

2021-10-13

Young children's understanding of justifications for breaking a promise

Kanngiesser, P

<http://hdl.handle.net/10026.1/18180>

10.1016/j.cogdev.2021.101127

Cognitive Development

Elsevier

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This is a preprint copy and may diverge slightly from the published version.

Please cite as: Kanngiesser, P., Mammen, M., & Tomasello, M. (2021). Young children's understanding of justifications for breaking a promise. *Cognitive Development*, 60, 101127.

<https://doi.org/10.1016/j.cogdev.2021.101127>

Young children's understanding of justifications for breaking a promise

Patricia Kanngiesser^{1*}, Maria Mammen^{2*}, Michael Tomasello^{3,4}

¹School of Psychology, University of Plymouth, Plymouth, UK

²Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

³Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

⁴Department of Psychology and Neuroscience, Duke University, Durham, US

*Both authors contributed equally to this manuscript

Corresponding authors: Patricia Kanngiesser (patricia.kanngiesser@plymouth.ac.uk), Maria Mammen (maria.mammen@lmu.de)

Author note: We would like to thank Gesa Volland, Susanne Mauritz, Benjamin Brückner, Jana Mahler, Undine Bork, Regina Sticker, Amelie Conrad und Kathrin Kern for their help with data collection and data coding; Cristina Zickert for the drawings and all nurseries and children for their friendly cooperation. We thank Jan K. Woike for pointing us towards Cicero's writings.

Abstract

There are sometimes legitimate reasons for breaking a promise when circumstances change. We investigated 3- and 5-year-old German children's understanding of promise breaking in prosocial (helping someone else) and selfish (playing with someone else) conditions. In Study 1 ($n = 80$, 50% girls), preschoolers initially kept their own promise in all conditions. When they eventually broke their promise, 3-year-olds' justifications mostly referenced salient events, whereas 5-year-olds also referenced social norms. In Study 2 ($n = 65$, 49% girls), 5-year-olds preferred others' promise-breaking more in prosocial than selfish conditions; 3-year-olds showed the reverse pattern. Three-year-olds' justifications focused on desires, whereas 5-year-olds focused on relevant events. Overall, 3-year-olds were able to offer justifications, but 5-year-olds started to distinguish what counted in the eyes of others as "good" and "bad" reasons for promise breaking.

Keywords: moral development, promises, reasons

Young children's understanding of justifications for breaking a promise

When we make a promise to someone, we now have an obligation to do what we have promised (Searle, 1969). At the same time, promises by others give us reason to expect that they will fulfil their commitment (Scanlon, 1990; Woods, 2016). Empirical evidence has consistently shown that adults usually keep their word even when it is personally costly for them (Bicchieri, 2016; Charness & Dufwenberg, 2006; Ostrom et al., 1992; Sally, 1995; van den Assem et al., 2011; Vanberg, 2008; Woike & Kanngiesser, 2019). To date, most experimental and theoretical work on promises has focused on how and why promises are obligatory. In the last decades, philosophers have argued, for example, that promises are binding because of a social contract (Rawls, 1955), because it is wrong to violate other's trust (Scanlon, 1990) or because of a joint commitment between the promisor and the promisee (Gilbert, 2011). While it appears clear-cut that keeping a promise is usually obligatory, the question of whether it can be permissible to break a promise intentionally has received considerably less attention.

Promises concern future actions of speakers and, hence, they are particularly vulnerable to changes in circumstances. Cicero argued that, in some situations, one even has a duty to break one's promise (Cicero & Miller, 1913). For example, breaking a promise to appear in court because one's child has suddenly fallen gravely ill or breaking a promise to return a sword to a friend who has a "fit of insanity" (Cicero & Miller, 1913, I.X., p. 33; III.XXV, p. 371). There are likely nuances in what counts as a legitimate change of circumstances to warrant promise breaking. For instance, you may accept that I break my promise to go swimming with you because a close family member has a health emergency and requires immediate help. However, you may be less inclined to accept my promise breaking if I stayed home to watch TV or helped another friend to do something that could easily be done at another time. These differences in permissibility may be due to an understanding of how different moral values are ranked in a community's cultural or moral common ground

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

(Tomasello, 2019). By extension, one would find it easier to provide credible and accepted justifications for promise breaking under some circumstances than others.

Developmental research has shown that young children already understand that commitments have implications for their own behaviour. From three years of age, children will keep their promise despite being tempted to play a fun game (Kanngiesser et al., 2017) and from five years of age, children will cheat less or tell the truth more frequently after promising to be honest (Heyman et al., 2015; Kanngiesser, Sunderarajan, & Woike, 2021; Lyon & Dorado, 2008; Quas, Stolzenberg, & Lyon, 2018; Talwar, Lee, Bala, & Lindsay, 2002). Moreover, starting at age three, children expect their partners to keep their commitments and, when those partners fail to honour their obligation, young children will protest (Kachel et al., 2017; Kanngiesser et al., 2017) or try to re-engage them (Gräfenhain et al., 2009).

More sophisticated judgments about promises emerge later in childhood (Keller et al., 1998; Keller & Edelstein, 1985; Hussar & Horvath, 2013). Five-year-olds, but not younger children, trust someone who has previously kept a promise more than someone who has previously broken a promise (Isella, Kanngiesser, & Tomasello, 2019; Rotenberg, 1980; but see, Kidd, Palmeri, & Aslin, 2013, for tentative evidence in younger children). In addition, 6-year-olds correctly infer that the failure to perform a promised action makes the promisee unhappy (Bernicot & Laval, 1996). Research has further shown that 6- and 7-year-olds distinguish between commitments that were broken intentionally and those that were broken unintentionally (Mant & Perner, 1988) and assign blame differently depending on speakers' intentions (Maas & Abbeduto, 2001). In summary, previous research has demonstrated that young children already understand the obligatory nature of promises¹ and that older children

¹Promises, as understood here, possess normative force, and entail an obligation to perform the promised act. When we write about "commitments" we hence refer to commitments in the normative sense. This is distinct

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

reason differently about intentionally and unintentionally broken promises. It is an open question, however, whether young children understand that there exist “good” and “bad” reasons for intentionally breaking a promise.

In other words, do children understand that, under some circumstances, promissory obligations can be superseded by other obligations? A good candidate for such an overriding obligation is helping behaviour. From the second year of life, infants help others in a variety of situations (Warneken, 2015). Three- to 5-year-olds prefer helpers over hinderers and evaluate them as nicer (Van de Vondervoort, & Hamlin, 2017). They also judge helping as obligatory across a range of situations, including “helping” someone to take another’s property (Dahl, Gross, & Siefert, 2020). This suggests that young children view helping as a moral obligation that even overrides other norms (e.g., norms not to steal). It is therefore possible that young children would consider it legitimate to break a promise for prosocial reasons (i.e., to help someone) and, conversely, consider it illegitimate to break a promise for selfish reasons (e.g., to play with someone). Alternatively, young children may view promises as categorical obligations that one should not break under any circumstances.

To address this question, we conducted two studies with 3- and 5-year-olds. This allowed us to investigate whether there would be age-related changes in children’s understanding of intentional promise-breaking and their justifications. Previous research has shown that 3-year-olds show a nascent understanding of commitments and promises in interactive first-party situations (Gräfenhain et al., 2009; Kachel et al., 2017; Kanngiesser et al., 2017). However, studies using third-party tasks have found a later emergence of promissory understanding such as selective trust in promise-keepers (Isella, Kanngiesser, & Tomasello, 2019) or the difference between intentionally and accidentally broken promises

from a “minimal” commitment approach that does not include a normative notion (Michael, Sebanz, & Knoblich, 2016).

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

(Maas & Abbeduto, 2001; Mant & Perner, 1988). Moreover, research on the development of moral reasoning has found that with age, children consider more, sometimes even competing, obligations in their judgments (Killen & Stangor, 2001; Smetana & Ball, 2017; Theimer et al., 2001) and provide more relevant and elaborate reasons (Köymen et al., 2016; Mammen et al., 2018; 2021).

In Study 1, we focused on children's *own* promise-breaking by investigating a) whether they would break a promise under prosocial or selfish conditions (between-subjects) and b) which justifications they would offer for their behaviour. In Study 2, we investigated children's preferences for *third parties* by presenting them with stories, in which one character promised to another and then had to choose whether to break this promise under prosocial or selfish conditions (between-subjects). We recorded a) whether children preferred to end the story with a broken promise and b) how they justified choosing the promise-breaking ending. We asked children to select an ending they preferred (and not what one should do) to minimize the explicit moral framing of the situation. This ensured comparability to Study 1 where promise breaking was not explicitly morally framed.

If children distinguish between legitimate and illegitimate reasons for promise-breaking, they should break their own promises (Study 1) and choose the promise-breaking ending (Study 2) more often in the prosocial than in the selfish condition. Children's justifications should track these patterns and we should observe more suitable justifications (e.g., references to norms) in the prosocial as compared to the selfish condition. It is possible that 3-year-olds would show an understanding of intentional promise breaking in first party tasks (Study 1), but not in third party tasks (Study 2) and that 5-year-olds would generally provide more suitable justifications for their own and other's actions than 3-year-olds.

