



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Young children mostly keep, and expect others to keep, their promises



Patricia Kanngiesser^{a,b,*}, Bahar Köymen^{c,b}, Michael Tomasello^{b,d}

^a Faculty of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany

^b Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, 01408 Leipzig, Germany

^c School of Health Sciences, University of Manchester, Manchester M13 9PL, UK

^d Department of Psychology and Neuroscience, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Article history:

Received 14 June 2016

Revised 3 February 2017

Keywords:

Promises

Obligations

Commitments

Social norms

Speech acts

Prosocial

ABSTRACT

Promises are speech acts that create an obligation to do the promised action. In three studies, we investigated whether 3- and 5-year-olds ($N = 278$) understand the normative implications of promising in prosocial interactions. In Study 1, children helped a partner who promised to share stickers. When the partner failed to uphold the promise, 3- and 5-year-olds protested and referred to promise norms. In Study 2, when children in this same age range were asked to promise to continue a cleaning task—and they agreed—they persisted longer on the task and mentioned their obligation more frequently than without such a promise. They also persisted longer after a promise than after a cleaning reminder (Study 3). In prosocial interactions, thus, young children feel a normative obligation to keep their promises and expect others to keep their promises as well.

© 2017 Elsevier Inc. All rights reserved.

Introduction

Promises express commitments to future actions. Austin (1975) was the first to systematically describe promises as speech acts, highlighting that they are more than mere words but also a way

* Corresponding author at: Faculty of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany.

E-mail address: patricia.kanngiesser@fu-berlin.de (P. Kanngiesser).

of acting with words. In many societies, saying the words “I promise to do A” or simply “I will do A” under the appropriate circumstances creates an obligation to do A (Searle, 1969). Yet, what makes promises binding and people keep their word is debated among philosophers (Gilbert, 2011; Prichard, 2002; Rawls, 1955; Scanlon, 1990) as well as behavioral scientists (Charness & Dufwenberg, 2006; Ellingsen & Johannesson, 2004; Vanberg, 2008). Some theorists view promises as social practices that play an important role in sustaining social coordination and cooperation (Bicchieri, 2002; Hume, 1890; Lewis, 1969/2002); indeed, experimental work has overwhelmingly shown that rates of cooperation in social dilemmas increase substantially when participants promise each other to cooperate (Orbell, Van de Kragt, & Dawes, 1988; Ostrom, Walker, & Gardner, 1992; Sally, 1995).

Studies with children have found that (English-speaking) children begin to talk about commitments (“I will do it”) and produce promises from 4 to 5 years of age (Astington, 1988a; Diessel, 2004). At around the same age, children adjust their behavior following an adult’s request to promise; Han Chinese children were less likely to cheat in a competitive game after having promised to follow the rules (Heyman, Fu, Lin, Qian, & Lee, 2015), and North American children revealed transgressions (i.e., reported that they had played with a forbidden toy) more frequently after having promised to tell the truth (Lyon & Dorado, 2008; Lyon, Malloy, Quas, & Talwar, 2008; Talwar, Lee, Bala, & Lindsay, 2002). Yet, the effect of promises is more variable, and potentially more context specific, in children under 5 years of age: Heyman et al. (2015) found that promises did not reduce cheating rates in 4-year-olds in a competitive game. Although Talwar et al. (2002, Experiment 3) found a significant main effect of promises on reducing lying behavior for 3- to 7-year-olds, closer inspection of the data shows that the lying rates decreased only marginally for 3- and 4-year-olds (3-year-olds: promises vs. discussion—47% lying vs. 50% lying; 4-year-olds: promises vs. discussion—67% lying vs. 71% lying).

The above findings suggest that from 4 to 5 years of age, children produce promises and start to keep their own promises, yet their abilities to correctly reason about promises have been found to develop later. Studies have shown that younger children do not distinguish between the speech act of promising (or committing) and the performance of the act, and younger children often prioritize outcomes when judging others’ behavior (e.g., Astington, 1988c; Kalish & Cornelius, 2007; Rotenberg, 1980). From 7 years of age, children reason that speakers are responsible for fulfilling what they promised (Astington, 1988c; Maas & Abbeduto, 1998) and take into account whether someone broke a promise (or a commitment) intentionally or accidentally (Maas & Abbeduto, 2001; Mant & Perner, 1988). But it is not until their teens that (English- and French-speaking) children begin to distinguish promises (“I promise I will play with you”) from other types of speech acts such as predictions (“I promise it will rain”) and assertions (“I promise, it rained”) (Astington, 1988b; Bernicot & Laval, 1996).

The overall developmental picture based on the findings to date is that children under 5 years of age mostly fail to keep their own promises and do not reason that others ought to keep their promissory obligations. However, two lines of evidence challenge this view: (a) studies on young children’s understanding of joint commitments in collaborative social activities and (b) studies on young children’s understanding of normativity.

First, recent work on children’s behavior in collaborative social activities has shown that 3-year-olds already behave in a committed manner and expect others to be committed to a joint endeavor. Specifically, 3-year-olds will reengage an adult when she stops to participate in a joint activity, will acknowledge their own leaving of a joint play activity, and will wait, help, and take over the partner’s role after collaboration (Gräfenhain, Behne, Carpenter, & Tomasello, 2009; Gräfenhain, Carpenter, & Tomasello, 2013). However, these studies elicited an explicit (verbal or nonverbal) agreement from children to collaborate, engaged the children in a collaborative activity, and compared this collaboration condition to a condition with individual activity (and no agreement). Thus, it is unclear whether the increased commitment in the collaboration condition (as compared with the individual condition) resulted from the explicit agreement, the collaborative activity, or both. In fact, other work has shown that collaboration by itself (without verbal agreements) makes preschoolers more committed to finishing a task—both when the partner is present (Hamann, Warneken, & Tomasello, 2012) and when the partner is absent (Butler & Walton, 2013). In our studies, we wanted to focus exclusively on the effect of verbal agreements (i.e., promises) on young children’s behavior. In each study, therefore,

we used the same prosocial tasks across all conditions and only varied whether a promise was given or not.

Second, one important characteristic of promises is that they are normatively binding (i.e., they entail an obligation that one should keep one's word), and research has shown that young children already comprehend the normativity of some social rules and will hold responsible others who do not abide to norms (Schmidt & Tomasello, 2012). For example, 3-year-olds will actively correct a puppet actor that fails to play a game correctly (Rakoczy, Warneken, & Tomasello, 2008) or will intervene if a puppet tries to take someone else's property (Rossano, Rakoczy, & Tomasello, 2011), often using norm-related language during their interventions (e.g., "You did that wrong," "You must not do this"). Moreover, preschoolers also start to respond to moral transgressions and judge that it is wrong to harm others (Smetana, 2013). To date, no study has investigated whether young children will also comprehend the normative implications of promising (e.g., that it is wrong to break a promise).

Therefore, we conducted a series of studies investigating children's understanding of promises focusing on their understanding of others' promissory obligation (Study 1) and their own commitment after having promised to do something (Studies 2 and 3). In all three studies, children were engaged in prosocial interactions (helping)—either with a puppet partner (Study 1) or with adult experimenters (Studies 2 and 3). We conducted the studies with 3-year-olds, who have been shown to comprehend both joint commitments and the normative implications of social rules (Gräfenhain et al., 2009; Schmidt & Tomasello, 2012). We also included 5-year-olds in all studies to explore whether a more nuanced understanding of promises may emerge later during the preschool years (e.g., Heyman et al., 2015).

