The *Pizzagame*: A virtual public goods game to assess cooperative behavior in children and adolescents.

Jan Keil[1], Andrea Michel[1,2], Fabio Sticca[3], Kristina Leipold[4], Annette M. Klein[1], Susan Sierau[1], Kai von Klitzing[1], and Lars O. White[1]

[1]*Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Leipzig*

[2] *Department of Developmental Psychology, Friedrich-Schiller-University of Jena*

[3] D*epartment of Empirical Educational Research, University of Konstanz/Thurgau University of Teacher Education*

[4]*Faculty of Business and Economics, Dresden University of Technology*

Corresponding author:

Jan Keil

Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Leipzig

Liebigstraße 20a

04103, Leipzig

E-Mail: jan.keil@medizin.uni-leipzig.de

Original article: Keil, J. et al. (2017). The Pizzagame: A virtual public goods game to assess cooperative behavior in children and adolescents. *Behavior Research Methods*, *49*(4), 1432-1443.

Abstract

Social dilemmas are characterized by conflicts between immediate self-interest and long-term collective goals. Although such conflicts lie at the heart of various challenging social interactions, we know little about how cooperation in these situations develops. To extend work on social dilemmas to child and adolescent samples, we developed an age appropriate computer task (the *Pizzagame*) with structural features of a public goods game (PGG). We administered the *Pizzagame* to a sample of 191 children aged 9 to 16. Participants were led to believe they were playing the game over the Internet with three sets of two same-aged, same-sex co-players. In fact, co-players were computer-generated and programmed to expose children to three consecutive conditions: (1) a cooperative strategy, (2) a selfish strategy, and (3) a divergent cooperative-selfish strategy. Supporting the validity of the *Pizzagame*, results revealed that children and adolescents displayed conditional cooperation, such that their contributions rose with increasing cooperativeness of their co-players. Moreover, while increasing age did not influence children and adolescents' cooperative behavior within each condition, older children adapted their behavior more substantially to parallel the strategies of their co-players across the different conditions. These results support the utility of the *Pizzagame* as a feasible, reliable and valid instrument for assessing and quantifying child and adolescent cooperative behavior. Moreover, these findings extend previous work showing that age influences cooperative behavior in the PGG.

Keywords: public goods game, cooperation, prosocial behavior, computerized task, child development

Original article: Keil, J. et al. (2017). The Pizzagame: A virtual public goods game to assess cooperative behavior in children and adolescents. *Behavior Research Methods*, *49*(4), 1432-1443.

Introduction

We can choose to pay our taxes, organize a reunion with our classmates, prepare a meal with the family, clean up with friends after a party, but we can also just rely on others to do the work and reap the benefits from their efforts without contributing. Such choices make up the fabric of our daily interpersonal interactions, many of which might be construed as social dilemmas. In these situations immediate self-interest runs contrary to potential long term collective gains (Van Lange, Joireman, Parks, & van Dijk, 2013). Interestingly, humans quite successfully manage to cooperate in such situations, which has enabled many of the accomplishments of modern societies like government, health care, or accessible education (De Dreu, 2013).

Numerous theories state reasons why and when cooperation emerges in social dilemmas (for an overview seeParks, Joireman, & Van Lange, 2013). However, despite substantial work on adults, we know relatively little about how the propensity to cooperate in these situations develops. Additionally, the few existing studies with children have produced inconsistent empirical results and contain several methodological drawbacks. Therefore, with the help of a novel, computerized, and developmentally appropriate instrument we aimed to overcome some of these drawbacks and promote our understanding of how cooperation develops in children and adolescents.

The Public Goods Game

A classic example of a social dilemma is the Public Goods Game (PGG; Hardin, 1968). Public goods refer to resources available to, and consumable by all group members, irrespective of how much an individual contributes to their provision (Olson, 1971 (1965)), such as a clean environment or public services (Gummerum, Hanoch, & Keller, 2008). Groups achieve their social optimum if everyone chooses to contribute to the public good and thus "pulls his/her weight", but the individual profits most by choosing a selfish strategy (i.e.,

Original article: Keil, J. et al. (2017). The Pizzagame: A virtual public goods game to assess cooperative behavior in children and adolescents. *Behavior Research Methods*, *49*(4), 1432-1443.

freeriding) and exploiting others (Dawes, 1980). Due to this conflict, inherent in ample social interactions, these social dilemmas are especially appealing for studying cooperation in groups (Parks et al., 2013).