Study 1

In this study, we investigated 3- and 5-year-olds' promise-breaking behaviour and their justifications for their own behaviour. The study set-up was similar to a previous study

(Kanngiesser et al., 2017), in which an adult experimenter (E1) elicited a promise from a child to continue a tedious task—cleaning up paper shreds—in E1's absence. Another adult experimenter (E2) then tried to lure children away from the cleaning task with a marble run. Kanngiesser et al. (2017) found that preschoolers resisted temptation and continued to clean for longer after giving a promise as compared to a no-promise control (Study 2) or a cleaning reminder (Study 3). In the current study, we contrasted two different promise conditions (between-subjects) and refrained from including an additional control as the previous study had already validated the promise task. In the selfish condition, E2 tried to lure children away by asking them to play with the marble run (similar to Kanngiesser et al., 2017). In the prosocial condition, E2 accidentally broke the marble run and asked children for help to rebuild it. All children were eventually induced to break their promise and then asked by E1 to justify their behaviour.

Methods

Participants. Forty 3-year-olds ($M = 3;10$, $Range = 3;6 - 3;12$; 20 girls) and 40 5-year-olds ($M = 5;9$, $Range = 5;6 - 5;12$; 20 girls) took part in the study. Half of the children per age group participated in the prosocial and the selfish condition, respectively. The sample size (20 children per cell) was chosen based on previous studies (Kanngiesser et al., 2017). Ten additional 3-year-olds and four additional 5-year-olds began the study but were excluded because they refused to participate (five 3-year-olds, three 5-year-olds), never broke their promise (three 3-year-olds), accidentally collapsed the marble track during familiarization (one 3-year-old, one 5-year-old), or did not understand the procedure (one 3-year-old). We recruited children through a database of parents who had given informed consent for their child to participate in studies on socio-cognitive development. They lived in a city in [removed for review] and came mostly from middle class families. We tested children individually in a quiet room in their kindergarten.

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

The study was conducted in accordance with the Declaration of Helsinki and the ethical guidelines of the German Psychological Society (DGPs) and the Association of German Professional Psychologists (BDP). The study did not involve any invasive techniques, ethically problematic procedures, or deception, and, therefore, did not require approval by an Institutional Review Board (see the regulations on freedom of research in the German Constitution, §5(3)).

Procedure. We used a procedure and set-up similar to a previous published study on children's commitment to promises (Kanngiesser et al., 2017). The study consisted of four parts: (1) a warm-up phase, (2) a promising phase, (3) a distraction phase, and (4) a reasons phase. Two female experimenters (E1, E2) conducted the study. For a detailed study script, see Supplementary Material.

Warm-up phase. As a warm-up, E1 and E2 played a puzzle game with the child. Once the child was warmed-up, E2 left the room on a pretend errand.

Promising phase. Next, E1 demonstrated a marble track to the child that was constructed out of individual building blocks (see Supplementary Figure S1). While the child played with the marble track, E1 accidentally spilt a bucket of shredded paper on a carpet located near the marble track (see Supplementary Figure S2). Next, E1 alerted the child to the accident and asked for the child's help to clean up the shredded paper. If the child hesitated, E1 repeated her request for help. After a short while, E2 returned and asked E1 to leave on a pretend errand (i.e., to talk to the nursery head teacher). Importantly, before E1 left the room, she elicited a promise from the child to continue cleaning (*Versprichst du mir, dass du weiter aufräumst?* 'Do you promise to keep cleaning?'). If necessary, she repeated the question to ensure that all children promised (by agreeing verbally or non-verbally).

Distraction phase. In this phase, E2 tried to induce promise-breaking by luring children away from their cleaning task. This was necessary as we wanted to elicit children's justifications for breaking their promise in the final phase. While E1 left the room, E2 sat

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

behind the marble track (facing the carpet and the child) and tried to distract the child from the cleaning task using a sequence of prompts. In the prosocial condition, E2 tried to prompt the child to help her rebuild the marble track that she had accidentally collapsed. In the selfish condition, E2 tried to prompt the child to play with the marble track. The prompts were taken from Kanngiesser et al. (2017) and slightly adjusted to ensure a close match between the prosocial and the selfish condition (see Table 1). E2 timed the prompts using a mobile phone timer (the timing was identical to Kanngiesser et al., 2017). If children did not leave the carpet within 60 seconds, E2 continued to ask the child for help to rebuild the marble track (prosocial condition) or to join her to play (selfish condition) until the child complied. All children, except for three 3-year-olds (not included in the data analyses), eventually broke their promise after repeated prompts from E2. Next, E2 and the child either rebuilt the marble track (prosocial condition) or played with the marble track (selfish condition) for about 1 minute.

Justification phase. After 1 minute had passed, E2 gave E1 a covert signal (i.e., coughing) to return. After E1 had returned, she used a series of prompt questions to elicit excuses from the child:

- (1) E1 expressed surprise that the child had stopped cleaning ('You are not cleaning?').
- (2) E1 mentioned the promise ('But you promised it!').
- (3) E1 repeated that the child was not cleaning ('You are not cleaning. Why?')
- (4) E1 asked what had happened ('What happened?').

After the final prompt, to resolve the situation amicably, E1 stated that the cleaning was almost done and that the child could play with the marble track, while E1 finished cleaning the carpet.

Data coding. We videotaped the study. A coder blind to the hypotheses of the study coded children's behaviour from videotape. The behaviours and responses of 25% of children

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

(equally distributed by age, gender, and condition) were scored by additional coders for reliability purposes.

Commitment. We assessed children's promise-keeping by calculating a commitment score for each child (see Kanngiesser et al., 2017). The commitment score was derived from the number of prompts that were needed until a child broke their promise. We scored promise-breaking if the child started building (prosocial condition) or playing with (selfish condition) the marble track or if the child left the carpet for more than 15s. The commitment score ranged from 0 to 5, with higher scores indicating greater commitment. Specifically, a score of 0 indicated that the child broke their promise immediately after E1 had left the room, and a score of 5 indicated that children broke their promise after the final prompt (see Table 1). It should be noted that all children were eventually made to break their promise. Agreement between two independent coders on 25% of the data was very good (equal-weighted $\kappa = .97$).

Returning-to-clean. While coding the video tapes, we noticed that some children returned to the carpet to continue cleaning when E1 re-entered the room. We thought this noteworthy behaviour that may indicate an attempt to make amends, possibly because children felt guilty for having broken their promise (Barrett, Zahn-Waxler, & Cole, 2008; Vaish et al., 2016; Vaish, 2018). As an exploratory approach, we coded whether children returned to the carpet to clean after having helped to rebuild (prosocial condition) or played with (selfish condition) the marble track. Agreement between two independent coders on 25% of the data was perfect ($\kappa = 1$).

Justifications. We transcribed and coded children's justifications for breaking their promise. We excluded the last question ('What happened?') from our analyses as this question did not directly elicit a justification but rather asked children to recount what had happened. We used a deductive-inductive approach to develop a coding scheme. In a first step, the authors read through children's utterances to identify the most frequently mentioned

YOUNG CHILDREN’S UNDERSTANDING OF PROMISE BREAKING

themes. This process was guided by theoretical considerations—for example, to get a sense of whether children viewed helping/playing as an obligation or whether they simply wanted to do it, we coded for normative statements (“Because I had to help.”) and desires (“Because I wanted to play.”), respectively. A detailed coding scheme was then developed, consisting of the following five categories: *accident*, *desire*, *normative*, *request* and *task* (for details, see Table 2). We scored each child’s answers to questions 1–3 together, that is, each child only received a binary code (yes/no) for each category (Note that responses could receive a score in more than one category). This allowed us to control for differences in verbal fluency across children. Responses that did not fall into any of these categories (including children who did not give any response) were coded as *irrelevant*.

A coder blind to condition and to the aim of the study, coded children’s responses. The coder was provided with the coding manual and in case they had any clarification questions consulted with the study authors. For reliability purposes, a second coder was first trained on a subset of children and then independently scored 25% of the data. There was very good to perfect agreement between the two coders for the six coding categories ($\kappa = 0.88$ – 1.00).

Table 1: Overview of the sequence of prompts used to distract the child in the two conditions in Study 1.

score	sequence	prosocial condition	selfish condition
0	0 seconds	E2 accidentally collapses part of the marble track, saying ‘Oh’.	E2 puts two marbles on the marble track, saying ‘Oh’.
1	15 seconds	E2 says ‘Oh no’. (not looking at the child)	E2 says ‘Oh great’. (not looking at the child)
2	30 seconds	E2 accidentally collapses further parts of the track, saying ‘Oh no, the marble track’. (<i>not looking at the child</i>)	E2 puts two marbles on the track, saying ‘Oh great, the marble track’. (<i>not looking at the child</i>)
3	45 seconds	E2 briefly makes some noises with blocks of the marble track, looks at the child and says, ‘The marble track is broken’.	E2 puts two marbles on the track, looks at the child and says, ‘The marble track is nice’.
4	60 seconds	E2 looks at the child and says, ‘Will you help?’.	E2 looks at the child and says, ‘Will you play?’.
5	>60 seconds	E2 repeats ‘Oh, come and help for a bit’.	E2 repeats ‘Oh, come and play for a bit’.