Study 1

In this study, we investigated whether 3- and 5-year-old children expect others to keep their promises. The 3- and 5-year-olds helped a puppet partner to retrieve sticker rewards that only the puppet could access. The puppet either promised to share the stickers (promise group) or only asked for help to retrieve them (control group). In both groups, the puppet then decided to keep all of the stickers to itself. Because the outcome was held constant in both groups (i.e., the child did not receive any stickers from the puppet), the meaning of the puppet's action differed due to the preceding social interaction; it signified either the breaking of a promise to share (promise group) or an act of stinginess (control group). We coded children's spontaneous responses to the puppet's behavior, focusing in particular on whether they would refer to the puppet's obligation or promise to share.

Method

Participants

In total, 40 3-year-olds ($M = 3;7$ [years;months], range = 3;0–3;11; 20 girls and 20 boys) and 40 5-year-olds ($M = 5;6$, range = 4;11–5;11; 20 girls and 20 boys) took part. An additional 10 3-year-olds and 9 5-year-olds were excluded because of failure to protest during the warm-up phase of the experiment (9 3-year-olds and 8 5-year-olds), inattentiveness (1 3-year-old), or experimenter error (1 5-year-old). Children were recruited through a database of parents who had agreed to have their children participate in developmental studies. They lived in a medium-sized German city. They were all native speakers of German and came mostly from middle-class families.

Procedure

Children interacted with a puppet partner that was animated by a female experimenter (E1). The experiment was divided into two phases: a warm-up phase and a test phase.

During the warm-up phase, a second female experimenter (E2) introduced children to the puppet, who then played a short warm-up game together. Next, children experienced three warm-up protest trials to test children's willingness to spontaneously correct the puppet. In each warm-up trial, the puppet made an instrumental mistake such as drinking from a cup upside down, brushing the floor instead of its teeth, or holding a brush the wrong way around when sweeping the floor. Children

had 30 s to correct the mistake while the puppet became increasingly puzzled about failing to achieve its goal. If children did not spontaneously correct the puppet, they were directly asked to do so at the end of each trial. Only children who spontaneously corrected the puppet in at least one of the three trials participated in the test phase. All other children were dropped from the experiment because they were judged to lack the motivation (or to be too inhibited) to correct the puppet.

For the 17 children who were excluded because of failure to protest during the warm-up phase, test trial data are available for 11 children (5 in the promise group and 6 in the control group). For the remaining 6 children, the procedure was stopped after absence of protest in the warm-up trials. For the 11 children with test trial data, there was one instance of protest (out of three trials) by a 3-year-old boy in the control group. The remaining children never protested, and only 2 children spoke to the puppet once (one asked whether the puppet was real, and the other asked the puppet to keep pulling on the rope). This means that there was a protest rate of 1/33 (3%) during the test phase for this subsample of children (i.e., children who did not protest during the warm-up). We believe that this warrants our decision to exclude these children from our final sample.

During the test phase, E2 introduced a box from which stickers could be retrieved (see Fig. 1). The stickers were placed on trays inside the box, and both partners needed to pull a rope together to move the trays and take out the stickers. There was one sticker tray on the child's side and one on the puppet's side. Children first participated in one *practice trial*, where E2 explained how the box worked and told children that they could collect the stickers in their bowl and take them home later (there were two stickers on each side).

Next, children participated in one *trust trial*, where the puppet had all four stickers on its side and either promised to share half of the stickers with the child and asked the child to pull the rope (*promise group*) or only asked the child to pull the rope (*control group*). In both groups, the puppet handed two stickers to the child to establish trust. This trial was followed by three *test trials*, where four stickers were always placed on the puppet's side. At the start of the trial, the puppet made children aware of the stickers. In the promise group the puppet then promised to share half of the stickers ("Ich gebe dir die Hälfte der Aufkleber ab. Das verspreche ich dir. Also versprochen!") [I will give you half of the stickers. I promise. Promised!]). In the control group the puppet did not say anything. In both groups, the puppet then asked the child to pull on the rope.

Crucially, after having retrieved the stickers, the puppet announced that the stickers were very nice and that it would keep them all. In the promise group, this statement signaled breaking of the puppet's promise to share with the child. In the control group, it simply stated the puppet's desire to keep all of the stickers. The child had 30 s to react and possibly protest against the puppet's behavior, after which



Fig. 1. The setup used in Study 1. The child sat on the pillow next to the puppet (here: the empty pillow). The box was covered with Plexiglas, and the child and puppet needed to pull the rope together to move the board with the stickers toward them. The stickers could then be retrieved through the hole in front of the puppet and the child, respectively.

the puppet gave the child two stickers, pretending that it had forgotten to do so earlier. This last part was included to depict the puppet as forgetful and friendly (and not as mean) and to help children to maintain some level of trust in subsequent trials. The procedure was repeated for each of the three test trials. The entire session (including the warm-up and test phase) lasted approximately 10–15 min.

Data coding and analyses

All sessions were videotaped. We coded children's behavior and transcribed their utterances verbatim from videotape. We focused on children's protest in the 30 s following the puppet's announcement that it would keep all of the stickers. We coded children's protest behavior according to the following categories:

1. *Promise normative*: Children refer to the puppet's obligation/promise to share stickers using normative language with deontic modal verbs such as "must" and "should"; for example, "Aber was man versprochen hat, das muss man auch halten" (But you must keep what you promised), "Das darfst du aber nicht" (You mustn't do this).
2. *Promise non-normative*: Children referred to the promise or its content in a non-normative manner; for example, "Du hast mir das versprochen" (You promised it to me), "Du hast doch gesagt, du gibst mir die Hälfte ab" (You said you would give me half), "Du wolltest welche abgeben" (You wanted to share some).
3. *Personal desire*: Children expressed their desire for the stickers by complaining about not getting any stickers or directly requested stickers from the puppet; for example, "Jetzt hab ich wieder keine" (Now, I don't have any again), "Kann ich auch mal auf meiner Seite welche haben?" (Can I have some on my side?), "Gib mir meine zwei Aufkleber" [Give me my two stickers].
4. *Simple*: Children said "nein" (no) without further task-related elaboration.

It was only scored whether protest occurred in a trial or not (we did not score the frequency of protest in a trial). We conducted analyses on two types of protest measures:

1. We analyzed children's overall protest in each trial, scoring if they protested in a trial or not (yes = 1, no = 0). We analyzed the binary overall protest data with generalized linear mixed models (GLMM), using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015; R Development Core Team, 2014). Protest (yes = 1, no = 0) was entered as the response variable. We included a two-way interaction of experimental group (promise vs. control) and age group (3- vs. 5-year-olds) as predictor as well as the control predictors gender (male vs. female) and trial number (*z*-transformed). Participant ID was entered as a random effect to account for the repeated-measures design (i.e., multiple trials per child). We conducted model comparisons with likelihood ratio tests using the analysis of variance (ANOVA) method in R. Specifically, we compared a full model (all predictors) with a null model (containing the control predictors gender and trial number). In addition, we calculated a protest sum-score for each child by adding overall protests across the three trials (range = 0–3). We compared for each age group whether protest sum-scores differed between the promise and control groups using nonparametric tests.
2. We analyzed the different types of children's protest. Here, we coded only the highest level of protest per trial, with normative promise protest being the highest category and simple protest being the lowest category. For the analyses, we calculated a sum-score for promise protest (normative and non-normative promise) and a sum-score for other protest (desire and simple) for each child, adding the number of trials with promise protest (range = 0–3) and the number of trials with other protest (range = 0–3), respectively. Categories were combined for analyses because of the low occurrence of normative promise protest and of simple protest. In addition, we recorded the number of children who protested at least once in the three trials and scored only the highest level of protest they used. Data were analyzed for each age group separately using nonparametric statistics.

All data were coded by the first author. A second coder recoded 25% of the data for reliability purposes ($n = 20$ children in total, equally distributed by age, experimental group, and gender). The irr

package in R was used for reliability calculations (Gamer, Lemon, Fellows, & Singh, 2012). Agreement between coders on protest categories was very good (equal-weighted $\kappa = .90$).

