peak ahead of easing slightly in later years (Heyman et al., 2003; Rholes & Ruble, 1984; see Boseovski, 2010 for a review). Aged 5, for example, children's impressions of an actor with a history of bad behaviour turn positive after hearing about a single good act (Rholes & Ruble, 1986). From this age, children also tend to interpret negative behaviours positively (Ruble et al., 1988) and ascribe norm-conforming desires to transgressors (Kalish & Shiverick, 2004) or claim their ignorance to the norm (Samland et al., 2016). But when transgressions involve some kind of moral offence, older preschoolers do not hesitate to evaluate offenders negatively. Five-year-olds, for example, selectively avoid transgressors and expect them to continue to cause harm (e.g., tear another picture) unless they apologise (Oostenbroek & Vaish, 2019; Waddington et al., 2022). This may be, in part, due to the attribution of hostile intentions (Choe et al., 2013; Dodge, 2006). Thus, late preschool represents an interesting period where children increasingly prefer to see others in a positive light while also retaining a certain degree of skepticism in their normative evaluations. Whether children at this age give ambiguous moral transgressors the benefit of the doubt remains an open question.

Another influence on children's propensity to assign benign intentions under ambiguous contexts could be whether this decision is being made individually or collectively. Adult studies, for example, suggest groups respond more rationally than individuals and reason about moral dilemmas in more complex ways (Kugler et al., 2012; Tassy et al., 2013). Much of the literature on children's intentionality judgments has focused on their responses in individual contexts (i.e., asking a single child what he/she thinks). Less is known about how they coordinate with others to yield collective judgments of intent, which feature commonly in society (e.g., juries). According to Piaget (1932), peer interactions represent a "safe harbor" for children to discuss and explore moral concepts. As such, preschool children usually engage in more active and in-depth reasoning with their peers than with their parents (Mammen et al., 2019; see also Mammen et al., 2018, 2021). Collective reasoning, especially among co-equal minds, is also argued to aid rational decision-making by enabling interactants to point out and correct for individual errors and mistakes (Conradt & Roper, 2003; but see Stanovich, 2011). Thus, reasoning alongside co-equal others might influence children's willingness to lend the benefit of the doubt and bring about different intentionality judgements.

In the present set of studies, we, therefore, investigated whether 5-year-olds give others the benefit of the doubt in their collective and independent moral judgments. Children evaluated moral transgressions either independently or jointly with an age-matched peer. In addition to their well-documented positivity and negativity bias at this age, we tested 5-year-olds as they reliably make intent-based normative judgments (Li & Tomasello, 2018) and understand mental state-based emotions such as surprise and happiness (Hadwin & Perner, 1991), and since children younger than 5-years often experience difficulties when talking about others' mental states during

collective reasoning (Hartwell et al., 2022; Köymen & Tomasello, 2018; Mammen et al., 2018, 2021; see Köymen & Tomasello, 2020 for a review).

1. Study 1

In Study 1, 5-year-old dyads witnessed moral transgressions (e.g., breaking someone's necklace) during which the transgressor's emotional response to the transgression varied: she either had a happy expression, a surprised expression or no expression at all. Dyads then decided on the cause of each transgression by selecting one of two cause pictures, one depicting an intentional transgression or one depicting an accidental one. If children give transgressors the benefit of the doubt, they would state the transgression was accidental more often in the surprised and no-expression conditions than in the happy condition. They would also produce more accidental justifications (e.g., "She didn't mean to") in the surprised and no-expression conditions than in the happy condition. Further, no difference between the surprised and no-expression conditions would be observed.

2. Method

2.1. Participants

The sample consisted of 108 5-year-olds (*M* = 5;6 [years; months], *Range* = 5;0–5;11, 54 girls, 54 boys) in 54 same-sex dyads. Children were recruited from a database of volunteer parents from a medium-sized German city and tested in their local kindergartens. Dyads were always formed of children from the same kindergarten group, meaning that they knew each other. Each dyad was randomly assigned to one of three conditions (18 dyads per condition). The sample size was based on previous studies with similar dyadic designs (e.g., Köymen et al., 2015; Köymen & Tomasello, 2018). One additional dyad was excluded due to equipment malfunction. Children were native German speakers with various socio-economic backgrounds. Parental consent was obtained before children's participation.

2.2. Materials

Each of the two warm-up and three experimental trials used a 3-page picture book. The first page was the *introduction* picture (e.g., someone carrying cups); the middle page was blank; the last page was the *outcome* picture (e.g., broken cups on the floor). Children were given two *cause* pictures to choose from for the middle page (see Appendix A).

2.3. Procedure

Testing took place in quiet rooms of nurseries. In the first warm-up trial, the experimenter (E) presented the first picture book. The *introduction* picture showed Max sitting at a table. Two *cause* pictures, Max eating a cookie and Max drinking juice were then introduced. Lastly, E revealed the *outcome* picture which contained cookie crumbs and asked children to place the correct cause picture on the blank middle page. E further asked, "You think Max ate a cookie because…?" to prime dyads to give reasons for their decisions. The second warm-up trial followed the same procedure. The *introduction* picture depicted Lisa and her sister. The *outcome* picture showed Lisa crying. The two *cause* pictures were Lisa's sister pulling Lisa's hair and Lisa's sister kissing Lisa. E drew children's attention to Lisa's facial expression in the *outcome* picture ("Look, Lisa is crying") to prime dyads to attend to facial expressions.

In the experimental trials, children played the same game but without E. In the first experimental trial, E introduced dyads to a picture book in which Anna was carrying a tray of her grandmother's cups (*introduction* picture) which later lay broken on the floor (*outcome* picture). E also introduced two cause pictures (see Appendix A). For the accidental cause picture, E said, "Anna tripped over the rug". For the intentional cause picture, E said, "Anna tilted the tray". E then instructed each dyad, "Tell each other what you see on the last page, then stick the right picture on the middle page. Did Anna do this intentionally or accidentally? The face on the last page gives you a clue" and left the room. Children were free to discuss the scenarios however they pleased. E waited until dyads had come to a collective agreement about which cause picture to use. Once a joint decision was reached, E returned and asked dyads to justify their chosen cause picture, "You chose [accidental/intentional] because…?". E gave no feedback.

The procedure was then twice repeated. In the second experimental trial, Tom ruined his friend's picture either accidentally (Tom knocked water over the picture while reaching for a pen) or intentionally (Tom poured water over the picture). In the third experimental trial, Caro broke her friend's necklace either accidentally (Caro caught the necklace on her hair pin) or intentionally (Caro pulled the necklace apart). The only difference between the three conditions was the facial expression of the protagonist in the *outcome* picture. The protagonist had a pleased face in the happy condition, a shocked face in the surprised condition, and a blank face in the no-expression condition. The stories used in the experimental trials were chosen because of their clear depiction of harm (i.e., broken cups with scattered shards of glass, a broken necklace with beads bestrewn, a ruined picture soaked in water) and because intentional and accidental "cause" pictures could be easily portrayed (i.e., tilting the tray of cups vs. tripping over a rug, pulling the necklace to break it vs. catching it on a hair pin, pouring water over the picture vs. knocking water over it). The order in which the stories and the cause pictures were presented was counterbalanced.