Table 2: Overview of the coding scheme for children’s excuses in Study 1.

Category	Description	Examples
Accident	Children refer to the accident or to the broken marble track.	(a) <i>Weil die Marmelbahn umgefallen ist.</i> ‘Because the marble track fell over.’
Desire	Children refer to their own desires, wishes, or intentions.	(b) <i>Das ist kaputt gegangen.</i> ‘This broke’ (a) <i>Weil ich das hier aufbauen wollte.</i> ‘Because I wanted to build this here.’ (b) <i>Ich hatte keine Lust, ich wollte nur mit das spielen!</i> ‘I wasn’t up for it, I only wanted to play with it.’
Normative	Children refer to an obligation using normative language, i.e. the verbs <i>müssen</i> ‘should’ or <i>sollen</i> ‘must’.	(a) <i>Na, weil ich mithelfen sollte.</i> ‘Because I had to help’ (b) <i>Weil ich damit spielen musste.</i> ‘Because I had to play with it.’
Request	Children refer to E2’s request to help or to play.	(a) <i>Weil sie gesagt hat, ich solle mal kurz mitspielen.</i> ‘Because she said that I should come and play for a bit.’ (b) <i>Die hat gefragt, ob ich mithelfen kann.</i> ‘She asked whether I can come and help.’
Task	Children refer to the cleaning task; to having done some of it or to having finished it.	(a) <i>Ich hab ja schon ganz viel gemacht [...]</i> ‘I have already done so much.’ (b) <i>Weil das langweilig ist.</i> ‘Because it is boring.’

Data analyses. We analysed children’s commitment scores using ordinal linear models in R (Christensen, 2018; R Core Team, 2018). We entered commitment score as response variable (ordered factor from 0 to 5), and age group (3- vs. 5-year-olds), condition (prosocial vs. selfish), and their two-way interaction as predictors. We entered gender (female vs. male) as a control predictor. We conducted model comparisons using likelihood ratio tests. Specifically, we compared (i) a full model (all predictors) with a null model (including only the control predictor gender), and (ii) a full model and a reduced model (excluding the two-way interaction of age group and condition). We determined p-values for individual predictors (or their interaction) in the final model using the `drop1` function in R (likelihood ratio tests). In addition, we analysed whether children returned to the carpet to clean using Fisher exact tests using the `exact2x2`-package in R (Fay, 2010). We further used Fisher exact tests to determine whether there was a difference between conditions (a) in the number of children producing relevant justifications and (b) in the number of children referring to each of the five

different relevant justifications. For de-identified data and complete documentation of R code and results, see: <https://osf.io/rc3tp/>.

Results

Commitment. Children were highly committed to the task in the two conditions (see Figure 1). Comparison of the full model with the predictors condition, age, and their two-way-interaction (and gender as a control predictor) to the null model with only gender, revealed no significant improvement in model fit for the full model, $\chi^2(3) = 4.84, p = 0.184$. Accordingly, none of the predictors in the full model was significant: condition X age interaction, $Est = 0.47, 95\% CI [-1.51, 2.43], \chi^2(1) = 0.22, p = 0.640$, and gender, $Est = -0.40, 95\% CI [-1.39, 0.57], \chi^2(1) = 0.66, p = 0.418$ (likelihood ratio tests; see Supplementary Information, Table S1).



Figure 1. Three- and 5-year-olds’ commitment scores in the prosocial and the selfish condition in Study 1. Higher commitment scores indicate that children continued the cleaning task for longer. The size of the bubbles indicates the number of children who were assigned the respective commitment score (range = 0–16). The median is indicated by the solid lines, and the second and third quartiles are indicated by the boxes.

Returning-to-clean. As an exploratory analysis, we analysed whether children returned to continue clean. We found no significant difference between conditions in the number of 3-year-olds who returned to clean (prosocial: $n = 2$, 10%; selfish: $n = 8$, 40%), $p = 0.065$, $OR = 5.74$, 95% $CI [1.00, 43.48]$ (Fisher exact test). Similarly, there was no significant difference between conditions for 5-year-olds (prosocial: $n = 7$, 35%; selfish: $n = 13$, 65%), $p = 0.113$, $OR = 3.34$, 95% $CI [0.82, 13.80]$.

Justifications. More 3-year-olds provided relevant justifications in the prosocial condition ($n = 15$ of 20, 75%) as compared to the selfish condition ($n = 8$ of 20, 40%), $p = 0.054$, $OR = 0.23$, 95% $CI [0.06, 1.00]$ (Fisher exact test). In contrast, most 5-year-olds provided relevant justifications in both conditions (prosocial: $n = 18$ of 20, 90%; selfish: $n = 15$ of 20, 75%), $p = 0.408$, $OR = 0.34$, 95% $CI [0.04, 2.07]$. Figure 2 provides a detailed overview of the frequencies of the different types of relevant justifications. In the prosocial condition, most 3-year-olds (75%) and about half of the 5-year-olds (45%) referred to the accident, but they never did so in the selfish condition—this difference was significant both for 3-year-olds, $p < 0.001$, $OR = 0$, 95% $CI [0.00, 0.09]$ and for 5-year-olds, $p = 0.001$, $OR = \infty$, 95% $CI [0.00, 0.30]$ (Fisher exact tests). Moreover, more 5-year-olds referred to norms in the prosocial condition (45%) than in the selfish condition (10%), $p = 0.031$, $OR = 0.14$, 95% $CI [0.02, 0.78]$. Three-year-olds showed no difference in their references to norms between the prosocial (35%) and the selfish condition (20%), $p = 0.480$, $OR = 0.47$, 95% $CI [0.11, 2.16]$. None of the other categories differed significantly between conditions for 3-year-olds, $p_s > 0.10$, and for 5-year-olds, $p_s > 0.60$, with 95% CIs for odds ratios always overlapping 1 (see Supplement, Rcode and output, for details).