Results and discussion

First, we investigated children's overall protest in the three test trials (see Fig. 2). We found that 3-year-olds protested in 40% of trials in the promise group and in 42% of trials in the control group, whereas 5-year-olds protested in 47% of trials in the promise group and in 33% of trials in the control group. We analyzed children's protest behavior with Generalized Linear Mixed Models (GLMM) and found that a full model did not have a significantly better fit to the data than a null model (including only gender and trial), $\chi^2 = 1.95$, $df = 3$, $p = .584$ (see Supplementary Table 1 in online supplementary material for details). Thus, neither experimental group nor age group significantly predicted children's overall protest behavior. Separate analyses for the two age groups using protest sum-scores showed no significant differences between the promise and control groups for 3-year-olds, $U = 210$, $p = .790$, and for 5-year-olds, $U = 164$, $p = .327$ (Mann–Whitney U tests). Together, these analyses indicate that 3- and 5-year-olds in the promise and control groups protested similarly about not receiving any sticker rewards.

However, we observed qualitative differences in children's protest behavior in the two groups. Fig. 2 reports the percentages of all four protest categories. Because normative promise protest and simple protest occurred in less than 10% of the trials, we used only two categories for the subsequent statistical analyses: promise protest (including normative and non-normative promise protest) and other protest (including desire and simple protest). Only children in the promise group used normative and non-normative promise protest such as “You promised it to me” or “One shouldn't break a promise” (promise protest collapsed: 3-year-olds in 17% of trials, 5-year-olds in 27% of trials), whereas this type of protest was completely absent in the control group. We summed the two types of promise protest across trials for each child and found a significant difference between the promise and control groups for 3-year-olds, $U = 150$, $p = .047$, and for 5-year-olds, $U = 100$, $p < .001$ (Mann–Whitney U tests). Children in the control group used other types of protest; they protested primarily about not getting any stickers or expressed their desire for more stickers using utterances such as “I have so few and she has so many” (other protest collapsed: 3-year-olds in 42% of trials, 5-year-olds in 33% of trials). These other types of protest occurred less often in the promise group (3-year-olds in 23% of trials, 5-year-olds in 20% of trials). There was a trend for a significant difference in other protest

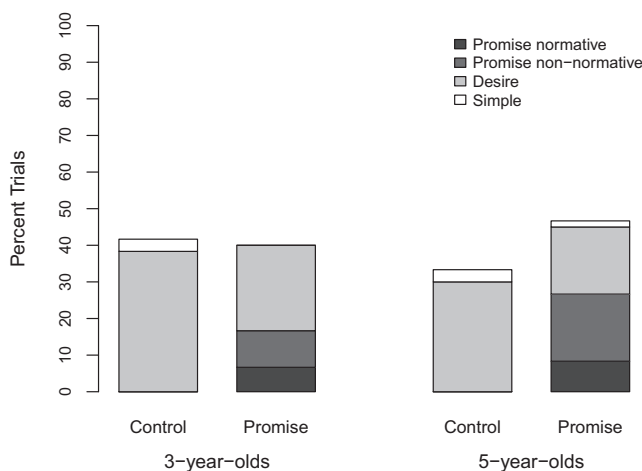


Fig. 2. Percentages of trials in Study 1 in which 3- and 5-year-olds in the promise and control groups used normative promise protest (dark gray bar), non-normative promise protest (medium gray bar), desire protest (light gray bar), and simple protest (white bar).

Table 1

Numbers of children who protested at least once in the three test trials in Study 1 (reporting only the highest level of protest that occurred).

Age (years)	Condition	Promise norm	Personal desire	No protest
3	Promise	5	5	10
	Control	0	13	7
5	Promise	10	2	8
	Control	0	8	12

sum-scores between the promise and control groups for 3-year-olds, $U = 258$, $p = .090$, and no significant difference for 5-year-olds, $U = 225$, $p = .437$ (Mann–Whitney U tests). Finally, we compared how many children protested at least once in the three trials, scoring only the highest level of protest (see Table 1). One quarter (25%) of 3-year-olds in the promise group used promise protest at least once (25% used other protest), and 65% of 3-year-olds in the control group used other protest at least once, $p = .008$ (Fisher exact test). Half (50%) of 5-year-olds in the promise group used promise protest at least once (10% used other protest), and 40% of 5-year-olds in the control group used other protest, $p < .001$.

These findings show qualitative differences in children's protest behavior in the control and promise groups; whereas children in the control group primarily voiced their desire for stickers, only children in the promise group referred to promising norms and considered the puppet responsible for not sharing. Thus, by 3 years of age, some children in our sample (25%) expected a promisor to perform the promised act, and these expectations became more pronounced in 5-year-olds (50%). It should be noted that 9 of 50 (18%) 3-year-olds and 8 of 49 (16%) 5-year-olds were excluded from the sample because of failure to protest in the warm-up, which potentially limits the generalizability of our findings.

It is possible that children protested differently in the control and promise groups because only in the promise group were children told that they would receive half of the stickers (the puppet never mentioned sharing in the control group). Prior work has shown that 3-year-olds share rewards equally after collaborating on a task that is identical to the one used in our study (Hamann, Warneken, Greenberg, & Tomasello, 2011; Warneken, Lohse, Melis, & Tomasello, 2011). This suggests that children in our study likely had an expectation to receive stickers in both conditions. This claim is supported by our finding that children's overall protest behavior did not differ between the promise and control groups, indicating that children in both groups were equally dissatisfied with not receiving any stickers. But crucially, only in the promise group did some children mention the puppet's obligation to share.

Earlier work had shown that it is not until 7 years of age (possibly somewhat earlier) that children reason that speakers are responsible for fulfilling promises (e.g., "I promise I will play with you") but not for fulfilling predictions (e.g., "I promise it will rain") (Astington, 1988c; Maas & Abbeduto, 1998). We show that preschoolers already assume that promises obligate a speaker to perform the promised act, although our study design does not allow us to draw any conclusions as to whether preschoolers differentiate promises and predictions. Yet, a series of related studies has shown that preschoolers direct their protest at actors for failing to follow a directive (e.g., "Put the red thing there!") and at speakers for failing to make accurate predictions or assertions (e.g., "The red thing is/will be there") (Lohse, Gräfenhain, Behne, & Rakoczy, 2014; Rakoczy & Tomasello, 2009). Taken together, this indicates that young children may have a more nuanced understanding of speech acts and speakers' responsibility than was previously assumed.

One of the questions left open is whether the different developmental onsets arise from methodological differences between studies (action-based methodologies vs. interviews about third-party stories) or are indicative of children's conceptual developmental (implicit action-based understanding vs. explicit knowledge about the correct use of promises). Future research could help to clarify this question by combining first-party action-based and third-party judgment-based methodologies in studying young children's understanding of different speech acts.

Study 2a

Whereas the findings from Study 1 indicate that from 3 years of age children expect others to keep their promises in a prosocial interaction, evidence to date suggests that children younger than 5 years mostly fail to keep their own promises (Heyman et al., 2015; Talwar et al., 2002). However, in these studies children were asked to promise to refrain from engaging in forbidden behaviors such as cheating or lying. It is an open question whether the findings indicate a general inability to keep promises in younger children (e.g., due to limited inhibitory control skills) or whether they indicate a context-specific inability (e.g., limited to forbidden behaviors). In fact, another line of evidence has shown that 3-year-olds behave committedly after a joint commitment to a collaborative task (Gräfenhain et al., 2009, 2013), suggesting that contexts could play a role in scaffolding children's promise keeping.

Therefore, we wanted to investigate whether young children will keep their own promissory obligations using a collaborative helping task. To get a graded measure of children's commitment to the helping task, we used increasingly persistent cues to distract children from the task. Specifically, we engaged 3- and 5-year-olds in a cleaning task with an adult experimenter. When the experimenter needed to leave the task, she either asked children a series of prompt questions to elicit a spontaneous promise to continue cleaning (promise group) or immediately left without asking any prompt questions (control group). A second experimenter then tried to entice children away from the cleaning task with an exciting toy. We expected children in the promise group to resist temptation longer than children in the control group.