2.4. Coding

Children's peer conversations and their responses to E's *why*-question (i.e., why they chose that particular cause picture) were transcribed verbatim. First, we coded which cause picture children chose (accidental vs. intentional). Next, we coded children's justifications for the cause of the mishap. Each trial received one of the following codes (one for children's peer conversations and one for their response to E's *why*-question):

1) Accidental, if children:
   o referred to the protagonist's facial expression as being sad/surprised (e.g., "She looks sad");
   o described the scene as an accident, such as:
      ▪ using unintentional verbs like "stumbled" or "tripped" in active voice (e.g., "She stumbles over here somewhere");
      ▪ providing excuses for the transgression (e.g., "Because she is running too fast");
      ▪ used passive voice to de-emphasize the role of the protagonist in the transgression (e.g., "The cups are broken", see Budwig, 1995; Köymen et al., 2014, 2015).
2) Intentional, if children:
   o referred to the protagonist's facial expression as being happy (e.g., "Because, look, he laughs");
   o described the scene as intentional, such as:
      ▪ using intentional verbs like "tear", "pour", "tilt/dump" in active voice (e.g., "He dumped water on it");
      ▪ used active voice without any accidental adverbs (e.g., "She broke it").
3) No justification, if children:
   o gave no justification or ambiguous/circular arguments (e.g., "My picture is correct").

Rarely, dyads provided both accidental and intentional justifications in a single trial. In those cases, we coded children's last justification since the latter often corrected the former. A second coder, blind to predictions, coded 22 % of the data (12 dyads; 4 per condition). Agreement was $\kappa = .88$ for children's peer conversations and $\kappa = .89$ for their responses to E's *why*-question, respectively.

3. Results

The studies were not preregistered. All deidentified data and statistical scripts are publicly available at: https://osf.io/x83za.

First, we analyzed whether children's choice of cause picture differed by condition. Using one-sample t-tests, we compared the number of trials (out of 3) in which children chose the accidental cause picture to chance (1.5) in each condition. This statistical approach enabled us to determine whether children preferred one option significantly more than the other within each condition. In the happy condition, children chose the accidental cause picture significantly below chance ($t(17) = -6.45$, $p < .001$, $d = 1.52$, see Fig. 1). In the surprised condition, children chose the accidental cause picture significantly above chance ($t(17) = 3.90$, $p = .001$, $d = 0.92$). In the no-expression condition, children chose the accidental cause picture at chance ($t(17) = -0.37$, $p = .719$, $d = 0.09$). We also compared across conditions using a between-subjects ANOVA. The response variable was the number of trials in which dyads chose the accidental cause picture. Predictors included condition (happy, surprised, no-expression) only, the main effect of which was significant ($F(2,51) = 17.58$, $p < .001$, $\eta^2 = .41$, see Fig. 1). As compared to dyads in the happy condition, dyads in the surprised and no-expression conditions chose the accidental cause picture significantly more often (Tukey HSD, $p < .001$; $p = .008$, respectively). Dyads in the surprised condition chose the accidental cause picture significantly more often than dyads in the no-expression condition (Tukey HSD, $p = .019$).[1]

Second, we analyzed whether children's spontaneous justifications in their peer conversations differed across conditions using Generalized Linear Mixed Models (GLMMs) with binomial error distribution. Forty-four dyads (out of 54) produced justifications in 79 out of 162 trials (49 %). The unit of analysis was each trial. The response variable was the binary measure of whether children produced intentional or accidental justifications in each trial. The full model included condition (happy, surprised, no-expression), trial order (1−3) and story type (cups, necklace, picture) as fixed factors, and the random factor of dyad. The null model included the same factors, except for condition. Model comparison revealed a significant main effect of condition ($\chi^2 = 30.74$, $df = 2$, $p < .001$) and story type ($\chi^2 = 7.60$, $df = 2$, $p = .022$). As compared to dyads in the happy condition, dyads in the surprised and no-expression conditions were more likely to produce accidental justifications ($z = 3.38$, $p < .001$; $z = 2.93$, $p = .003$, respectively), whereas no difference was observed between the surprised and no-expression condition ($z = 1.52$, $p = .129$, see Fig. 2). Children provided more spontaneous accidental justifications for the cups story than for the necklace story ($z = 2.52$, $p = .012$), whereas neither differed from the picture story (z's < 1.49, p's > .135).

Third, we analyzed whether children's justifications to E's *why*-question differed across conditions. All dyads produced justifications in 161 out of 162 trials (99 %). The models were the same as in the previous analysis. Model comparison revealed a significant main effect of condition ($\chi^2 = 57.21$, $df = 2$, $p < .001$). As compared to dyads in the happy condition, dyads in the surprised and no-expression conditions were more likely to produce accidental justifications ($z = 4.58$, $p < .001$; $z = 4.00$, $p < .001$, respectively). Dyads produced more accidental justifications in the surprised condition than in the no-expression condition ($z = 2.33$, $p = .020$, see