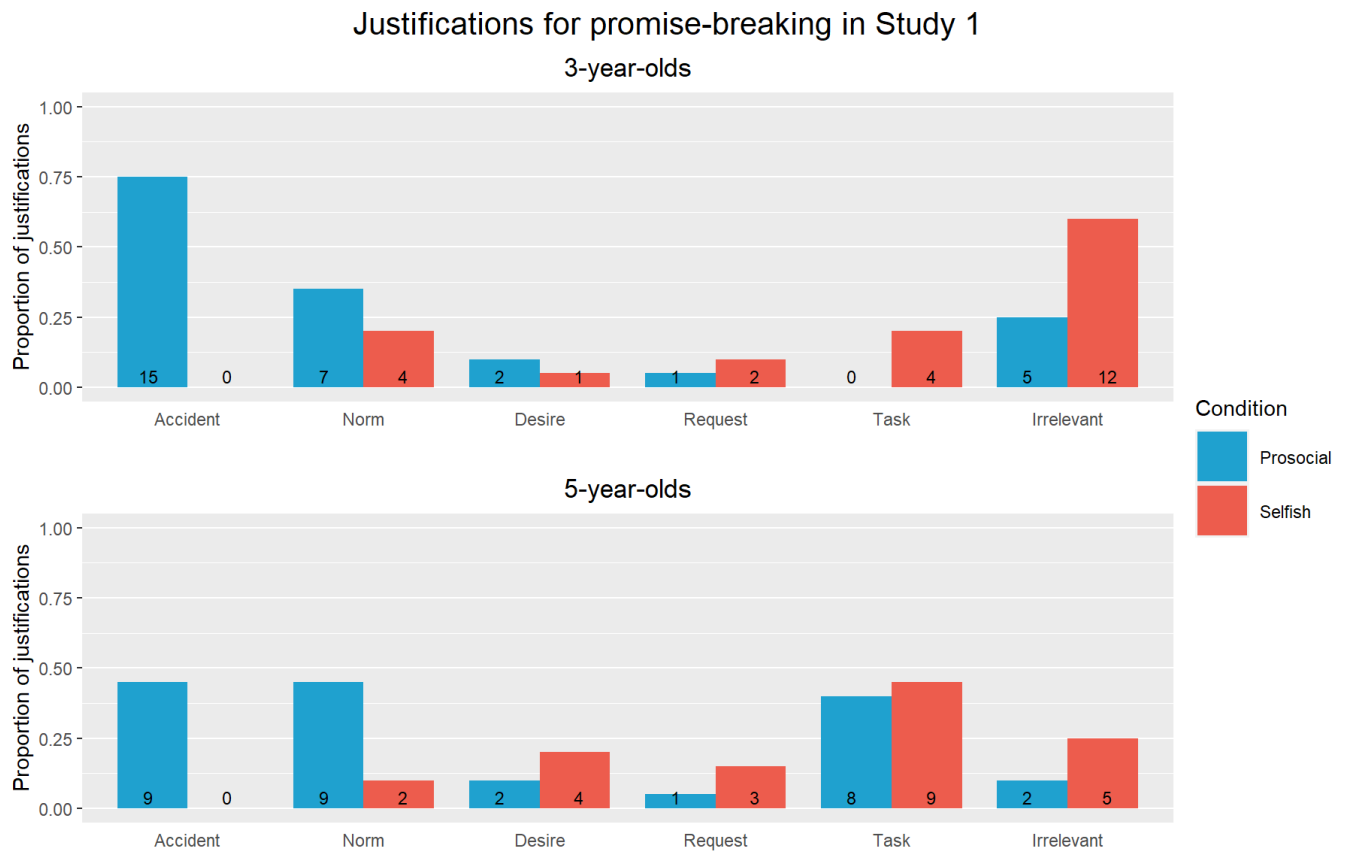


Figure 2. Three- and 5-year-olds’ justifications in the prosocial and the selfish condition in Study 1. The bars indicate different relevant responses (accident, norm, desire, request, task) and irrelevant responses (on the right). For relevant responses, we scored each category only once per child (though it should be noted that responses could be scored in multiple categories). The number of scores per category are printed at the bottom of each bar.

Discussion

We found that both 3- and 5-year-olds were highly committed in both conditions. This extends a previous study that had found high levels of commitment to promises in preschool children with a set-up comparable to the selfish condition in the current study (Kanngiesser et al., 2017, Studies 2 and 3). It shows that children not only resist temptations to break their promises in selfish circumstances (to play a fun game with someone), but also in prosocial circumstances (to help someone). Children’s strong commitment likely masked any differences between conditions that we had expected to find in this study. It should be noted that if children are simply tempted away from the cleaning task without an explicit commitment (see Kanngiesser et al., 2017, Studies 2 and 3), they abandoned the task more

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

quickly. This indicates that explicit commitments and not simply joint activities with an adult led to children's reluctance to abandon their cleaning task. In an exploratory analysis, we found no significant difference between conditions in the number of children who spontaneously resumed their cleaning when E1 returned from their errand. Further studies are needed to investigate what spontaneous behavioural (e.g., reparative behaviours, gaze aversion) and emotional reactions (e.g., shame, guilt) children would show after having broken a promise (Barrett, Zahn-Waxler, & Cole, 2008; Mascolo & Fischer, 2007; Vaish, 2018).

Three- and 5-year-olds provided relevant justifications in both conditions, but fewer 3-year-olds did so in the selfish condition (note though that the condition effect only approached significance in this age group). We further found that children in both age groups frequently referred to the accident in the prosocial condition. Yet, only 5-year-olds mentioned norms significantly more often in the prosocial than in the selfish condition. Taken together, this suggests that 3-year-olds were already able to provide relevant justifications for their own promise-breaking; particularly, when it involved salient events such as the accidental collapse of a marble track. Five-year-olds displayed a more sophisticated understanding by referencing norms under the appropriate circumstances (i.e., when justifying helping someone; "I had to help."), but rarely when it would have been inappropriate (i.e., when justifying playing). These findings are in line with previous research showing that children's reasoning becomes more sophisticated throughout the preschool years (Köymen et al., 2016; Mammen et al., 2018) and that with increasing age, children are better able to judge which pieces of information are sufficient and informative for an interlocutor (e.g., Bartsch et al., 2007; Perner & Leekam, 1986; Saylor et al., 2006).

Both conditions involved an attractive marble run and a series of closely matched distraction cues, with the crucial difference being whether the adult experimenter needed help fixing the marble track or invited the child to play. It is possible that differences between

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

conditions were too subtle to elicit variation in children's readiness to abandon the cleaning task. For example, children may have perceived helping to fix the marble track as an intermediate step to later playing with it. It should be noted though that children's justifications differed between conditions, indicating that they noticed relevant features of the situation (e.g., breaking of the marble track). Future studies could increase the contrast between conditions and, for example, increase the need of the person who requires help.

Study 1 focused on children's behaviour in a first party situation: children clearly felt bound by their promises and were able to provide some justifications. However, children may have felt obligated to keep their promise because the promisee could have returned any time and they may have been worried about negative consequences of their promise breaking. Moreover, justifications for one's own behaviour may differ from how one reasons about what others should or should not do (Eisenberg-Berg & Neal, 1981). We therefore conducted a second study in which we asked children about their preferences and justifications for third parties.

Study 2

In this study, we investigated 3- and 5-year-olds' preferences for third parties' promise-breaking. We closely modelled the scenarios in this study after the first party situations in Study 1. We told children short, illustrated stories in which one character (C1) made a promise to another character (C2) to continue a boring task (e.g., cleaning up leaves) in their absence. A third character (C3) then tried to lure the promiser (C1) away from their task either by asking for help (prosocial condition) or by suggesting to play together (selfish condition; between-subjects). Children then selected one of two story endings by indicating which one they liked more: in one ending, C1 kept their promise and in the other, C1 broke their promise. To minimize the moral framing of the choice question, we asked children to select an ending they preferred. This ensured that the stories were comparable to Study 1 where promise breaking was not explicitly morally framed (i.e., children in Study 1 were eventually

made to break their promise, but they were not asked whether they should or should not do it). To investigate what motivated children's decisions, we also asked them to provide justifications for their choices.

Methods

Participants. Thirty-three 3-year-olds ($M = 3;9$, $Range = 3;6-3;12$, 16 girls) and thirty-two 5-year-olds ($M = 5;9$, $Range = 5;6-5;12$, 16 girls) took part in the study. Half of the children per age group participated in the prosocial and the selfish condition, respectively. The sample size was chosen based on previous studies (e.g., Mammen et al., 2018). Two additional 3-year-olds and two additional 5-year-olds began the study but were excluded because of experimenter error (two 5-year-olds, one 3-year-old), or because they did not talk to the experimenter (one 3-year-old). We recruited children in the same manner as in Study 1. We tested children individually in a quiet room in their kindergarten.

Procedure. The study was divided in two phases: a warm-up phase and a story-telling phase. Two experimenters conducted the study: E1 operated a lifelike puppet and E2 operated a frog puppet (see Supplementary Figures S6 and S7). For a detailed script, see Supplementary Materials.

Warm-up phase. E1 and E2 played a puzzle or ball game with the child to familiarise the child with the two puppet characters. Next, E1 conducted a reasoning warm-up task similar to Mammen et al. (2018). In this task, E1 and the child deliberated on how to organize a birthday party for a toy cat (see Supplementary Figure S4). E1 elicited reasons from the child ('What do you think?', 'Why?') and provided reasons herself throughout the task (e.g., 'Yes, I think Mimi [the cat's name] should get the wool because cats like to play with this.'). The stories in the warm-up had no relation to the stories in the main phase of the study.

Story telling phase. In the main phase of the study, we presented children with two different picture-stories (order counterbalanced across children). Both stories depicted the following situation (see Supplementary Materials, Figure S5): (i) the protagonist (C1) and a

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

second characters (C2) engaged in an activity (e.g., cleaning up), (ii) C2 needed to leave and asked the protagonist to promise to continue the activity, and (iii) a third character (C3) appeared and either asked the protagonist for help (prosocial condition) or to play (selfish condition). For each story, E1 presented two endings (order counterbalanced across children): (1) the protagonist kept their promise and continued the activity (same in both conditions) and (2) the protagonist broke their promise and helped (prosocial condition) or played (selfish condition) with C3. We matched the gender of the characters to the gender of the participating child.

We presented children with the stories as follows (for the script, see Electronic Supplement). In the first part, E1 suggested to create a story for the frog (operated by E2, waiting outside) and presented children with the first three pictures of story, reading the accompanying text. E1 then showed the child the two story-endings and asked them to select one ending by asking “What do you think, which ending do you like more? You can only choose one ending.”. After children had decided on an ending, E1 continued with a series of questions about the chosen ending: ‘What does [the character] do here?’ (action comprehension check), ‘Why does s/he [do activity]?’ (justification), and ‘Why do you like this ending?’ (justification).

In a second part, E2 (operating a frog puppet) re-entered the room and asked children to re-tell the story. To check whether children remembered the promise, E2 asked ‘What are they doing?’ (promise comprehension check) when they reached the second picture of the story (C1 promising to C2). Regardless of whether the children gave the correct answer or not, E2 always stated ‘Ah, they are promising something to each other’. Finally, E2 asked a series of questions about the final picture (i.e., the ending selected by the child): ‘What does the character do here?’ (action comprehension check) and ‘Why does the character [do activity]?’ (justification).

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

Data coding. We videotaped the study and coded from videotape which story-ending children chose. We calculated a promise-breaking score by summing their choice of the promise-breaking ending across the two stories (range 0-2).