Method

Participants

In total, 39 3-year-olds ($M = 3;9$, range = 3;6–3;11; 19 girls and 20 boys) and 40 5-year-olds ($M = 5;9$, range = 5;5–5;11; 20 girls and 20 boys) took part in the study. An additional 2 3-year-olds were excluded, 1 due to a technical error (video camera not switched on) and 1 who never committed to the task in the promise group. Children were recruited in the same manner as in Study 1.



Fig. 3. Setup used in Study 2. E1 accidentally spilled paper shreds on the carpet (pictured on the left) and asked the child to help clean up. E2 later sat in front of the marble track on the right and tried to distract the child from the cleaning task.

Procedure

Children were tested individually in a quiet room of their nursery. Half of the children were assigned to the promise group and the other half to the control group. Children in both groups first played a familiarization warm-up puzzle with the two female experimenters (E1 and E2). Next, E1 showed the child a marble track in one corner of the room while E2 left the room (pretending to look for the kindergarten head teacher). The marble track was colorfully decorated and made jingle sounds when marbles were thrown in (see Fig. 3). While the child was distracted playing with the marble track, E1 walked backward and pretended to spill a bucket of shredded paper on a carpet. After the accident, E1 drew the child's attention to her mishap, explained that she would need a clean carpet later on, and asked the child whether she or he could help with the cleanup. This means that children in both groups (control and promise) were initially asked to commit to helping E1 with the cleaning task. E1 and the child spent about 15 s cleaning the carpet before E2 entered the room and told E1 that the kindergarten head teacher wanted to talk to her about something important. This was introduced to give E1 a credible excuse to leave the room.

Before she left the room, E1 asked children in the promise group a series of prompt questions to elicit a spontaneous commitment to continue the cleaning task. The following questions were asked in a fixed sequence: (1) "Was sollen wir machen? Hast du eine Idee?" (What should we do? Do you have any idea?); (2) "Wer räumt hier den weiter auf?" (Who will keep cleaning?); (3) "Könntest du das vielleicht machen?" (Could you maybe do it?); (4) "Versprichst du mir, dass du weiter aufräumst? Also versprochen?" (Do you promise to keep cleaning? Promised, okay?). If children spontaneously committed after one of the earlier prompt questions, all remaining prompt questions except the promise question (Question 4) were omitted. That is, all children in the promise group were eventually asked to promise. E1 then announced that she would be back soon. In the control group, E1 omitted all prompt questions and simply stated that she would be back soon.

Next, children in both groups took part in the distraction phase of the experiment. While E1 left the room, E2 sat down in front of the marble track with her back toward the child. She then spent 60 s trying to entice the child away from the cleaning task using increasingly explicit distraction cues (one cue every 15 s as follows). After 15 s, E2 simply commented that the track was great ("Ah, toll"). After 30 s, she threw a marble in the track (which produced a jingle sound), commenting that this is how the marble track worked ("Ah, so funktioniert das"). After 45 s, she threw another marble in the track, turned to look at the child, and said that the marble track was great ("Oh schau mal, das ist eine tolle Marmelbahn"). Finally, after 60 s, E2 turned again toward the child and asked whether the child wanted to play ("Spielst du mit?"). If the child refused to play, E2 asked her why not ("Warum denn nicht?"). At this point E1 returned, thanked the child for cleaning, and allowed the child to play with the marble track. The entire test session lasted approximately 10 min.

Data coding and analyses

The study was videotaped, and children's behavior was later coded from videotape. For children in the promise group, we coded how many prompt questions were asked before children committed to the cleaning task using a score from 0 to 4 (with higher scores indicating increasing levels of prompting). For example, children received a score of 0 if they volunteered to help before E1 had asked the first prompt question, and they received a score of 4 if they committed only after being asked to promise (Question 4: "Do you promise to keep cleaning?").

For children in both groups, we coded how committed they were to continue the cleaning task. We assigned a commitment score from 0 to 5 based on the number of cues E2 needed to use in order to distract the child from her or his task, with higher scores indicating greater levels of commitment. Distraction was scored if the child started playing with the marble track (e.g., by inserting a marble) or if the child stopped cleaning and left the carpet for more than 15 s (we scored the distraction cue at the start of the 15 s). For example, children received a score of 0 if they left the carpet immediately after E1 had left the room, a score of 1 if they left after the first distraction cue, and a score of 5 if they kept cleaning even after the final distraction cue (i.e., after E2 asked whether they wanted to play and, if not, why not).

We also coded whether children referred to their promise or the cleaning task at any point during the distraction phase using the following coding scheme:

- (a) *Promise*: Children explicitly refer to their promise to clean; for example, “Hab ich versprochen” (I promised).
- (b) *Normative cleaning*: Children use normative language when referring to their commitment to do the cleaning task; for example, “Aber ich muss das aufräumen” (But I have to clean this) and “Ich muss das hier alles fertig machen” (I have to finish all of this).
- (c) *Simple (non-normative cleaning)*: Children refer to their cleaning without using normative language; for example, “Ich räume hier auf” (I am cleaning) and “Ich mach die Schnipsel” (I am doing the paper shreds).

If children used different types of protest, we coded only the highest level, with promise being the highest category and simple being the lowest category.

Data were analyzed in R using the exactRankTests package for all nonparametric tests, the ordinal package for all ordinal linear models, the Kendall package for correlations, and the irr package for reliability calculations (Christensen, 2015; Gamer et al., 2012; McLeod, 2011). We used the Kendall package (Kendall function) and calculated Kendall rank correlations because our data were ordered and contained ties. The Kendall function (a) computes a tau-b (τ_B) coefficient that is suited for tied data and (b) approximates a *p* value for tied data that is an adequate approximation of an exact *p* value. In the ordinal linear models, we entered commitment scores as response variable (ordered factor from 0 to 5), age group (3- vs. 5-year-olds), experimental group (promise vs. control), and their two-way interaction as predictors and gender (female vs. male) as a control predictor. We conducted model comparisons with likelihood ratio tests using the ANOVA method in R. Specifically, we compared a full model (all predictors) with a null model (control predictor gender) as well as a full model and a reduced model (excluding certain predictors).

A second coder scored 25% of the data for reliability purposes (equally distributed by age group, experimental group, and gender). Agreement between observers was good for children’s responses to the prompt question in the promise group, that is, at which question children agreed to do the cleaning task (weighted $\kappa = .71$, $n_{\text{reliability}} = 10$ children). All other reliability analyses were calculated together with data from study 2b because the same reliability coder coded the three groups ($N = 30$ children in total, equally distributed by age group, experimental group, and gender). Agreement between observers was good for children’s commitment scores during the distraction phase (equal-weighted $\kappa = .80$) and for children’s spontaneous mentioning of their commitment (equal-weighted $\kappa = .74$).

Results and discussion

First, we analyzed how many prompt questions children in the promise group needed before committing to the cleaning task. The vast majority of 3-year-olds (90%) and of the 5-year-olds (80%) committed only after they were asked the third prompt question (“Could you maybe do it?”; see Table 2 for details). The numbers of prompt questions did not differ significantly between 3-year-olds ($M = 3.00$, $SD = 0.33$) and 5-year-olds ($M = 2.75$, $SD = 0.85$), $U = 209$, $p = .401$ (Mann–Whitney *U* test). There was no significant correlation between children’s willingness to commit to the task (i.e., the number of prompt questions asked) and how committed they were to completing the task during the distraction phase (3-year-olds: $\tau_B = -0.23$, $p = .303$; 5-year-olds: $\tau_B = 0.08$, $p = .737$; Kendall correlation). No child in the promise group spontaneously used the word “versprechen” (promise)

Table 2

Numbers of children in the promise group in Study 2a who spontaneously committed to continue with the cleaning task in response to the sequence of prompt questions.