---

[1] In the Supplementary Material, children's choice of cause picture was further analyzed using Generalized Linear Mixed Models (GLMMs), which revealed the same results.
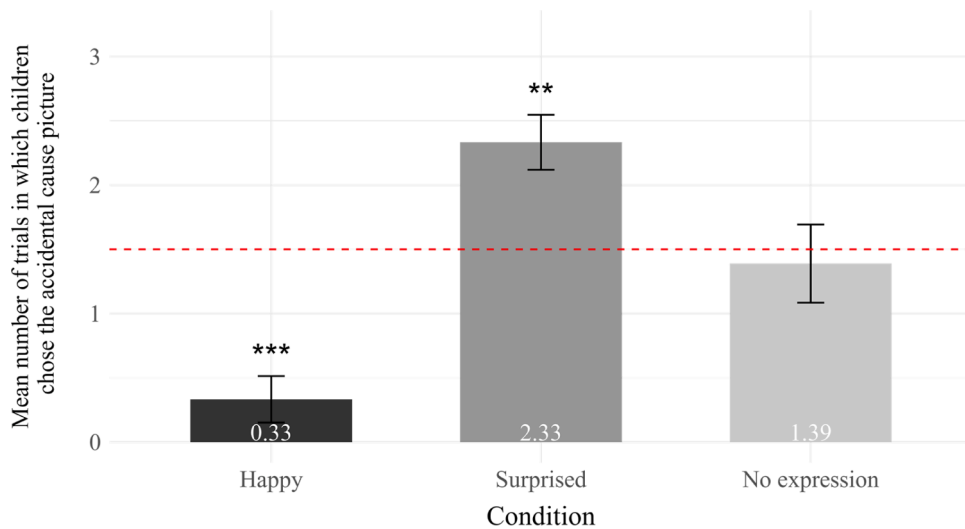
Fig. 1. The mean number of trials in which children chose the accidental cause picture. Error bars show standard error. ** $p < .01$, *** $p < .001$ compared to chance.
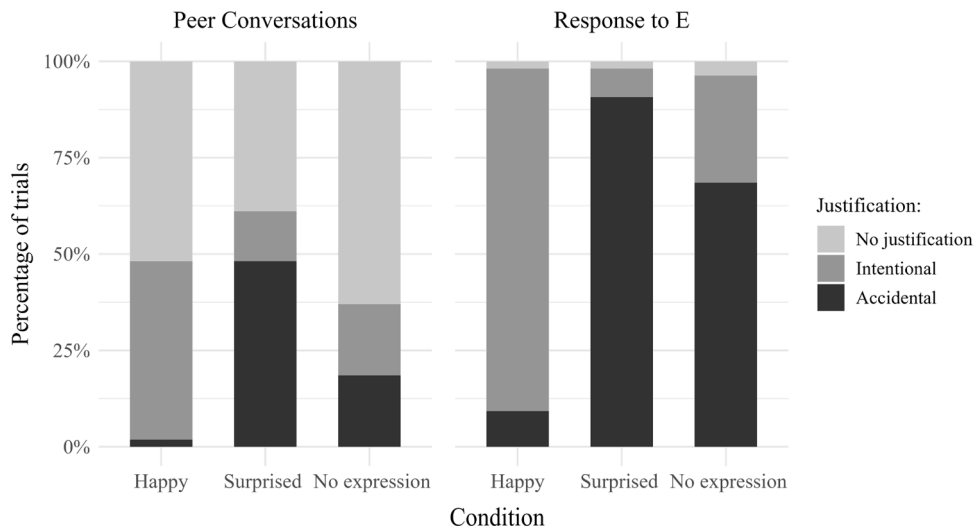
Fig. 2. The percentage of trials in which children produced accidental and intentional justifications in their peer conversations and responses to the experimenter (E).

Fig. 2).

4. Discussion

Five-year-old dyads appropriately inferred the transgressors' intentions based on their emotional responses to the harm caused. In line with existing work (Nummer-Winkler & Sodian, 1988; Yuill, 1984; Yuill & Perner, 1988), when the transgressor was pleased or surprised, dyads correctly chose the cause picture depicting an intentional or accidental harm, respectively. However, when the transgressor showed no emotion and all intent-based information was withheld, children performed at chance. Although dyads in the no-expression condition selected the accidental cause picture more often and gave more accidental justifications (both in their peer conversations and responses to E's *why*-question) than dyads in the happy condition, they did not quite follow the same pattern as dyads in the surprised condition, who gave more accidental justifications and selected the accidental cause picture further still. Thus, whether children gave the no-expression transgressor the benefit of the doubt remained unclear.

Children produced justifications much less frequently in their peer conversations (43 % of trials) than in their responses to E's *why*-question (99 %). There were several potential factors that reduced children's talk and their production of reasons. First, from the dyads' perspective, the task simply involved choosing one of two cause pictures. The physical nature of the task meant dyads often

rushed to deposit a picture onto the middle page of the picture book without necessarily discussing it. Alternatively or additionally, peers may have relied on their common ground about which cause picture was most appropriate (i.e., the correct picture was obvious to both partners, at least in the happy and surprised conditions) thus reducing the need to produce justifications for their proposals (Köymen et al., 2016; Mammen et al., 2018). The inclusion of E's *why*-question therefore enabled us to understand how the majority of dyads made and reached their decisions where we may not have known otherwise, and speaks to the prevalence with which children use common ground assumptions to streamline their justificatory reasoning.

Two further factors likely affected children's responses in the no-expression condition. First, having the intentional cause picture as an option might have reminded children that ill-intentioned people exist, whereas this idea may not have crossed their minds had it not been visually reinforced. Second, the scenario in the latter warm-up trial possibly primed children's responses in the no-expression condition. In this scenario, Lisa was shown crying because her sister had pulled her hair. With no emotional markers to inform whether the harm was (un)intentional, children in the no-expression condition might thus have chosen the intentional cause picture more often than they would have otherwise. Thus, we addressed these limitations in a follow-up study.

5. Study 2

In Study 2, we made three changes to the procedure. First, we created new material for the warm-up trials. These depicted neutral (non-moral) acts so as not to bias children's later judgments. Second, we removed the cause pictures from the experimental trials and asked children to come up with a cause for the transgressions themselves. This way, children could reason about the cause of the mishap without relying on visual aids. Finally, we dropped the surprised condition because (1) if children give the benefit of the doubt, there would be no difference between the no-expression and surprised conditions, and (2) children at this age already recognise that shocked expressions imply accidental harm (e.g., Waddington et al., 2023; Yuill & Perner, 1988). We predicted that if children give transgressors the benefit of the doubt, they would state the transgression was accidental significantly above chance in the no-expression condition and would do so more often than dyads in the happy condition. They would also produce more accidental justifications in the no-expression condition than in the happy condition.

6. Method

6.1. Participants

The sample consisted of 72 5-year-olds ($M = 5;5$, *Range* = 5;0–5;11, 36 girls, 36 boys) in 36 same-sex dyads who knew each other from attending the same kindergarten group. Each dyad was randomly assigned to one of two conditions (18 dyads per condition). The sample size aligned with previous research using similar dyadic designs (e.g., Köymen et al., 2015; Köymen & Tomasello, 2018). Three additional dyads were excluded due to equipment malfunction (2) and experimental error (1). Recruitment was the same as Study 1, though in a different German city. Children had not previously participated in Study 1, were native German speakers, with various socio-economic backgrounds.

6.2. Materials

For the warm-up trials, two new picture books were created (see Appendix B). For the experimental trials, the same picture books from Study 1 were used without the cause pictures.

6.3. Procedure

Testing took place in quiet rooms of nurseries. The structure of the warm-up trials was the same as Study 1, only the stories differed. In the first warm-up trial, the *introduction* picture showed Moritz and his mom on her birthday. Two *cause* pictures, Moritz and his dad baking a cake and them buying a cake were then introduced. Lastly, the *outcome* picture of a birthday cake followed. In the second warm-up trial, the *introduction* picture depicted Lisa and her sister donning rubber boots. The *outcome* picture showed two pairs of wet boots. The two *cause* pictures were Lisa and her sister jumping into puddles and them building a dam.

In the first experimental trial, E introduced the cups story taken from Study 1. Dyads were first shown the *introduction* picture, "Anna is carrying cups" then the *outcome* picture, "The cups are broken". In the happy condition, E added, "And look, Anna is smiling". In the no-expression condition, E added, "And look, you cannot see Anna's face". E then asked, "What do you think? Why are the cups broken?" and left the room. Again, children were free to discuss the scenarios however they pleased. When dyads reached agreement, E returned and asked, "Why do you think the cups are broken?" After dyads answered, E replied, "Do you think she wanted to break them or did that happen accidentally?" The procedure was then twice repeated for the necklace and picture story.