We also transcribed children's answers to E1's and E2's questions verbatim. There were six questions in total: two action comprehension checks, one promise comprehension check, and three justification questions. Children's responses were coded by a coder unfamiliar with the hypotheses of the study according to the following coding scheme. We scored children's replies to the comprehension questions as either correct or incorrect (for details, see Table 3). We scored children's answers to the justifications questions by distinguishing between *relevant* justifications and *irrelevant* justifications. We scored justifications as *relevant* if they fell into one or more of the following categories (utterances could be scored in more than one category): *desire*, *normative*, *request*, *evaluation*, and *environment* (for details, see Table 4). We coded as irrelevant justifications responses that did not fall into any of the categories, including no responses, "don't know"-answers, or off-topic replies (e.g., "Because we are making a book for the frog"). The coding categories were developed using an inductive-deductive approach with the aim to mirror the coding categories of Study 1 as much as possible. However, the study-set up and scenarios differed to an extent that made some codes from Study 1 inapplicable and required new codes. For example, no salient accident occurred in Study 2, instead children referred to states in the environment (e.g., "The ball is on the hedge."; codes as "environment"). Moreover, children frequently used evaluations (e.g., "I like it best.") to justify why they chose a particular ending.

A coder blind to the aim of the study, coded children's responses. The coder was provided with the coding manual and in case they had any clarification questions consulted with the study authors. For reliability purposes, a second coder was first trained on a subset of children and then independently scored 26% of the children (equally distributed by age, gender, and condition) for reliability analyses. Agreement between two independent coders

YOUNG CHILDREN’S UNDERSTANDING OF PROMISE BREAKING

was good to perfect for most coding categories (relevant, desire, request, evaluation, environment, promise comprehension; $\kappa = 0.64\text{--}1.00$). For action comprehension, coders agreed on 66 out of 68 codes (97% agreement), but reliability calculations returned $\kappa = 0.49$ as the coding was very highly skewed towards “correct”. For norm justification, coders agreed on 98 of 102 codes (96% agreement), but norm codes were extremely rare for both coders, resulting in $\kappa = -0.02$.

Table 3: Overview of the coding of children's replies to comprehension questions in Study 2

Category	Description	Examples
Action comprehension	When children correctly describe/mention the action, this was coded as <i>correct</i>	<i>Der hebt die Blätter auf.</i> 'He is picking up the leaves.'
Promise comprehension	Statements of ignorance and false recollections were coded as <i>incorrect</i>	<i>Weiß nicht.</i> 'I don't' know' <i>Der schläft jetzt ein.</i> 'He's going to sleep'
	When children refer to the promise, this was coded as <i>correct</i> . NB: indirect reference to the content of the promise without using the word “(to) promise”, was also coded as <i>correct</i>	<i>Die versprechen sich was.</i> 'They promise something' to each other. ' <i>Der sagt, dass er die Blätter aufräumt.</i> 'He said he will clean up the leaves'
	Statements of ignorance and false recollections were coded as <i>incorrect</i>	<i>Weiß nicht.</i> 'I don't' know' <i>Die verabschieden sich</i> 'They say goodbye '

Data analyses. For each child, we summed whether they chose the promise-breaking ending in the two stories (range: 0-2). To test for effects of condition and age (and control for gender), we ran an ANOVA of the sum-scores using the afex-package in R (Singmann, et al. 2021). We also conducted follow-up t-tests (two-tailed) to further explore group differences.

We focused our analyses of children’s justifications on those children who chose the promise-breaking ending (for completeness, justifications of children who chose the promise-keeping ending can be found in the Supplementary Materials). We collapsed children’s responses across the three justification questions and scored each coding category as yes/no (binary) if it was mentioned at least once by a child. Children received only one binary score per category, irrespective of whether they had chosen the promise-breaking ending once or

twice. This approach is similar to the one chosen in Study 1. We used Fisher exact tests (exact2x2-package) to determine whether there was a difference between conditions (a) in the number of children producing relevant reasons and (b) in the number of children referring to each of the five different relevant reasons. We conducted all analyses in R (R Core Team, 2018). De-identified data and complete documentation of R code and results can be found here: <https://osf.io/rc3tp/>.

Table 4: Overview of the coding scheme for children’s justifications in Study 2

Category	Description	Examples
Norm	Children use normative statements describing the characters’ actions including the verbs <i>müssen</i> ‘should’ or <i>sollen</i> ‘must’.	<i>Der muss helfen.</i> ‘He has to help.’
Desire	Children refer to the character’s desires or wishes.	<i>Der wollte das.</i> ‘He wanted to.’
Request	Children refer to the character being asked to do something or that the other character told the protagonist to do something	<i>Die hat das gesagt.</i> ‘She said it.’
Evaluation	Children evaluate the story without referring to the protagonists in the story.	<i>Mir gefällt das am besten.</i> ‘I like it best.’; <i>Das ist schöner.</i> ‘It is nicer.’
Environment	Children refer to the circumstances of the action without directly mentioning the protagonist or his/her actions.	<i>Weil der Ball auf der Hecke liegt.</i> ‘Because the ball is on the hedge.’

Results

Choices. Children’s choices of the promise-breaking ending differed across ages and conditions (see Figure 3). We found a significant two-way interaction of condition and age, $F(1, 60) = 10.07, p = 0.002$, partial $\eta^2 = 0.14$, and a significant effect of gender, $F(1, 60) = 4.60, p = 0.036$, partial $\eta^2 = 0.07$. On average, girls chose the promise-breaking ending more often ($M = 1.25, SD = 0.84$) than boys ($M = 0.85, SD = 0.71$). There were no significant main effects of condition, $F(1, 60) = 0.06, p = 0.801$, partial $\eta^2 = 0.001$, or age, $F(1, 60) = 0.60, p$

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

= 0.442, partial $\eta^2 = 0.01$. Follow-up t-tests for each group revealed that 3-year-olds chose the promise-breaking ending significantly more often in the selfish condition ($M = 1.25$, $SD = 0.58$) than in the prosocial ($M = 0.71$, $SD = 0.77$), 95% $CI = [-1.03, -0.06]$, $t(31) = -2.28$, $p = 0.030$. Five-year-olds showed the opposite pattern and chose the promise-breaking ending significantly more often in the prosocial condition ($M = 1.44$, $SD = 0.81$) than in the selfish condition ($M = 0.81$, $SD = 0.83$), 95% $CI = [0.03, 1.22]$, $t(30) = 2.15$, $p = 0.040$.

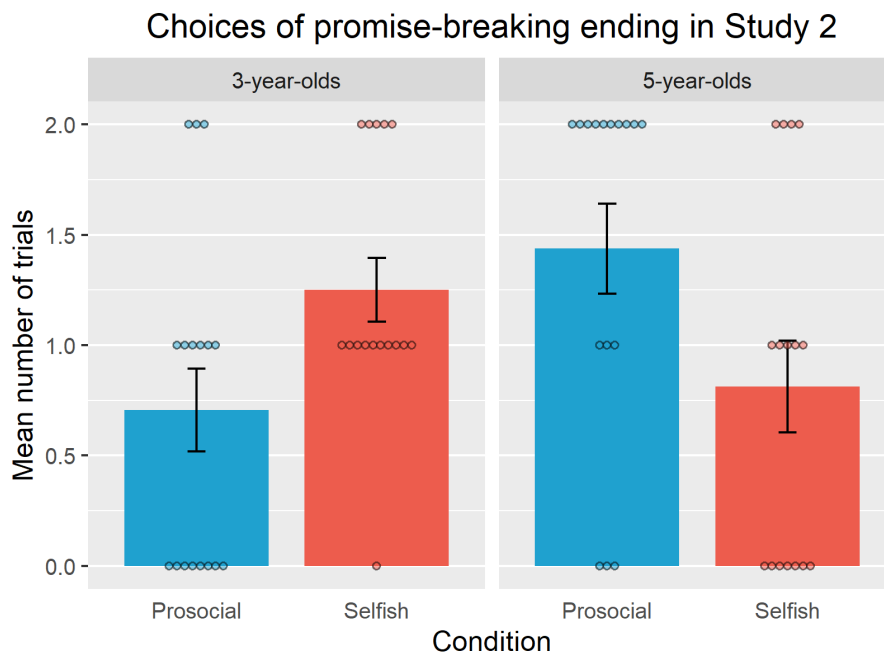


Figure 3. Mean number of trials in which 3- and 5-year-olds chose the promise-breaking ending in Study 2, split by condition. Dots indicate choices of individual children; bars indicate standard errors.

Comprehension questions. Most 3-year-olds and all 5-year-olds answered at least three of four action comprehension questions correctly (3-year-olds: 55% all correct, 30% three correct; 5-year-olds: 91% all correct, 9% three correct). About half of the 3-year-olds most 5-year-olds gave at least one correct answer to the promise comprehension question (3-year-olds: 21% both correct, 24% one correct; 5-year-olds: 66% both correct, 22% one correct).