Age (years)	(0) No question	(1) What should we do?	(2) Who will clean?	(3) Could you do it?	(4) Do you promise?
3	0	0	1	17	1
5	1	1	1	16	1

Note. All children were eventually asked to promise. Questions are abbreviated.

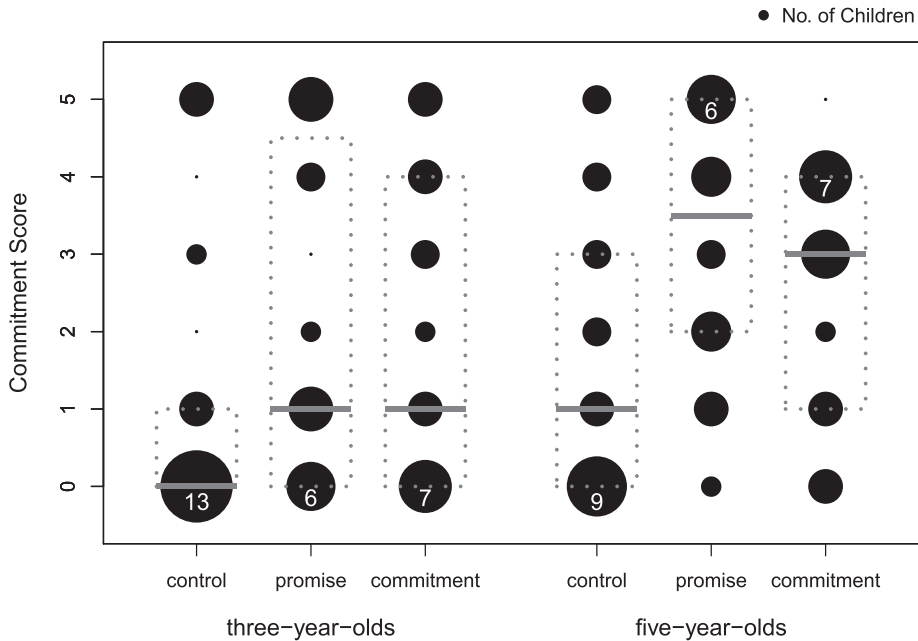


Fig. 4. The 3- and 5-year-olds' commitment scores in the control and promise groups in Study 2a and in the commitment group in Study 2b. Higher scores indicate greater commitment to the cleaning task. The numbers of children who reached the respective commitment scores are indicated by the areas of the bubbles (range = 0–13). The mode is indicated by the white numerals, the median is indicated by the solid gray line, and the second and third quartiles are indicated by the dotted gray boxes.

to commit to the cleaning task, but all children in this group were eventually asked by E1 to promise to complete the task. Children in the control group were never asked any prompt questions.

Next, we analyzed children's behavior during the distraction phase, finding that children in the promise group were more committed to continue the cleaning task than children in the control group (see Fig. 4). We analyzed children's commitment scores using ordinal linear regression models, finding that a full model had a significantly better fit to the data than a null model (including only the control predictor gender), $\chi^2 = 16.28$, $df = 3$, $p < .001$ (see [Supplementary Table 2](#) for details). Next, we dropped the nonsignificant interaction term between age group and experimental group from the full model, which did not change the model's fit significantly, $\chi^2 = 0.01$, $df = 1$, $p = .917$. The reduced model revealed that children in the promise group had significantly higher commitment scores than children in the control group, $Z = 3.39$, $p < .001$, and that 5-year-olds had significantly higher commitment scores than 3-year-olds, $Z = 2.03$, $p = .042$. In addition, there was a trend for girls to have higher commitment scores than boys, $Z = 1.82$, $p = .069$. Separate comparisons for each age group revealed a nonsignificant difference between the control and promise groups for 3-year-olds, $U = 126$, $p = .055$, and a significant difference for 5-year-olds, $U = 100$, $p = .005$ (Mann–Whitney U tests). Taken together, these findings indicate that promises resulted in higher levels of commitment to complete the task, with 5-year-olds being clearly more committed and 3-year-olds showing some signs in this direction.

During the distraction phase, some children spontaneously mentioned their commitment to the cleaning task. There were no significant age differences, $p > .999$ (Fisher exact test). In the promise group, 14 children (36%) mentioned their commitment to the task (see [Table 3](#) for further details). The majority of those children used normative language (e.g., “Aber ich muss das aufräumen” [But I have to clean here]; $n_{\text{normative}} = 11$), and 2 of those children explicitly referred to their promise. In contrast, significantly fewer children ($n_{\text{total}} = 6$ [15%] with $n_{\text{normative}} = 4$) mentioned their commitment in the control group, $p = .041$ (Fisher exact test). Overall, children's justifications revealed that more

Table 3

Numbers of children who spontaneously mentioned their commitment to the cleaning task during the distraction phase of Studies 2 and 3.

Study	Age (years)	Condition	Promise	Normative	Simple	Total
Study 2	3	Promise	1	5	0	19
		Control	0	1	1	20
		Commit w/out promise	0	2	2	19
	5	Promise	1	6	1	20
		Control	0	3	1	20
		Commit w/out promise	0	1	1	20
Study 3	3	Promise	0	8	2	20
		Control	0	2	1	20
	5	Promise	1	6	4	20
		Control	0	3	0	20

children in the promise group than in the control group viewed the cleaning task as a normative obligation.

To summarize, we found that 3- and 5-year-olds who were asked to promise to continue a tedious cleaning task stayed committed for longer and mentioned their commitment more frequently than children who were not asked to promise. Even though we were unable to elicit spontaneous promises from children, the adult experimenter's request of a promise ("Do you promise to clean?") induced a promissory commitment. Children's ability to keep elicited promises, thus, emerges earlier than previously found (e.g., Heyman et al., 2015). There are a number of possible explanations for why our findings differ from the findings of previous studies. First, it is conceivable that our cooperative helping task reduced incentives for young children to defect on their promise as compared with the competitive games or transgression situations of previous studies (Heyman et al., 2015; Lyon & Dorado, 2008). Second, irrespective of the task context, our series of distraction cues may have provided us with a more fine-grained measure of children's behavior than the dichotomous measure (e.g., cheat yes/no) used in other studies. Third, our study may have differed from previous studies in the extent to which young children needed to exercise self-control in the face of competing desires. Future studies could, for example, directly compare promise-keeping abilities in competitive versus cooperative contexts these different accounts.

One question that we were not able to address (because we failed to elicit spontaneous promises) is whether children distinguish between spontaneous and elicited promises. Research with adults has shown that promises that are uttered spontaneously in free-flowing communication are more effective in enhancing trust and cooperation than "bare promises" (i.e., sending a prewritten message with the words "I promise"; Charness & Dufwenberg, 2006, 2010). Moreover, developmental work on children's helping behavior has revealed that 5-year-olds (but not 3-year-olds) helped longer with a task when they were given a choice to help as compared with being asked to help (Rapp, Engelmann, Herrmann, & Tomasello, 2017). Together, these findings tentatively suggest that spontaneous promises may create an even stronger obligation for older preschoolers than elicited promises. This remains an interesting avenue for future research to investigate.