The order in which the stories were presented was counterbalanced, as was the cause question after each trial (accidental cause question first vs. second).

6.4. Coding

Coding was the same as in Study 1. A second coder, blind to predictions, coded 22 % of the data (8 dyads; 4 per condition). Agreement was $\kappa = .78$ for children's peer conversations and $\kappa = .85$ for their responses to E's *why*-question, respectively.

7. Results

First, we analyzed whether children's choice of cause differed by condition. Using one-sample t-tests, we compared the number of trials (out of 3) in which children gave an accidental response to chance (1.5) in each condition. In the happy condition, children gave an accidental response at chance level ($t(17)$ = -1.79, p = .092, d = 0.42, see Fig. 3). In the no-expression condition, children gave an accidental response significantly above chance ($t(17)$ = 12.50, p < .001, d = 2.95). We also compared across conditions using a between-subjects ANOVA. The response variable was the number of trials in which children responded with an accidental cause. Predictors included condition (happy, no-expression) only, the main effect of which was significant ($F(1,34)$ = 39.30, p < .001, η^2 = .54). Children gave an accidental response significantly more often in the no-expression condition than in the happy condition (see Fig. 3).[2]

Second, we analyzed whether children's spontaneous justifications in their peer conversations differed across conditions using GLMMs with binomial error distribution. All dyads produced justifications in 102 out of 108 trials (94 %). The models were the same as in Study 1, except the fixed factor, condition, consisted of two levels (happy, no-expression). Model comparison revealed a significant main effect of condition (χ^2 = 16.56, df = 1, p < .001) and story type (χ^2 = 10.13, df = 2, p = .006). Dyads in the no-expression condition produced more accidental justifications than dyads in the happy condition (see Fig. 4). Dyads gave more accidental justifications for the cups story than for the necklace story (z = 2.82, p = .005), whereas neither differed from the picture story (z's < 1.70, p's > .090).

Third, we analyzed whether children's justifications to E's *why*-question differed across conditions. All dyads produced justifications across all three trials. The models were the same as in the previous analysis. Model comparison revealed a significant main effect of condition (χ^2 = 23.68, df = 1, p < .001) and story type (χ^2 = 14.77, df = 2, p < .001). Dyads in the no-expression condition produced more accidental justifications than dyads in the happy condition (see Fig. 4). Dyads gave more accidental justifications for the cups story than for the necklace story (z = 2.96, p = .003), whereas neither differed from the picture story (z's < 1.89, p's > .059).

To identify whether the procedural changes in Study 2 resulted in global increases in children's accidental judgments as compared to Study 1, we ran an additional exploratory between-subjects ANOVA. The response variable was the number of trials in which dyads responded with an accidental cause. Predictors included study (Study 1, Study 2), condition (happy, no-expression), and their interaction. A marginally significant interaction effect was observed ($F(1,68)$ = 3.22, p = .077, η^2 = .05). Dyads gave more accidental responses in the no-expression condition of Study 2 than Study 1 (Tukey HSD, p < .001). In the happy condition, dyads' accidental responses did not vary between Studies 1 and 2 (Tukey HSD, p = .055).

8. Discussion

When dyads produced a possible cause for the transgressions themselves, children gave more accidental judgments overall. In line with literature on children's tendency to describe transgressions in less negative ways (e.g., Kalish & Shiverick, 2004), dyads in the no-expression condition, and to a lesser extent in the happy condition, were more likely to attribute an accidental cause compared to Study 1. Importantly, however, children stated that the transgression was an accident significantly more often in the no-expression condition than in the happy condition. In fact, of the 18 dyads in the no-expression condition, 17 stated that the transgression was accidental across all three trials. Children's performance in the present no-expression condition was largely equivalent to that of the surprised condition in Study 1 (see Figs. 1, 3).

The methodological changes in Study 2 not only shifted children's accidental judgments in the no-expression condition from chance to above chance, but also the happy condition from below chance to near chance (see Fig. 3). This raises the question of whether children judge happy transgressors to have intentionally transgressed as robustly as suggested (e.g., Waddington et al., 2023). Though more accidental judgments were made in the happy condition of Study 2, further analysis suggested this increase was not significantly different from the rates found in the corresponding condition of Study 1.

Children also gave more spontaneous justifications in Study 2 than Study 1, even though the kinds of justifications they produced followed a similar pattern. Dyads were more likely to give accidental justifications in the no-expression condition (at near ceiling) than in the happy condition. The removal of the cause pictures from Study 1 perhaps gave children more freedom to discuss the intentional structure of the moral acts before them, while also making it harder for them to rely on common ground assumptions (e.g., Köymen et al., 2016). Study 2 thus elicited more accidental judgments and justifications than Study 1.

Both Studies 1 and 2 focused on children's collective moral judgments. To further examine the robustness of their tendency to lend the benefit of the doubt under ambiguous contexts, we observed children making independent moral judgments in Study 3.

9. Study 3

We ran an "independent" version of Study 2, in which children were interviewed individually by an experimenter and asked to decide whether the witnessed transgressions were intentional or accidental. We predicted that if children give transgressors the benefit of the doubt, they would state the transgression was accidental significantly above chance in the no-expression condition and would do so more often than in the happy condition. Children would also produce more accidental justifications for the experimenter in the no-