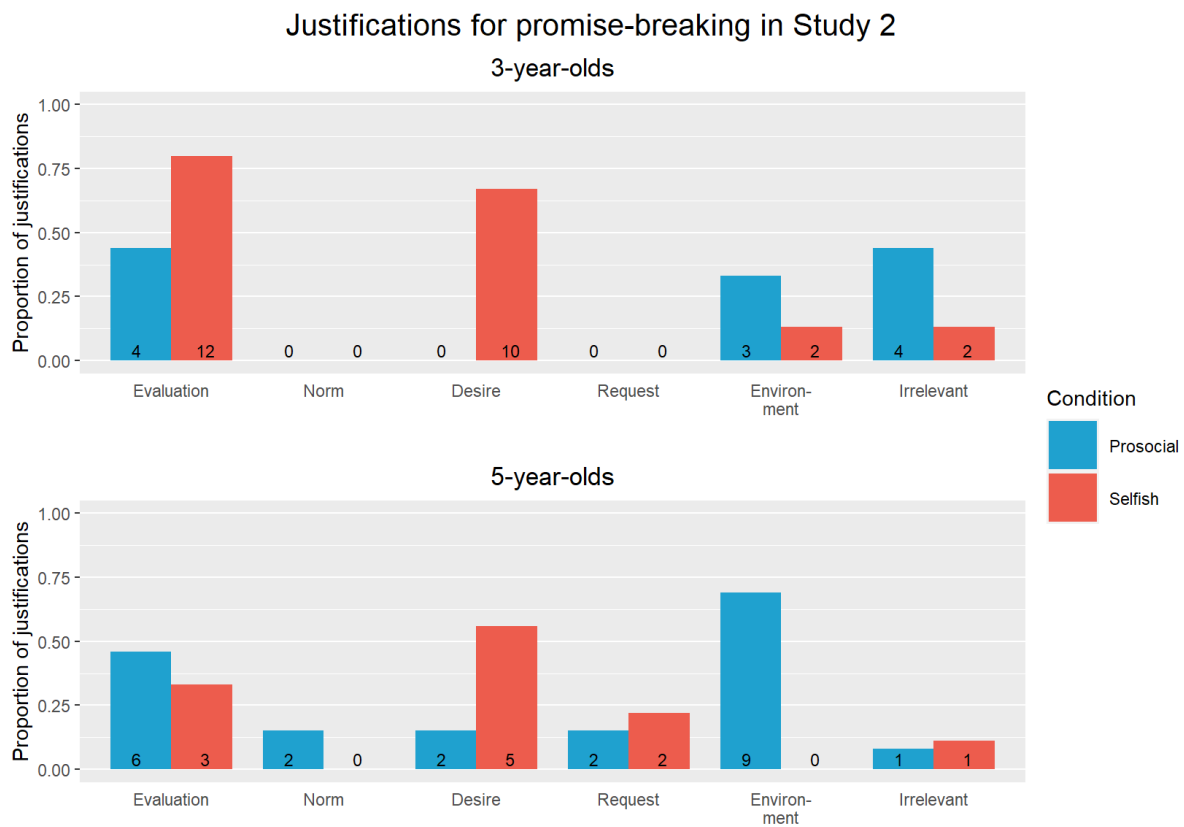


Figure 4. Three- and 5-year-olds’ justifications for choosing the promise-breaking ending in the prosocial and the selfish condition in Study 2. The bars indicate the different relevant responses (evaluation, norm, desire, request, environment) and irrelevant responses (right side). For relevant responses, we scored each category only once for the three questions per story (though it should be noted that responses could be scored in multiple categories). The number of scores per category are printed at the bottom of each bar.

Justifications. We split the data set according to children’s choices of story-endings (promise-keeping vs. promise-breaking). Here we will focus our analyses on children’s justifications in the 68 trials in which they chose the promise-breaking ending, since they are core to this study (for completeness, analyses of justifications for promise-keeping are presented in the Supplementary Materials). We first examined whether children gave relevant justifications for choosing the promise-breaking ending in the two conditions. More 3-year-olds provided relevant justifications in the selfish condition ($n = 13$ of 15 [87%]) than in the prosocial condition ($n = 5$ of 9 [56%]), $p = 0.150$, $OR = 4.81$, $95\% CI [0.68, 46.29]$ (Fisher exact test). Almost all 5-year-olds provided relevant justifications at least once in both conditions (prosocial: $n = 12$ of 13 [92%]; selfish: $n = 8$ of 9 [89%]), $p > .999$, $OR = 0.70$, $95\% CI [0.02, 28.80]$. Next, we turn to analysing children’s reasons for choosing the promise-

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

breaking ending (see Figure 4 for details). Three-year-olds referred to the characters' desire significantly more often in the selfish ($n = 10$ of 15 [67%]) than in the prosocial condition ($n = 0$), $p = 0.002$, $OR = Inf$, 95% CI [2.63, Inf] (Fisher exact test). All other comparisons were non-significant for 3-year-olds, $p_s > .099$, with all 95% CI s for odds ratios overlapping 1. Five-year-olds referred significantly more often to environmental circumstances in the prosocial ($n = 9$ of 13 [69%]) than in the selfish condition ($n = 0$), $p = 0.002$, $OR = 0$, 95% CI [0.00, 0.38]. All other comparisons were non-significant, $p_s > 0.074$, with 95% CI s for odds ratios overlapping 1.

Discussion

To sum up, 5-year-olds chose promise breaking for prosocial reasons (i.e., helping someone) but not for selfish reasons (i.e., playing) when asked about third parties. Most 5-year-olds produced relevant justifications and they referred to situational circumstances (e.g., 'Because otherwise she won't reach the ball') more often in the prosocial than in the selfish condition. This indicates that 5-year-olds preferred helping someone in need over keeping a promise to continue some menial tasks and viewed a promissory obligation as more important than a personal desire to play. In addition, they could point to mitigating circumstances when justifying promise-breaking to help others. In contrast, 3-year-olds chose the promise-breaking ending more often in the selfish as compared to the prosocial condition. To complete this picture, we found that 3-year-olds referred more frequently to characters' desires (e.g., "Because he wanted to play with the ball") in the selfish condition than in the prosocial condition.

Overall, our findings reveal a developmental shift between three to five years of age in children's preferences and justifications about when it is acceptable to break a promise. At age five years, children began to make their decisions in line with what others may view as acceptable circumstances such as helping someone in need. Three-year-olds, however, appear to be guided primarily by others' desires—possibly because younger children are better at

ascribing desires than beliefs to others (Rakoczy, 2010). It is also possible that our results reflect a changing understanding of the choice question (“Which one do you like more?”): 3-year-olds might have chosen based on what they expect to happen (i.e., people behave selfishly), while 5-year-olds might have incorporated normative considerations (i.e., what people *should* do). In line with previous findings, our results suggest that children’s ability to arrive at moral judgments improves throughout the preschool years—especially when they have to consider more than one aspect simultaneously such as an actor’s intentions, direct and indirect consequences of actions or conflicting rules in a dilemma situation (Jambon & Smetana, 2014; Oppenheimer et al., 1997; Smetana et al., 2012).

General Discussion

We investigated children’s understanding of intentional promise breaking through their own behaviour, through their preferences for third party behaviour, and their justifications. Specifically, we focused on whether prosocial obligations would override promissory obligations. We found that 3- and 5-year-olds were reluctant to break their own promises in the prosocial and the selfish condition (Study 1). A more nuanced picture emerged in the third-party context (Study 2): 5-year-olds favoured promise-breaking under prosocial but not under selfish conditions. However, 3-year-olds showed the inverse pattern. In both studies, 5-year-olds provided more justifications that were relevant for the respective context than 3-year-olds. These findings have several implications.

First, our study extends previous findings (Kanngiesser et al., 2017) by showing that 3-year-olds already keep their own promises even when faced with competing prosocial obligations. Second, we demonstrate that from five years of age, children do not only distinguish between intentionally and accidentally broken promises in third party tasks (Maas & Abbeduto, 2001; Mant & Perner, 1988), but also between “good” and “bad” reasons for intentionally breaking a promise. This also adds to previous literature on children’s understanding of prosocial obligations and the situations in which they supersede other social

rules (Dahl, Gross, & Siefert, 2020). Third, our findings indicate that 5-year-olds, but probably not 3-year-olds, understand what justifications are acceptable to others, perhaps because older children recognize that these notions reside in the common ground of the group (Tomasello, 2019). Fourth, children's prosocial and moral behaviours, preferences for third parties, and justifications are rarely studied together (Paulus, Nöth, & Wörle, 2018; Smith, Blake, & Harris, 2013). Our findings suggest that the development of children's behavioural responses to and judgements about promises may follow different developmental trajectories.