One developmental pattern emerged in our study: Irrespective of condition, 5-year-olds were more committed to the task than 3-year-olds. Similarly, Gräfenhain et al. (2009) found that 4-year-olds took longer to leave a (joint or parallel) game than 3-year-olds. In addition, studies on the development of children's self-control have shown that with increasing age children are better able to regulate their actions (e.g., stop repeating a previously successful action that is no longer effective) and attention (e.g., focus despite a distracting noise) (Herrmann, Misch, Hernandez-Lloreda, & Tomasello, 2015). Thus, older children are better at focusing on their current task/game and are less easily distracted to explore a more attractive activity, which may explain the developmental findings in our study. The possibility remains that these observed age differences are specific to the distracter task we used (i.e., playing with an attractive toy) and that a different task—for example, a commitment to overcome an obstacle or persist on a difficult puzzle—may reveal no age-related changes in overall commitment.

So far, we have always used the word *versprechen* (promising) when investigating preschoolers' understanding of promissory commitments. Although saying "I promise" or "Do you promise?" is usually the most unambiguous way of signaling a commitment, utterances such as "I will do it" and "Will you do this?" can fulfill the same function (Searle, 1969). In fact, developmental work has shown that 3- to 10-year-old French children view future-directed speech acts ("I will do it") and assertions with a predictive content ("It will be done") as promises (Bernicot & Laval, 1996). However, under some circumstances, such as when statements containing "promise" and "will" provide conflicting evidence (e.g., regarding the location of a hidden toy), older but not younger children trust "promise" statements more than "will" statements (Lyon & Evans, 2014). To further investigate whether the word "promise" is necessary to elicit task commitment in our study, we conducted a follow-up study with 3- and 5-year-olds where we elicited commitments without asking children to promise.

Study 2b

In this follow-up to Study 2a, we explored preschoolers' commitment to the cleaning task when commitments are elicited without the word "promise". We repeated the experiment from Study 2a but excluded the last prompt question and only asked children, "Could you maybe do it?" We chose this question because the majority of children had committed to the cleaning task after this prompt in Study 2a.

Method

Participants

In total, 19 3-year-olds ($M = 3;9$, range = 3;6–3;11; 10 girls and 9 boys) and 20 5-year-olds ($M = 5;9$, range = 5;4–5;11; 11 girls and 9 boys) took part in the study. An additional 2 3-year-olds were excluded because they were too shy and did not fully participate in the study. An additional 6 5-year-olds were excluded because they spontaneously committed to the cleaning task before the last prompting question was asked. Children were recruited in the same manner as in the previous studies and were tested in a quiet room in their local kindergarten.

Procedure

The procedure was identical to that in Study 2a with the exception that we did not ask children to promise to continue cleaning (commitment-without-promise group). Specifically, E1 asked all of the prompting questions up to "Could you maybe do it?" in quick succession (Question 3). Children who spontaneously committed to the task before the final prompt question were excluded from the study (6 5-year-olds). The study was conducted by the same two female experimenters as in Study 2a.

Data coding and analyses

Data coding and analyses and reliability coding were identical to those in Study 2a. We included the data from Study 2a in all statistical analyses. Results of the reliability analyses are reported in Study 2a ("Data coding and analyses" section).

Results and discussion

An overview of children's commitment scores can be found in Fig. 4. We analyzed children's commitment scores using ordinal linear regression models. A full model (with age group, experimental group, and their two-way interaction) had a significantly better fit to the data than a null model (including only the control predictor gender), $\chi^2 = 19.39$, $df = 5$, $p = .002$ (see Supplementary Table 3). Removing the nonsignificant two-way interaction between experimental group and age group did not significantly change the reduced model's fit to the data, $\chi^2 = 0.33$, $df = 2$, $p = .850$. The reduced model revealed that children in the current study had significantly higher commitment scores than children in the control group (Study 2a), $Z = 2.60$, $p = .009$, but commitment scores did not differ significantly from those of children in the promise group (Study 2a), $Z = 1.20$, $p = .232$. We also found that

5-year-olds had significantly higher commitment scores than 3-year-olds, $Z = 2.19$, $p = .029$, and girls had significantly higher commitment scores than boys, $Z = 2.57$, $p = .010$. Thus, eliciting a commitment from children without using the word “promise” had a similar effect on young children’s behavior as eliciting a promise.

Moreover, we coded whether children spontaneously mentioned their commitment and found that only 6 children (15%) did so in the current study (see Table 3). These were fewer children than in the promise group (Study 2a), $p = .068$, and a similar number of children compared with the control group (Study 2a), $p > .999$. It is possible that the omission of the word “promise” made the normative dimension of the commitment less salient for children.

Our findings suggested that children are committed to the task even when the word “promise” is omitted. This is in line with previous work on French-speaking children’s understanding of promises and future-directed commissive speech acts (“I will do it”; Bernicot & Laval, 1996), and it supports theoretical work arguing that using the word “promise” is not a necessary condition for an utterance to count as a verbal commitment (Searle, 1969).

In Studies 2a and 2b, children were always asked a series of prompt questions to elicit a commitment. We had introduced the prompt questions to evoke spontaneous promises from children, but the possibility remains that the questions simply reminded children of the cleaning task and, thus, resulted in higher commitment compared with the control group. To control for this alternative explanation of our findings, we conducted a third study with 3- and 5-year-olds in which we contrasted a promise to clean (without repeated prompts) with a cleaning reminder (see Heyman et al., 2015, for a similar control).

Study 3

In this study, we investigated 3- and 5-year-olds’ promissory commitment to help in a collaborative cleaning task, contrasting a promise to clean (promise group) with a cleaning reminder (control group). We expected that children in the promise group would be more committed to the task than children who only received a cleaning reminder.

Method

Participants

In total, 40 3-year-olds ($M = 3;8$, range = 3;5–3;10; 21 girls and 19 boys) and 40 5-year-olds ($M = 5;8$, range = 5;4–5;11 years; 20 girls and 20 boys) took part in the study. An additional 5 3-year-olds were excluded because they did not want to participate (3 children) or to promise (2 children). One additional 5-year-old was excluded for needing to use the bathroom during the study. Children were recruited in the same manner as in the previous studies and were tested in a quiet room in their local kindergarten.

Procedure

The procedure was identical to that in Study 2 with the following exceptions:

1. In the promise group, children were directly asked to promise to clean without using the series of prompt questions (“Versprichst du mir, dass du hier aufräumst?” [Do you promise to clean here?]). In the control group, they were reminded that there was some cleaning left to do (“Es gibt hier noch was zum Aufräumen” [There is something to clean here]).
2. We used a different marble track (a wooden one; see Supplementary Fig. 1) and a new team of experimenters (E1 was male and E2 was female).

Data coding and analyses

Data coding and analyses and reliability coding were identical to those in Study 2. Some children spontaneously made generic normative statements during the distraction phase (e.g., “Der Teppich

muss ganz sauber sein" [The carpet needs to be completely clean]). We did not score these utterances in the normative cleaning category because they did not refer to the child's own cleaning obligation.

A second coder scored 25% of the data ($n = 20$ children) for reliability purposes (equally distributed by age group, experimental group, and gender). There was good agreement between observers for children's commitment scores during the distraction phase (equal-weighted $\kappa = .86$) and for children's spontaneous mentioning of their commitment (equal-weighted $\kappa = .95$).

Results and discussion

When we compared a direct promise with a cleaning reminder (control), we found that children in the promise group had higher commitment scores than children in the control group (see Fig. 5). Ordinal linear regression models revealed that a full model had a significantly better fit to the data than a null model (including only the control predictor gender), $\chi^2 = 14.47$, $df = 3$, $p = .002$ (see Supplementary Table 4 for details). Next, we dropped the nonsignificant age group and experimental group interaction from the full model, which did not change the model's fit significantly, $\chi^2 = 0.06$, $df = 1$, $p = .800$. The reduced model revealed that children in the promise group had significantly higher commitment scores than children in the control group, $Z = 3.39$, $p < .001$. There was a trend for 5-year-olds to have higher commitment scores than 3-year-olds, $Z = 1.66$, $p = .097$. Separate analyses for the two age groups showed a significant difference between the promise group and the control group for both 3-year-olds, $U = 130$, $p = .048$, and 5-year-olds, $U = 107$, $p = .007$ (Mann-Whitney U tests). These findings confirm that from 3 years of age children's commitment to continue the cleaning task increases after an elicited promise and that older children overall persist longer on the task than younger children.