---

[2] In the Supplementary Material, children's choice of cause was further analyzed using Generalized Linear Mixed Models (GLMMs), which revealed the same results.
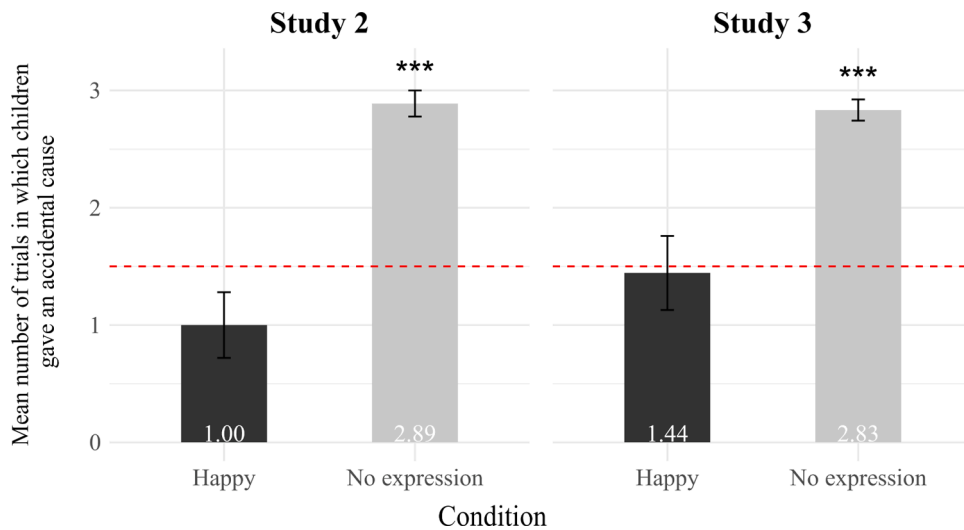
Fig. 3. The mean number of trials in which children gave an accidental cause when reasoning collectively (Study 2) or independently (Study 3). Error bars show standard error. *** $p < .001$ compared to chance.
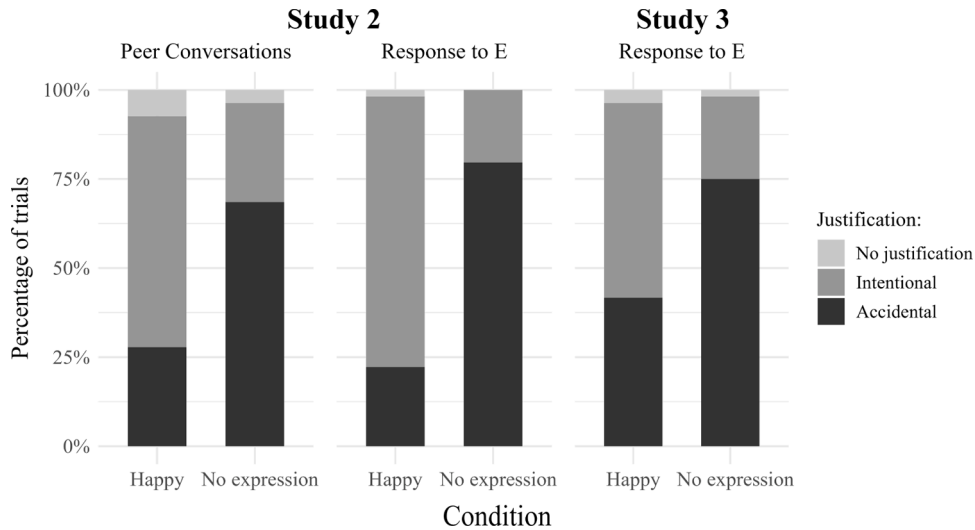
Fig. 4. The percentage of trials in which children produced accidental and intentional justifications in their peer conversations in Study 2 and responses to the experimenter (E) in Studies 2 and 3.

expression condition than in the happy condition. We further compared children's patterns of responding across Studies 2 and 3. Specifically, we explored whether children's choices and justifications differed when reasoning with a peer (Study 2) versus when reasoning independently (Study 3).

10. Method

10.1. Participants

The sample consisted of 36 5-year-olds ($M = 5;5$, *Range* = 5;0–5;11, 18 girls, 18 boys). Children were randomly assigned to one of two conditions (18 children per condition). The sample size was determined to enable for cross-study comparisons. One additional child was excluded due to experimental error. Recruitment was the same as Study 2. Children had not participated in earlier studies, were native German speakers, with various socio-economic backgrounds.

10.2. Materials

The same picture books from Study 2 were used.

10.3. Procedure

The procedure was the same as Study 2, except children participated individually. In the experimental trials, E introduced the story, the *introduction* picture, and then the *outcome* picture. Next, E asked the child, "What do you think? Why is [the object] broken?" and "Do you think she wanted to break [the object] or did that happen accidentally?" The procedure was then twice repeated for the other two stories.

10.4. Coding

Coding was the same as Study 1. A second coder, blind to predictions, coded 100 % of the data. Agreement was $\kappa = .74$ for children's responses to E's *why*-question.

11. Results

First, we analyzed whether children's choice of cause differed by condition. Using one-sample t-tests, we compared the number of trials (out of 3) in which children gave an accidental response to chance (1.5) in each condition. In the happy condition, children's responses did not differ from chance, meaning they provided accidental and intentional responses equally often ($t(17) = -0.18$, $p = .862$, $d = 0.04$, see Fig. 3). In the no-expression condition, children gave an accidental response significantly above chance ($t(17) = 14.75$, $p < .001$, $d = 3.48$). We also compared across conditions and whether or not children reasoned with a peer partner (Studies 2 and 3) using a between-subjects ANOVA. The response variable was the number of trials in which children responded with an accidental cause. Predictors included condition (happy, no-expression), partner (dyad in Study 2, independent in Study 3), and their interaction. Only the main effect of condition was significant ($F(1,68) = 54.14$, $p < .001$, $\eta^2 = .22$). Children and dyads gave an accidental response significantly more often in the no-expression condition than in the happy condition (see Fig. 3).[3]

Finally, we analyzed whether children's justifications differed across conditions and partners using GLMMs with binomial error distribution. The unit of analysis was each trial. The response variable was the binary measure of whether children produced intentional or accidental justifications in each trial. The full model included condition (happy, no-expression), partner (dyad in Study 2, experimenter in Study 3), their interaction, trial order ($1-3$) and story type (cups, necklace, picture) as fixed factors, and the random factor of dyad. The null model included trial order, story type and the random factor of dyad. The full model improved the fit ($\chi^2 = 30.06$, $df = 3$, $p < .001$). However, the interaction between condition and partner was not significant ($\chi^2 = 0.75$, $df = 1$, $p = .385$). The reduced model without this interaction term revealed significant main effects of condition ($\chi^2 = 23.52$, $df = 1$, $p < .001$), partner ($\chi^2 = 7.65$, $df = 1$, $p = .006$) and story type ($\chi^2 = 18.19$, $df = 2$, $p < .001$). Children gave more accidental justifications in the no-expression condition than in the happy condition (see Fig. 4). Children also gave more accidental justifications when interviewed by an adult experimenter (Study 3) than when they conversed with a peer (Study 2). Children gave more accidental justifications for the cups story than for the necklace story ($z = 3.78$, $p < .001$) and picture story ($z = 2.89$, $p = .004$), whereas the necklace and picture stories did not differ ($z = -1.13$, $p = .258$).

12. Discussion

The results of Study 3 replicated the findings of Study 2 in an independent context, showing that young children give transgressors the benefit of the doubt in both their collective and individual moral reasoning. When interviewed alone, children in the no-expression condition routinely attributed accidental causes to the witnessed transgressions. In comparison, children's judgments in the happy condition were at chance, meaning they gave accidental and intentional causes equally frequently. Although similar response patterns were observed across both studies, children in Study 3 were numerically (though not significantly) less likely to attribute intentional harm when making moral judgments independently in the happy condition (see Fig. 3). This was further reflected in the justifications they provided for their judgments: children produced more accidental justifications when talking to an adult experimenter (Study 3) than when talking to a peer (Study 2).