What could explain these different developmental trajectories in our study? Previous work has shown that from age two years, children show situational compliance and often do what their parents tell them to do (Kochanska & Aksan, 1995, Kuczynski & Kochanska, 1990). In Study 1, children likely experienced increased pressure to comply because they interacted with an unfamiliar adult and explicitly promised to continue cleaning in their absence. Moreover, another adult attempted to lure the child away from the cleaning task and previous work has shown that young children judge adults as more competent than peers and are less likely to assist adults in instrumental helping tasks (Ulber & Tomasello, 2020; Youniss, 1980). Combined, these factors might have increased children's reluctance to abandon their task in Study 1 and might have masked any difference between conditions that we had expected to find. In contrast, the third-party situation in Study 2 allowed for a more detached view, but also proved more taxing for 3- than for 5-year-olds. This is also evident in the finding that 3-year-olds struggled more than 5-year-olds in recalling the promise in our stories (for a similar finding, see, Isella, Kanngiesser, & Tomasello, 2019).

We now turn to children's justifications for their own and for third parties' behaviour. Their justifications provide insight into what might have driven children's decisions in the two conditions. There are two pieces of data to consider here: (1) whether children produced relevant responses at all and (2) what type of responses they provided. First, more 3-year-olds provided relevant responses in the prosocial condition than in the selfish condition in Study 1

and this pattern was reversed in Study 2 (though note that this difference was not statistically significant in either of the studies). Most 5-year-olds provided relevant responses in both conditions in Study 1 and 2. This (tentatively) indicates that contextual factors exert a greater influence on 3-year-olds than on 5-year-olds and that, for example, very salient events such as the breaking of the marble track (Study 1, prosocial condition) aided 3-year-olds in their reasoning. Previous research has found that children's reasoning becomes more flexible and sophisticated throughout the preschool period (Jambon & Smetana, 2014; Köymen et al., 2019; Mammen et al., 2018; Mercier, 2011; Smetana et al., 2012). Excusing promise-breaking in the selfish condition of Study 1 was arguably demanding and with no salient good excuses available, children had to generate alternative explanations. Five-year-olds did this more frequently than 3-year-olds and explained, for example, that they have already finished the task (even though they had not) or shifted the blame to the other experimenter. This is in line with previous work showing that with age, children become more sophisticated in deceiving others about their own rule transgressions (Evans et al., 2011; Evans & Lee, 2013; Talwar & Lee, 2008).

Second, the types of responses that children produced in our two studies differed between conditions and ages. In Study 1, 3- and 5-year-olds frequently referred to salient environmental events (i.e., the accident) when justifying their promise-breaking in the prosocial condition (but never in the selfish condition). In Study 2, only 5-year-olds referenced environmental factors more often in the prosocial than in the selfish condition (in line with choosing the promise-breaking ending in this condition). This suggests some commonality in how children justify promise-breaking across first- and third- party contexts by pointing to obvious circumstances ('Because the marble track fell over. '; 'Because she kicked the ball up the hedge'). Yet only 5-year-olds in Study 1 referred to norms in the prosocial condition when excusing their promise-breaking. References to norms in the third-person context of Study 2 may have been absent due to pragmatic reasons: If it is well

YOUNG CHILDREN’S UNDERSTANDING OF PROMISE BREAKING

established in the common ground of the interlocutors why an actor should help, underlying rules such as ‘One has to help’ do not need to be explicitly stated (Mammen et al., 2018). In the first-person context of Study 1, 5-year-olds were over-informative by explicitly stating the norm—probably to strengthen their excuse for breaking one norm (promise) in favour of another (helping). By explicitly stating why they broke the promise, 5-year-olds could not only avoid possible punishment, but also maintain their reputation as a “moral” person. Previous studies have found that from five years of age, children care about their reputation and behave more prosocially when their reputation is at stake (Engelmann et al., 2013; Fu et al., 2016; Rapp et al., 2019).

Moreover, 3-year-olds in Study 2 referred more often to the protagonists’ desire in the selfish than in the prosocial condition (5-year-olds showed a similar, but non-significant pattern, probably because fewer children chose the promise-breaking ending in the selfish condition). Even though such a desire might not constitute a legitimate reason for promise-breaking, it is a probable one—after all, there were no other apparent reasons for breaking the promise in the selfish condition. Our findings are in line with previous research showing that children from age two to three years are able to reason about others’ simple desires (Moore et al., 1995; Rakoczy, 2010; Wellman & Woolley, 1990). In the first-person context of Study 1 children did not systematically refer to their own desires, probably because this would not have counted as a good excuse. This could indicate that preschoolers understand what constitutes a legitimate justification in different contexts. Previous research into children’s evaluation of others’ arguments found that children did not differentiate between legitimate and illegitimate reasons for rule deviations until age eight (Schmidt et al., 2016). Our studies expand these findings by showing that preschoolers may already have a nascent understanding of legitimate excuses for their own rule deviations.

Limitations

Our studies have some limitations. First, to match the prosocial and the selfish conditions, both conditions involved social interactions. Specifically, in the selfish conditions a third person asked the child (Study 1) or the actor (Study 2) to play. However, playing with someone who is alone could also be viewed as prosocial act, and young children often prefer interacting or cooperating with others (Rekers et al., 2011; see, Brownell, 2011, for a review). These factors might have influenced children's behaviour and choices in the selfish conditions beyond considerations about whether it is legitimate to break a promise. Future studies could use non-social selfish conditions to control for this possibility.

Second, in Study 2 we focused our analyses on children's justifications of promise-breaking, but due to the preceding choice between a promise-breaking and promise-keeping ending, our sample sizes differed between conditions and age groups. Future studies could only present children with different promise-breaking scenarios and ask them to rate whether it was OK to break the promise and to justify their ratings. This would also explicitly frame promise-breaking as a moral question and stress that competing obligations exist.

Finally, children in our study lived in a city in Germany and came mainly from middle-class backgrounds. Previous research suggests that children's understanding of interpersonal commitments in middle childhood and adolescence varies across cultural contexts (Keller et al., 1998; Miller & Bersoff, 1992). We expect judgements about when it is legitimate to break a promise to vary across cultural contexts depending on the moral common ground of that cultural group. Future research could investigate how children's understanding of the permissibility of promise-breaking develops in different contexts and whether developmental trajectories are similar to the one we found for German middle-class children.

Conclusion

Our study shows that through the preschool years, children develop a growing understanding of how different obligations are ranked (Tomasello, 2019) and that a change in

circumstances can justify promise breaking (Cicero & Miller, 1913). Moreover, with age they become more skilled at providing relevant justifications for their own and other's behaviour. However, 3- and 5-year-olds in our study were not fully able to incorporate these considerations into their own behavioural repertoire. Our findings highlight that the preschool years are a crucial period of moral development where children begin to understand what count as "good" and "bad" reasons for reneging on an obligation and become increasingly aware of the common ground of their group.

References

- Barrett, K., Zahn-Waxler, C., & Cole, P. (2008). Avoiders vs. amenders: Implications for the investigation of guilt and shame during toddlerhood? *Cognition & Emotion*, 7, 481-505.
- Bartsch, K., London, K., & Campbell, M. D. (2007). Children's attention to beliefs in interactive persuasion tasks. *Developmental Psychology*, 43(1), 111.
<http://dx.doi.org/10.1037/0012-1649.43.1.111>
- Bernicot, J., & Laval, V. (1996). Promises in French children: Comprehension and metapragmatic knowledge. *Journal of Pragmatics*, 25(1), 101–122.
[https://doi.org/10.1016/0378-2166\(94\)00075-5](https://doi.org/10.1016/0378-2166(94)00075-5)
- Bicchieri, C. (2016). Covenants without Swords: Group Identity, Norms, and Communication in Social Dilemmas. *Rationality and Society*.
<https://doi.org/10.1177/1043463102014002003>
- Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology*, 2(2), 193–211. <https://doi.org/10.1007/s13164-011-0056-1>
- Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6), 1579–1601. <https://doi.org/10.1111/j.1468-0262.2006.00719.x>
- Cicero, Marcus Tullius. & Miller, Walter (1913). *De officiis*. London (England): Heinemann.
- Dahl, A., Gross, R. L., & Siefert, C. (2020). Young children's judgments and reasoning about

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

- prosocial acts: Impermissible, suberogatory, obligatory, or supererogatory?. *Cognitive Development*, 55, 100908. <https://doi.org/10.1016/j.cogdev.2020.100908>
- Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, 16(6), 952–958. <https://doi.org/10.1111/desc.12086>
- Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology*, 49(10), 1958–1963. <https://doi.org/10.1037/a0031409>
- Evans, A. D., Xu, F., & Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology*, 47(1), 39–49. <https://doi.org/10.1037/a0020787>
- Fay, M. P. (2010). Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics*, 11(2), 373-374. <https://dx.doi.org/10.1093/Fbiostatistics/kxp050>
- Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2016). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science*, 19(2), 275–283. <https://doi.org/10.1111/desc.12304>
- Gilbert, M. (2011). Three dogmas about promising. In H. Sheinman (Ed.), *Promises and agreements: Philosophical essays* (pp. 80–108). New York: Oxford University Press.
- Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, 45(5), 1430–1443. <https://doi.org/10.1037/a0016122>
- Heyman, G. D., Fu, G., Lin, J., Qian, M. K., & Lee, K. (2015). Eliciting promises from children reduces cheating. *Journal of Experimental Child Psychology*, 139, 242–248. <https://doi.org/10.1016/j.jecp.2015.04.013>
- Hussar, K., & Horvath, J. (2013). But You Promised: Children's Judgments of Broken Promises. *Psychology*, 04(12), 1046–1050. <https://doi.org/10.4236/psych.2013.412152>
- Isella, M., Kanngiesser, P., & Tomasello, M. (2019). Children's Selective Trust in Promises. *Child Development*, 90(6), e868–e887. <https://doi.org/10.1111/cdev.13105>