Children spontaneously mentioned their commitment to the task during the distraction phase. There was no significant age difference, $p > .999$ (Fisher exact test). In the promise group, 21 children (53%) mentioned their commitment to the task (see Table 3 for further details). The majority of those children used normative language ($n_{\text{normative}} = 14$), and 1 child explicitly referred to the promise. In

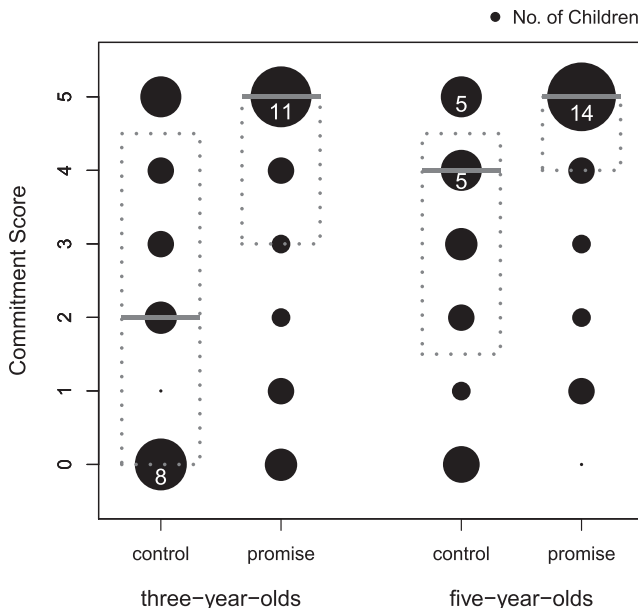


Fig. 5. The 3- and 5-year-olds' commitment scores in the control group (cleaning reminder) and the promise group in Study 3. Higher scores indicate greater commitment to the cleaning task. The numbers of children who reached the respective commitment scores are indicated by the areas of the bubbles (range = 0–14). The mode is indicated by the white numerals, the median is indicated by the solid gray line, and the second and third quartiles are indicated by the dotted gray boxes.

contrast, significantly fewer children ($n_{\text{total}} = 6$ [15%], of which $n_{\text{normative}} = 5$) mentioned their commitment in the control group, $p < .001$ (Fisher exact test).

These results support and extend our previous findings. From 3 years of age, children were more committed to a collaborative cleaning task after an elicited promise than after a cleaning reminder. This supports our claim that children's higher commitment scores in the promise and commitment groups in Study 2 were a result of their verbal commitments and not simply due to the prompt questions reminding children of the cleaning task. Moreover, our results extend to a younger age group the findings of Heyman et al. (2015), who showed that 5-year-olds refrained more often from cheating on a card game after an elicited promise than after a rule reminder. Commitment scores in this study were overall higher than those in Study 2, which may be due to variations in procedural details (e.g., wording, materials, experimental team); importantly, however, significant differences in commitment between the control and promise groups were replicated.

It is possible that the control condition in this study—reminding children that there was something to clean—as well as the control condition in Study 2a created some uncertainty on the part of children about what behavior was expected of them. For example, children may have assumed that the adult experimenter expected them to continue with the cleaning task. Indeed, previous work has shown that collaborative or joint activities induce commitments in young children (Butler & Walton, 2013; Hamann et al., 2012). Thus, it is likely that children in the control conditions in Studies 2a and 3 felt some commitment to the joint cleaning tasks. Importantly, however, our findings in both studies clearly show that promises or explicit verbal commitments create normative obligations that are stronger than the (implicit) commitments elicited by collaborative activities.

General discussion

We investigated German preschoolers' understanding of others' promissory obligations and their ability to keep their own promises, and we demonstrated across three studies that preschoolers understand that promises entail a normative obligation to fulfill the promised act. Specifically, they protested and referred to promise norms when a puppet partner broke a sharing promise (Study 1) and helped an adult longer in a collaborative cleaning task after an elicited promise or verbal commitment (Studies 2 and 3). Our findings support the view that children begin to comprehend promissory obligations in social interactions at around 3 years of age, an earlier age than what is documented in the literature (e.g., Astington, 1988b; Heyman et al., 2015). Furthermore, they suggest that promises in prosocial or collaborative situations may help young children internalize a sense of obligation to helping others and that these promises have important normative implications.

We propose that children's early understanding of promises develops through their participation in joint social activities. Their understanding of promises is preceded by their understanding of joint intentions, actions, and goals that emerges during the first 18 months of life (Bruner, 1975; Carpenter, 2009; Tomasello, Carpenter, Call, Behne, & Moll, 2005). By 3 years of age, preschoolers who are engaged in joint activities have been found to acknowledge their leaving of a task to their partner, to help their partner achieve her or his goal, and to persist longer in a challenging task, indicating a sense of commitment to the joint endeavor (Butler & Walton, 2013; Gräfenhain et al., 2009, 2013; Hamann et al., 2012; Warneken, Gräfenhain, & Tomasello, 2012). In addition, our findings show that when promises are uttered during joint activities, young children are committed to their own promises and expect that others ought to keep theirs.

Interestingly, preschoolers in our studies participated in joint activities in both the control condition and the promise condition; they jointly retrieved rewards with a puppet or cleaned a carpet with an adult. Based on the findings cited above, one would expect that in the control condition children may have developed a sense of commitment to the joint endeavor. However, when promises were involved in addition to joint activities, children referred more frequently to promise norms or normative obligations, indicating an understanding of the normative implications of promising (beyond a mere sense of commitment). Thus, whereas joint activities may lead to expectations about one's own and others' behavior, such as expecting someone to play her or his part in a joint activity and to contribute to the joint goal (Michael, Sebanz, & Knoblich, 2015), promises possess a strong

normative force (i.e., that one *ought* to keep them). That is, promises or explicit verbal commitments create normative obligations that are stronger than the implicit commitments elicited by joint activities.

Arguably, we investigated children's understanding of promises only in interpersonal contexts and did not study whether preschoolers will refer to or enforce promise norms as uninvolved third parties. The later behavior is one indication that children understand that norms apply not only to themselves but also to other members of their group. There is a growing body of developmental work showing that from 3 years of age Western children actively enforce social and moral norms toward third parties in a variety of contexts such as game rules and ownership norms (Rakoczy & Schmidt, 2013; Rakoczy et al., 2008; Rossano et al., 2011; Tomasello & Vaish, 2013). This suggests that preschoolers understand in principle that norms apply to all individuals within similar contexts. Furthermore, we take the fact that preschoolers in the promise conditions in our studies used normative language (e.g., "I have to clean here," "You shouldn't do this") and not just statements of personal liking or disliking (e.g., "I want to clean," "I want more stickers") as evidence of a normative understanding of promises.

That promises and verbal commitments possess normative force is remarkable—or as Prichard (2002, p. 257) stated, "Once call some act a promise and all question of whether there is an obligation to do it seems to have vanished." There are two mechanisms that may explain why people keep their word: external sanctions and internalized norms. The external sanctions account is dismissed by scholars on the grounds that many studies have shown that adults keep their promises in the absence of third-party sanctions (Orbell et al., 1988; Ostrom et al., 1992); instead, different psychological mechanisms have been suggested to explain this behavior such as an aversion to letting others down (Charness & Dufwenberg, 2006) or a preference for keeping one's word (Ellingsen & Johannesson, 2004). Our findings that young children already reference norms and obligations (and not adult authority or fear of punishment) support the psychological view and suggest that promise norms begin to be internalized early in development—although our data do not allow us to differentiate between different mechanistic accounts (e.g., disappointing others, preference for word keeping, fear of others' disapproval).