One explanation for this slight change in tendency to give the benefit of the doubt could be due to the company children kept. Specifically, whether children were able to reason dyadically with a peer or did so individually in the presence of an adult. Reasoning alongside a "co-equal mind" is known to influence children's decision-making and bring about different judgements than when reasoning with an adult (Piaget, 1932; Mammen et al., 2019). Whether collective reasoning helps or hinders decision-making as a whole, however, is subject to ongoing debate since it can either amplify (Stanovich, 2011) or diminish mistakes made by group members (Conradt & Roper, 2003). Anecdotally, children's joint reasoning in the current studies was largely complementary as it enabled dyads to point out to one another information that may have been initially overlooked. For example, one child failed to notice

---

the transgressor's facial expression while engaging with the task. Noticing this omission, her partner drew her attention to the expression which resulted in a more informed conclusion being drawn, in that all available information had been accounted for. Although children often arrived at the same conclusions whether reasoning collectively or independently, such adjustments in peer discussions suggest children may benefit from reasoning with another co-equal mind when making moral judgments. A second complementary account could be that children were less comfortable "accusing" the transgressor of intentional harm when reasoning independently than when reasoning together with a partner as part of a "we" in which responsibility for the decision was shared (Tomasello, 2019).

13. General discussion

Different theories propose (Strohminger et al., 2017) and oppose (Ross, 1977) that individuals give the benefit of the doubt as a psychological tendency. Across three studies, we investigated whether and under what circumstances children lend others the benefit of the doubt. Our results suggest that when transgressors were without expression and all intention information was withheld, children concluded the harms were accidental, especially in Studies 2 and 3. The leniency shown to the expressionless transgressor was further reflected in children's justifications. If the transgressor had no expression, children often downplayed her intentions (e.g., "She stumbled") and her role in the harm (e.g., "The cups are broken"), and even came up with excuses on her behalf (e.g., "Because she was running too fast"). Thus, in the absence of intent-based information, 5-year-olds assigned positive intentions and gave transgressors the benefit of the doubt.

Across all three studies, children were more likely to ascribe benign intentions to the expressionless transgressor than to the happy transgressor. However, in Studies 2 and 3, children's accidental judgements in the happy condition were at chance level, meaning accidental causes were assigned as often as intentional ones (see Fig. 3). Although children were more inclined to give the expressionless transgressor the benefit of the doubt, this tendency thus also extended, albeit to a lesser extent, to the transgressor pleased with her actions. Giving the benefit of the doubt to happy agents may suggest that 5-year-olds do not consider positive expressions following harm to necessarily signify negative intent (Waddington et al., 2023) and may, instead, appraise joy more in the context of goal-fulfilment; of simply achieving a desired outcome (e.g., Nummer-Winkler & Sodian, 1988; Yuill, 1984; Yuill et al., 1996). However, had children considered happiness to be unrelated to the transgressor's intentions, we would expect the rate of accidental attributions to be consistent across the happy and no-expression conditions, which was never the case. Rather, happy transgressors were more often criticized by children, both in the intentional judgments they made and in the justifications they gave (e.g., "She broke it"). Thus, children were not wholly against ascribing ill intentions. However, their desire to see others optimistically at this age might be so pervasive that it applies, at least somewhat, to contented transgressors.

Our findings have a number of implications. First, whether young children are positively- or negatively-valenced in their social evaluations is subject to ongoing debate. On the one hand, children exhibit a positivity bias in the personality judgments they make (Boseovski & Lee, 2006; Rholes & Ruble, 1986) and in the stories they re-tell (Kalish & Shiverick, 2004; Samland et al., 2016). On the other hand, young children show a robust negativity bias in their social-emotional evaluations (see Vaish et al., 2008 for a review) and in their perceptions of unapologetic transgressors (Oostenbroek & Vaish, 2019; see also Waddington et al., 2022). Our findings in the no-expression condition support the view that young children prefer to see the best in others in the context of ambiguous moral intent. In Studies 2 and 3, participants even extended this positivity bias, albeit to a lesser extent, to transgressors who seemed pleased with their actions, demonstrating the lengths children will go to at this age to shine a positive light on others (Boseovski, 2010).

Second, research involving intention ascription has focused almost exclusively on children's responses in individual contexts (Cushman et al., 2013; Killen et al., 2011; Proft & Rakoczy, 2019; Waddington et al., 2023). In the present set of studies, we examined children's intentionality judgements when they were made collectively with a peer, and how these judgements compared when made independently. Reasoning alongside others features commonly in society (e.g., juries) and has been found to influence children's moral decision-making (e.g., Piaget, 1932; Mammen et al., 2019). However, our results showed that children's responses were largely consistent across both contexts. That said, a slight aversion to attributing intentional harm was present when reasoning alone (see Figs. 3, 4). Children's peer discussions also showed signs of the benefits of collective reasoning, in that partners actively pointed out one another's mistakes (Conradt & Roper, 2003). Notwithstanding these nuances, children's tendency to offer transgressors the benefit of the doubt operates across both individual and collective domains of reasoning.

Third, we make methodological contributions. Several studies have previously examined children's moral judgments without specifying the transgressor's emotional response (whether they felt pleased or surprised) in the stories told to participants (e.g., Cushman et al., 2013; Killen et al., 2011). But failing to describe a transgressor's emotions is not the same as systematically removing them. For one, different children may attribute different emotions when they are left unspecified, whereas explicitly removing them ensures that children are reasoning from the same input conditions. By deliberately omitting the transgressor's emotions, the present set of studies help to overcome a potentially confounding source of variability in children's action interpretations and intention ascriptions. Norm violators in previous studies have further been depicted or were instructed to show neutral emotion in their expressions to make their intentions appear perceptually ambiguous (e.g., Oostenbroek & Vaish, 2019; Waddington et al., 2022). These "neutral" expressions, however, are seldom perceived as neutral (Russell & Fehr, 1987; Albohn & Adams, 2021) and often include traces of negative emotion (Lee et al., 2008). Removal of the transgressor's facial expression in the present studies thus enabled us to observe children's intent-based inferences and tendency to offer the benefit of the doubt under truly dispassionate conditions. Moreover, in most of these studies, children were presented with an entire story before being asked for their appraisals (Cushman et al., 2013; Killen et al., 2011; Proft & Rakoczy, 2019; Waddington et al., 2022, 2023). Here, children influenced the narrative for themselves. That is, children were presented with the beginning and end of each story and asked to determine how the outcome came about.

This "bridge-the-gap" paradigm represents a new and novel approach to investigating the reasoning behind children's moral judgments.