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

- Jambon, M., & Smetana, J. G. (2014). Moral complexity in middle childhood: Children's evaluations of necessary harm. *Developmental Psychology, 50*(1), 22–33.
<https://doi.org/10.1037/a0032992>
- Kachel, U., Svetlova, M., & Tomasello, M. (2017). Three-year-olds' reactions to a partner's failure to perform her role in a joint commitment. *Child Development*.
<https://doi.org/10.1111/cdev.12816>
- 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>
- Kanngiesser, P., Sunderarajan, J., & Woike, J. K. (2021). Keeping them honest: Promises reduce cheating in adolescents. *Journal of Behavioral Decision Making, 34*(2), 183–198.
- Keller, M., & Edelstein, W. (1985). *Reasoning about Promise-Keeping: The Early Development of Interpersonal-Moral Concern*. <https://eric.ed.gov/?id=ED261803>
- Keller, M., Edelstein, W., Schmid, C., Fang, F., & Fang, G. (1998). Reasoning about responsibilities and obligations in close relationships: A comparison across two cultures. *Developmental Psychology, 34*(4), 731–741. <https://doi.org/10.1037/0012-1649.34.4.731>
- Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational sacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*, 109–114. <https://doi.org/10.1016/j.cognition.2012.08.004>
- Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development, 72*(1), 174–186.
<http://dx.doi.org/10.1111/1467-8624.00272>

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

- 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., O'Madagain, C., Domberg, A., & Tomasello, M. (2019). Young children's ability to produce valid and relevant counter-arguments. *Child Development, cdev.13338*. <https://doi.org/10.1111/cdev.13338>
- Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect, 32*(7), 738-748.
- Maas, F. K. (2008). Children's Understanding of Promising, Lying, and False Belief. *Journal of General Psychology, 135*(3), 301–322. <https://doi.org/10.3200/GENP.135.3.301-322>
- 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>
- Mant, C. M., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology, 24*(3), 343–351. <https://doi.org/10.1037/0012-1649.24.3.343>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(02), 57–74. <https://doi.org/10.1017/S0140525X10000968>
- Michael, J., Sebanz, N., & Knoblich, G. (2016). The sense of commitment: A minimal approach. *Frontiers in Psychology, 6*, 1968. <https://doi.org/10.3389/fpsyg.2015.01968>
- Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved?. *Journal of Personality and Social Psychology, 62*(4), 541. <https://doi.org/10.1037/0022-3514.62.4.541>

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Development, 10*(4), 467–482.

[https://doi.org/10.1016/0885-2014\(95\)90023-3](https://doi.org/10.1016/0885-2014(95)90023-3)

Oppenheim, D., Emde, R. N., Hasson, M., & Warren, S. (1997). Preschoolers face moral dilemmas: A longitudinal study of acknowledging and resolving internal conflict. *International Journal of Psycho-Analysis, 78*, 943-957.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a Sword: Self-Governance Is Possible. *American Political Science Review, 86*(2), 404–417.

<https://doi.org/10.2307/1964229>

Paulus, M., Nöth, A., & Wörle, M. (2018). Preschoolers' resource allocations align with their normative judgments. *Journal of Experimental Child Psychology, 175*, 117–126.

doi:10.1016/j.jecp.2018.05.001

Perner, J., & Leekam, S. R. (1986). Belief and quantity: Three-year olds' adaptation to listener's knowledge. *Journal of Child Language, 13*(02).

<https://doi.org/10.1017/S0305000900008072>

Quas, J. A., Stolzenberg, S. N., & Lyon, T. D. (2018). The effects of promising to tell the truth, the putative confession, and recall and recognition questions on maltreated and non-maltreated children's disclosure of a minor transgression. *Journal of Experimental Child Psychology, 166*, 266-279.

Rakoczy, H. (2010). Executive function and the development of belief–desire psychology.

Developmental Science, 13(4), 648-661. doi.org/10.1111/j.1467-7687.2009.00922.x

Rapp, D. J., Engelmann, J. M., Herrmann, E., & Tomasello, M. (2019). Young children's reputational strategies in a peer group context. *Developmental Psychology, 55*(2),

329–336. <https://doi.org/10.1037/dev0000639>

Rawls, J. (1955). Two concepts of rules. *Philosophical Review, 64*, 3–32.

Rekers, Y., Haun, D. B. M., & Tomasello, M. (2011). Children, but not chimpanzees, prefer

to collaborate. *Current Biology*, 21(20), 1756–1758.

<https://doi.org/10.1016/j.cub.2011.08.066>

Rotenberg, K. J. (1980). 'A Promise Kept, a Promise Broken': Developmental Bases of Trust.

Child Development, 51(2), 614–617. JSTOR. <https://doi.org/10.2307/1129307>

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, 7(1), 58–92.

Saylor, M. M., Baird, J. A., & Gallerani, C. (2006). Telling others what's new: Preschoolers' adherence to the given-new contract. *Journal of Cognition and Development*, 7(3), 341–379. https://doi.org/10.1207/s15327647jcd0703_7

Scanlon, T. (1990). Promises and practices. *Philosophy & Public Affairs*, 19, 199–226.

Schmidt, M. F. H., Svetlova, M., Johe, J., & Tomasello, M. (2016). Children's developing understanding of legitimate reasons for allocating resources unequally. *Cognitive Development*, 37, 42–52. <https://doi.org/10.1016/j.cogdev.2015.11.001>

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S. (2021). afex: Analysis of Factorial Experiments. R package version 0.28-1. <https://CRAN.R-project.org/package=afex>

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Smetana, J. G., & Ball, C. L. (2017). Young Children's Moral Judgments, Justifications, and Emotion Attributions in Peer Relationship Contexts. *Child Development*. <https://doi.org/10.1111/cdev.12846>

Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental Changes and Individual Differences in Young Children's Moral Judgments. *Child Development*, 83(2), 683–696. <https://doi.org/10.1111/j.1467-8624.2011.01714.x>

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

- Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLOS ONE*, 8(3), e59510. <https://doi.org/10.1371/journal.pone.0059510>
- Talwar, V., & Lee, K. (2008). Social and Cognitive Correlates of Children's Lying Behavior. *Child Development*, 79(4), 866–881. <https://doi.org/10.1111/j.1467-8624.2008.01164.x>
- Theimer, C. E., Killen, M., & Stangor, C. (2001). Young children's evaluations of exclusion in gender-stereotypic peer contexts. *Developmental Psychology*, 37(1), 18–27. <https://doi.org/10.1037//0012-1649.37.1.18>
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. The Belknap Press of Harvard University Press.
- Ulber, J. and Tomasello, M. (2020). Young children's prosocial responses towards peers and adults in two social contexts. *Journal of Experimental Child Psychology*, 198, 104888. <https://doi.org/10.1016/j.jecp.2020.104888>
- Vaish, A., Carpenter, M., & Tomasello, M. (2016). The Early Emergence of Guilt-Motivated Prosocial Behavior. *Child Development*, 87(6), 1772–1782. <https://doi.org/10.1111/cdev.12628>
- van den Assem, M. J., van Dolder, D., & Thaler, R. H. (2011). Split or Steal? Cooperative Behavior When the Stakes Are Large. *Management Science*, 58(1), 2–20. <https://doi.org/10.1287/mnsc.1110.1413>
- Van de Vondervoort, J. W., & Hamlin, J. K. (2017). Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *Journal of Experimental Child Psychology*, 164, 136-151. <https://doi.org/10.1016/j.jecp.2017.07.004>
- Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations¹. *Econometrica*, 76(6), 1467–1480. <https://doi.org/10.3982/ECTA7673>

YOUNG CHILDREN'S UNDERSTANDING OF PROMISE BREAKING

- Warneken, F. (2015). Precocious prosociality: Why do young children help?. *Child Development Perspectives*, 9(1), 1-6. <https://doi.org/10.1111/cdep.12101>
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, 35(3), 245–275. [https://doi.org/10.1016/0010-0277\(90\)90024-E](https://doi.org/10.1016/0010-0277(90)90024-E)
- Woike, J. K., & Kanngiesser, P. (2019). Most People Keep Their Word Rather Than Their Money. *Open Mind*, 3, 68–88. https://doi.org/10.1162/opmi_a_00027
- Woods, J. (2016). The Normative Force of Promising. *Oxford Studies in Normative Ethics*, 6.
- Youniss, J. (1980). *Parents and peers in social development: A Sullivan-Piaget perspective*. Chicago: University of Chicago Press.