The finding that young children do already understand some normative implications of promising does not imply that there are no developmental changes in promise-related behaviors or that children always keep their promises. For one, we found that across conditions older children were more committed to our cleaning task (Studies 2 and 3), which is most likely related to developmental changes in children's abilities for self-control (Herrmann et al., 2015). This explanation has been championed by other authors (Heyman et al., 2015) to explain why younger children in their study were not affected by promises. We suggest that in addition to self-control, contextual factors (e.g., cooperative vs. competitive situations, task demands) influence whether young children avoid the temptation to renege on their promises. Furthermore, with age children's reasoning about promises becomes more sophisticated, and they increasingly consider intentions when judging whether it is okay for others to break a promise or not (Maas & Abbeduto, 2001; Mant & Perner, 1988). One may also expect that older children's decisions to break their own promises could be influenced by whether the children believe that they have a good reason to do so (e.g., it is okay to break a promise in order to help someone in immediate need, but it is not okay to break a promise in order to play a fun game). Even in the absence of good reasons for breaking a promise, children (and adults) may sometimes break promises to enhance their self-interest or the interests of those close to them.

One final caveat is that most of the work on promises and norms mentioned in this article (including our own) focuses on children and adults in large-scale Western societies. One cross-cultural study on promises has found that expectations about when it is obligatory to comply with promises and preference for how to communicate promises vary across German, Chinese, and Tongan participants (Beller, Bender, & Song, 2009). Moreover, past work has documented cross-cultural variation in whether children prioritize general normative principles such as justice (U.S. children) or personal obligations (Indian children) when judging moral dilemmas (Miller & Bersoff, 1992). Thus, it is plausible that similar cultural variation may exist with regard to whether promises are understood as societal norms or (inter)personal obligations. Future work is needed to better understand promise-related practices and beliefs in different societies and how these shape children's developing understanding of promises.

Acknowledgments

We thank Carl Bartl, Liane Doerr, Valentina Dondoli, Isabelle Gaillande-Mustoe, Jana Jurkat, Esther Klonowski, Solveig Jurkat, Julia Lang, Eva Siegert, Nadine Schumann, Benjamin Sklarek, Anne Tomm, and Gesa Volland for their help in running the study and transcribing and coding the data. Thanks also go to Magherita Isella for helpful comments on earlier versions of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jecp.2017.02.004>.

References

- Astington, J. (1988a). Children's production of commissive speech acts. *Journal of Child Language*, *15*, 411–423.
- Astington, J. (1988b). Children's understanding of the speech act of promising. *Journal of Child Language*, *15*, 157–173.
- Astington, J. (1988c). Promises: Words or deeds? *First Language*, *8*, 259–270.
- Austin, J. L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Beller, S., Bender, A., & Song, J. (2009). Conditional promises and threats in Germany, China, and Tonga: Cognition and emotion. *Journal of Cognition and Culture*, *9*, 115–139.
- Bernicot, J., & Laval, V. (1996). Promises in French children: Comprehension and metapragmatic knowledge. *Journal of Pragmatics*, *25*, 101–122.
- Bicchieri, C. (2002). Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society*, *14*, 192–228.
- Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*, 1–19.
- Butler, L. P., & Walton, G. M. (2013). The opportunity to collaborate increases preschoolers' motivation for challenging tasks. *Journal of Experimental Child Psychology*, *116*, 953–961.
- Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science*, *1*, 380–392.
- Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*, 1579–1601.
- Charness, G., & Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, *107*, 281–283.
- Christensen, R. H. B. (2015). *Ordinal—Regression models for ordinal data*. Retrieved from <<http://cran.r-project.org/package=ordinal/>>.
- Diessel, H. (2004). *Acquisition of complex sentences*. Cambridge, UK: Cambridge University Press.
- Ellingsen, T., & Johannesson, M. (2004). Promises, threats, and fairness. *The Economic Journal*, *114*, 397–420.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *irr: Various coefficients of interrater reliability and agreement*. Retrieved from <<http://cran.r-project.org/package=irr/>>.
- 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*, 1430–1443.
- Gräfenhain, M., Carpenter, M., & Tomasello, M. (2013). Three-year-olds' understanding of the consequences of joint commitments. *PLoS One*, *8*(9), e73039.
- Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*, 328–331.
- Hamann, K., Warneken, F., & Tomasello, M. (2012). Children's developing commitments to joint goals. *Child Development*, *83*, 137–145.
- Herrmann, E., Misch, A., Hernandez-Lloreda, V., & Tomasello, M. (2015). Uniquely human self-control begins at school age. *Developmental Science*, *18*, 979–993.
- 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.
- Hume, D. (1890). *A treatise of human nature*. London: Longmans, Green.
- Kalish, C., & Cornelius, R. (2007). What is to be done? Children's ascriptions of conventional obligations. *Child Development*, *78*, 859–878.
- Lewis, D. (2002). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press/Blackwell. (Original work published 1969).
- Lohse, K., Gräfenhain, M., Behne, T., & Rakoczy, H. (2014). Young children understand the normative implications of future-directed speech acts. *PLoS One*, *9*(1), e86958.
- 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*, 738–748.
- Lyon, T. D., & Evans, A. D. (2014). Young children's understanding that promising guarantees performance: The effects of age and maltreatment. *Law and Human Behavior*, *38*, 162–170.
- Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, *79*, 914–929.

- Maas, F., & Abbeduto, L. (1998). Young children's understanding of promising: Methodological considerations. *Journal of Child Language*, 25, 203–214.
- Maas, F., & Abbeduto, L. (2001). Children's judgments about intentionally and unintentionally broken promises. *Journal of Child Language*, 28, 517–529.
- Mant, C., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology*, 24, 343–351.
- McLeod, A. I. (2011). *Kendall: Kendall rank correlation and Mann-Kendall trend test*. Retrieved from <<https://cran.r-project.org/web/packages/Kendall/index.html>>.
- Michael, J., Sebanz, N., & Knoblich, G. (2015). The sense of commitment: A minimal approach. *Frontiers in Psychology*, 6. <http://dx.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, 541–554.
- Orbell, J. M., Van de Kragt, A. J., & Dawes, R. M. (1988). Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology*, 54, 811–819.
- Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86, 404–417.
- Prichard, H. A. (2002). The obligation to keep a promise. In J. MacAdam (Ed.), *Moral writings* (pp. 257–265). Oxford, UK: Clarendon.
- R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <<http://www.R-project.org>>.
- Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, 7, 17–21.
- Rakoczy, H., & Tomasello, M. (2009). Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition*, 113, 205–212.
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, 44, 875–881.
- Rapp, D. J., Engelmann, J. M., Herrmann, E., & Tomasello, M. (2017). The impact of choice on young children's prosocial motivation. *Journal of Experimental Child Psychology*. Advance online publication. doi:<http://dx.doi.org/10.1016/j.jecp.2017.01.004>.
- Rawls, J. (1955). Two concepts of rules. *Philosophical Review*, 64, 3–32.
- Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition*, 121, 219–227.
- Rotenberg, K. J. (1980). "A promise kept, a promise broken": Developmental bases of trust. *Child Development*, 51, 614–617.
- Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, 7, 58–92.
- Scanlon, T. (1990). Promises and practices. *Philosophy & Public Affairs*, 19, 199–226.
- Schmidt, M. F. H., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21, 232–236.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.
- Smetana, J. G. (2013). Moral development: The social domain theory view. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology. Body and mind* (1, pp. 832–864). Oxford, UK: Oxford University Press.
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior*, 26, 395–415.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, 64, 231–255.
- Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica*, 76, 1467–1480.
- Warneken, F., Gräfenhain, M., & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children's understanding of joint intentions in collaborative activities. *Developmental Science*, 15, 54–61.
- Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science*, 22, 267–273.