By showing that 5-year-olds give transgressors the benefit of the doubt, the present set of studies make a clear contribution to the literature on children's moral reasoning. That said, our study is not without its limitations. One potential concern might be our coding scheme, in which short justifications pertaining to the transgressor's facial expression (e.g., "She looks sad"), children's use of passive language (e.g., "The cups are broken"), and negligence-related excuses (e.g., "Because she is running too fast") were considered evidence of accidental judgments. It is true that passive voice does not necessarily imply a lack of intentionality. It does, however, clearly shift attention away from the agent (or, at least, attempts to) thereby de-emphasizing her active role in the harm caused (Budwig, 1995; Köymen et al., 2014). Research further shows that preschool-aged children only take negligence into account for accidental transgressions and not intentional ones (Nobes et al., 2009). Referring to the transgressor's carelessness thus implies an accidental judgement. Additionally, children are known to provide partners with optimally informative reasons which are based on common ground assumptions; for example, arguing that an individual ought to be punished because "He stole" without further elaborating that stealing is wrong (Mammen et al., 2018; see also Köymen et al., 2016). In fact, it would be pragmatically odd for individuals to be overly informative in their reasoning (e.g., explaining why stealing is wrong, see Clark, 1996; Grice, 1989). In the current studies, then, dyads presumably relied on what they perceived to be in their moral common ground (e.g., that looking "sad" indicates a lack of intent) to inform their peer conversations. Future research may wish to consider exploring the effects that optimally and overly informative reasons have on children's dyadic reasoning. A more thorough investigation into children's use of active and passive language to convey others' intentions is also advised.

Another methodological caveat is the extent to which the independent paradigm in Study 3 was truly independent. To enable comparison between Studies 2 and 3, the role of and input from the experimenter remained consistent across both studies. The picture books were introduced and attention was drawn to the transgressors' facial expressions (or lack thereof) before then eliciting children's views (e.g., "Why is [the object] broken?"). At no other point did the experimenter advise, revise nor otherwise influence the child's decision-making. Thus, children did, or so we might argue, reason about the scenarios independently. But in administering the study, some degree of social interaction between the child and experimenter was inevitable thus limiting the extent to which the task was strictly individual. More discreet methods for collecting children's responses could be used in future work comparing their independent and collective reasoning; for example, implementing a "thinking-out-loud" procedure.

Of further concern were the story-specific effects observed in each study. Children's tendency to lend the benefit of the doubt both in their judgements and justifications often differed across stories. One possible explanation for this change in tendency could be due to the content of the vignettes and the nature of the harms therein. Across the different stories, for example, the relationship between the victim and transgressor varied. In the cups story, the broken cups were said to belong to the transgressor's grandmother. For the picture and necklace stories, the ruined objects belonged to the transgressor's friend. Moreover, certain transgressions were perhaps easier for children to identify and produce an accidental cause for the damage done (e.g., stumbling over the rug while carrying the cups). Broader contextual factors thus appear to influence the extent to which children give transgressors the benefit of the doubt. Future research examining children's general reasoning biases in their intention ascriptions may wish to ensure that vignettes are as closely matched as possible.

The purpose of the present set of studies was to identify whether preschoolers offer the benefit of the doubt in their independent and collective moral judgments. However, since we only tested one age group, this leaves open the ontogenetic question of how children's tendency to give others the benefit of the doubt changes with age. We do know, however, that children's evaluations of third-party transgressions and their emotional consequences continue to develop into the school years (e.g., Cushman et al., 2013; Killen et al., 2011; Nummer-Winkler & Sodian, 1988). Their forgiveness of accidental harms also increases between ages 5 and 10 (Amir et al., 2021). Additionally, children might begin to offer the benefit of the doubt from an earlier age than presently tested. Already by age 3, children are fairly reluctant to attribute negative traits to others (Boseovski & Lee, 2006). Future research should, therefore, test a broader age range to examine whether and to what extent children's tendency to lend the benefit of the doubt varies across different stages of development.

The present findings show young children give transgressors the benefit of the doubt when seen damaging other's property. Another avenue for future work could thus be to examine the extent to which our findings generalise across other kinds of transgressive acts, especially those that may be harder to attribute accidental causes (e.g., bullying, name calling). Moreover, how our results compare to actions that have neutral or positive outcomes (e.g., being accidentally helpful) could be further explored. Based on current and existing evidence (Boseovski & Lee, 2006; Kalish & Shiverick, 2004), children are likely to ascribe benign intentions when given the opportunity.

Building on the present findings, another interesting direction for future research is to investigate how children's tendency to give the benefit of the doubt relates to their moral evaluations of the act and the agent. When juries offer defendants the benefit of the doubt, this ordinarily implies that the defendant should not be punished so harshly (i.e., a reduced sentence), if at all. In the present studies, children were asked to infer the transgressor's intentions and rather assumed an accidental cause when no intention information was made available. Children were not asked, however, to evaluate the transgressor's blameworthiness nor whether she ought to be punished. Developmental research going back to Piaget (1932) and forward to Cushman et al. (2013) suggest young children often focus on the outcomes of actions than on the actors' intentions when assigning blame and punishment. A plausible implication of this pattern of responding is that children might not respond in the same way a traditional jury would; they might judge a transgressor who causes damage as "wrong" and/or deserving of punishment even if they acknowledge that it was caused accidentally. However, other studies have found preschool children feature others' mental states in their moral judgments (Vaish et al., 2010) and advocate for accidental transgressors to be punished less than intentional ones (e.g., Nobes et al., 2009). In either case, the present studies show

children give transgressors the benefit of the doubt in their intentionality judgments, but cannot tell us whether these judgments then translate to less blame and punishment.

14. Conclusion

In summary, to assume good intentions and give the benefit of the doubt is a sensible social strategy which binds individuals and groups together. When intentions cannot be easily established, we showed that preschoolers prefer to ascribe positive intent and see the best in others, both in their individual and collective reasoning. Much like when juries lack clear evidence, 5-year-olds, too, lend the benefit of the doubt.

CRediT authorship contribution statement

Marina Proft: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Owen Waddington:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Bahar Köymen:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cogdev.2025.101634.

Data availability

All deidentified data and statistical scripts are publicly available at: https://osf.io/x83za.

References

Albohn, D. N., & Adams, R. B. (2021). Emotion residue in neutral faces: implications for impression formation. *Social Psychological and Personality Science, 12*, 479–486. https://doi.org/10.1177/1948550620923229

Amir, D., Ahl, R. E., Parsons, W. S., & McAuliffe, K. (2021). Children are more forgiving of accidental harms across development. *Journal of Experimental Child Psychology, 205*, Article 105081. https://doi.org/10.1016/j.jecp.2020.105081

Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development, 59*, 1611–1622. https://doi.org/10.2307/1130675

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": an examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4*, 212–218. https://doi.org/10.1111/j.1750-8606.2010.00149.x

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–513. https://doi.org/10.1037/0012-1649.42.3.500

Boseovski, J. J., & Lee, K. (2008). Seeing the world through rose-colored glasses? Neglect of consensus information in young children's personality judgments. *Social Development, 17*, 399–416. https://doi.org/10.1111/j.1467-9507.2007.00431.x

Budwig, N. (1995). *A Developmental-functionalist Approach to Child Language*. Mahwah, NJ: Erlbaum.

Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology, 49*, 2245–2256. https://doi.org/10.1037/a0032293

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Conradt, L., & Roper, T. J. (2003). Group decision-making in animals. *Nature, 421*, 155–158. https://doi.org/10.1038/nature01294

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127*, 6–21. https://doi.org/10.1016/j.cognition.2012.11.008

Dodge, K. A. (2006). Translational science in action: hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*, 791–814. https://doi.org/10.1017/S0954579406060391

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hadwin, J., & Perner, J. (1991). Pleased and surprised: children's cognitive theory of emotion. *British Journal of Developmental Psychology, 9*, 215–234. https://doi.org/10.1111/j.2044-835X.1991.tb00872.x

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–559. https://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13*, 923–929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Hartwell, K., Brandt, S., Boundy, L., Barton, G., & Köymen, B. (2022). Young children's use of meta-talk to make rational collaborative decisions. *Child Development, 93*, 1061–1071. https://doi.org/10.1111/cdev.13750

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, NY: Wiley.

Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development, 74*, 516–534. https://doi.org/10.1111/1467-8624.7402013

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development, 19*, 401–416. https://doi.org/10.1016/j.cogdev.2004.05.004

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*, 197–215. https://doi.org/10.1016/j.cognition.2011.01.006

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*, 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*, 423–429. https://doi.org/10.1037/dev0000089

Köymen, B., Schmidt, M. F., 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

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

Köymen, B., & Tomasello, M. (2020). The early ontogeny of reason giving. *Child Development Perspectives, 14*, 215–220. https://doi.org/10.1111/cdep.12384

Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*, 471–482. https://doi.org/10.1002/wcs.1184

Lee, E., Kang, J. I., Park, I. H., Kim, J. J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157*, 77–85. https://doi.org/10.1016/j.psychres.2007.02.005

Lefebvre, J. P., & Krettenauer, T. (2020). Is the true self truly moral? Identity intuitions across domains of sociomoral reasoning and age. *Journal of Experimental Child Psychology, 192*, Article 104769. https://doi.org/10.1016/j.jecp.2019.104769

Li, J., & Tomasello, M. (2018). The development of intention-based sociomoral judgment and distribution behavior from a third-party stance. *Journal of Experimental Child Psychology, 167*, 78–92. https://doi.org/10.1016/j.jecp.2017.09.021

Lucca, K., Yuen, F., Wang, Y., Alessandroni, N., Allison, O., Alvarez, M., … Hamlin, J. K. (2025). Infants' social evaluation of helpers and hinderers: a Large-Scale, Multi-Lab, coordinated replication study. *Developmental Science, 28*, Article e13581. https://doi.org/10.1111/desc.13581

Malle, B. F., & Bennett, R. E. (2002). People's praise and blame for intentions and actions: implications for the folk concept of intentionality. *Technical Reports of the Institute of Cognitive and Decision Sciences*. Eugene, OR: University of Oregon.

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*, 254–262. https://doi.org/10.1037/dev0000424

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

Mammen, M., Köymen, B., & Tomasello, M. (2021). Young children's moral judgments depend on the social relationship between agents. *Cognitive Development, 57*, Article 100973. https://doi.org/10.1016/j.cogdev.2020.100973

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin, 40*, 203–216. https://doi.org/10.1177/0146167213508791

Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology, 104*, 382–397. https://doi.org/10.1016/j.jecp.2009.08.001

Nummer-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development, 59*, 1323–1338. https://doi.org/10.2307/1130495

Oostenbroek, J., & Vaish, A. (2019). The emergence of forgiveness in young children. *Child Development, 90*, 1969–1986. https://doi.org/10.1111/cdev.13069

Perner, J., Priewasser, B., & Roessler, J. (2018). The practical other: teleology and its development. *Interdisciplinary Science Reviews, 43*, 99–114. https://doi.org/10.1080/03080188.2018.1453246

Piaget, J. (1932). *The Moral Judgement of the Child*. London: Routledge & Kegan Paul.

Proft, M., & Rakoczy, H. (2019). The ontogeny of intent-based normative judgements. *Developmental Science, 22*, Article e12728. https://doi.org/10.1111/desc.12728

Ruble, D. N., Newman, L. S., Rholes, W. S., & Altshuler, J. (1988). Children's "naïve psychology": the use of behavioral and situational information for the prediction of behavior. *Cognitive Development, 3*, 89–112. https://doi.org/10.1016/0885-2014(88)90032-9

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*, 550–560. https://doi.org/10.2307/1129966

Rholes, W. S., & Ruble, D. N. (1986). Children's impressions of other persons: the effects of temporal separation of behavioral information. *Child Development*, 872–878. https://doi.org/10.2307/1130364

Roessler, J., & Perner, J. (2013). Teleology: belief as perspective. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*. Oxford: Oxford University Press.

Ross, L. D. (1977). The intuitive psychologist and his shortcomings: distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*. New York, NY: Academic Press.

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General, 116*, 223–237. https://doi.org/10.1037/0096-3445.116.3.223

Samland, J., Josephs, M., Waldmann, M. R., & Rakoczy, H. (2016). The role of prescriptive norms and knowledge in children's and adults' causal selection. *Journal of Experimental Psychology: General, 145*, 125–130. https://doi.org/10.1037/xge0000138

Stanovich, K. (2011). *Rationality and the Reflective Mind*. New York, NY: Oxford University Press.

Strohminger, N., Knobe, J., & Newman, G. (2017). The true self: a psychological concept distinct from the self. *Perspectives on Psychological Science, 12*, 551–560. https://doi.org/10.1177/1745691616689495

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, 4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

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

Umscheid, V. A., Smith, C. E., Warneken, F., Gelman, S. A., & Wellman, H. M. (2023). What makes voldemort tick? Children's and adults' reasoning about the nature of villains. *Cognition, 233*, Article 105357. https://doi.org/10.1016/j.cognition.2022.105357

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development, 81*, 1661–1669. https://doi.org/10.1111/j.1467-8624.2010.01500.x

Vaish, A., Carpenter, M., & Tomasello, M. (2011). Young children's responses to guilt displays. *Developmental Psychology, 47*, 1248–1262. https://doi.org/10.1037/a0024462

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

Waddington, O., Jensen, K., & Köymen, B. (2022). Boundaries of apologies: children avoid transgressors who give the same apology for a repeat offence. *Cognitive Development, 64*, Article 101264. https://doi.org/10.1016/j.cogdev.2022.101264

Waddington, O., Proft, M., Jensen, K., & Köymen, B. (2023). Five-year-old children value reasons in apologies for belief-based accidents. *Child Development, 94*, 143–153. https://doi.org/10.1111/cdev.13893

Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *British Journal of Developmental Psychology, 2*, 73–81. https://doi.org/10.1111/j.2044-835X.1984.tb00536.x

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology, 24*, 358–365. https://doi.org/10.1037/0012-1649.24.3.358

Yuill, N., Perner, J., Pearson, A., Peerbhoy, D., & Van den Ende, J. (1996). Children's changing understanding of wicked desires: from objective to subjective and moral. *British Journal of Developmental Psychology, 14*, 457–475. https://doi.org/10.1111/j.2044-835X.1996.tb00718.x