In the corresponding laboratory situation of a PGG players receive an initial endowment of a resource and must then decide how much they want to keep for themselves or pool toward a public good. The latter is subsequently multiplied by a set factor and equally redistributed among all group members. Importantly, paralleling other social dilemmas, the outcome for an individual in this situation does not only depend on what he/she does but also on what everyone else does (strategic interdependency, Gross & Heinrichs, 2010).

Although standard rational choice assumptions suggest that individuals would opt for the selfish strategy, various experiments demonstrated that individuals behave much more cooperatively (Colman, 2003). Thus, they contribute a considerable amount in the first round (40-60% of their resources) while contributions steadily dwindle thereafter. Freeriding paired with the tendency for conditional cooperation (cooperate if others also cooperate) is thought to partly account for this pattern (Fischbacher, Gächter, & Fehr, 2001).

Age and gender differences

Despite extensive experimental research with adults in PGG’s (for reviews see Chaudhuri, 2011; Ledyard, 1995; Zelmer, 2003), few studies to date have examined child or adolescent cooperative behavior in these situations. Previous work shows that 5-year-olds playing a simplified PGG begin to adopt conditionally cooperative strategies (Vogelsang, Jensen, Kirschner, Tennie, & Tomasello, 2014). Moreover, an experimental study on 5- to 12-year-olds suggests that they initially contribute about the same share as adults, but then increase contributions which subsequently plateau and finally decline (Harbaugh & Krause, 2000). This curvilinear trajectory diverges from the continuous decrease in adults. However, inconsistent results and methodological drawbacks limit the conclusions we can draw from previous developmental work in the following ways:

First, studies analyzing the influence of age on cooperative behavior yield inconsistent results. While two studies reported that older children show more cooperative behavior by contributing more to the public good (Fan, 2000; Sally & Hill, 2006), a study by Cipriani, Giuliano, and Jeanne (2007) found that younger children contributed more than older ones. At the same time, other work shows that while older children initially contribute more, they also decrease contributions toward freeriding more readily than younger children (Harbaugh & Krause, 2000). Yet, this work cannot clarify whether older children freeride more readily regardless of other players' strategies or, alternatively, whether they freeride mainly when others freeride, but would also cooperate more when others cooperate (conditional cooperation).

Second, equivocal findings also abound in relation to gender. The literature on adults reports gender differences in opposite directions. While some studies show that cooperation in PGGs is more characteristic of females, others identified males as more cooperative (Croson & Gneezy, 2009). Similar inconsistencies exist for children. Whereas Harbaugh and Krause (2000) found no significant gender effects on cooperative behavior over the course of the game, in the study of Vogelsang, Jensen, Kirschner, Tennie, and Tomasello (2014) boys freerode more than girls. In a study of Cipriani, Giuliano, and Jeanne (2007) girls also tended to show more cooperative behavior.

In part, these inconsistencies may be attributable to a lack of systematic control for strategic interdependence. In most work on children, experimenters group multiple subjects together whose outcomes thus depend on each other's choices. To be sure, this approach may benefit ecological validity by exposing subjects to "real-life" strategies of other co-players. Yet, greater ecological validity often entails some sacrifice of experimental control and causal inference. For example, the behavior of a child interacting with cooperative co-players is difficult to compare to the behavior of another child interacting with selfish co-players. Although some work with children has employed computer-generated co-players (Leipold, Vetter, Dittrich, Lehmann-Waffenschmidt, & Kliegel, 2013; McClure et al., 2007; Sally & Hill, 2006), these studies either implemented responsive algorithms that tethered co-players' contributions to the previous decision of the subject (McClure et al., 2007; Sally & Hill, 2006) or divided subjects into different experimental groups, with each subgroup facing different strategies (Leipold et al., 2013). Thus, little or no work uses experimental designs where each subject faces identical strategies of other co-players.

Additionally, abstractness of most experimental designs (e.g., playing for money or tokens) arguably makes it harder for children to understand the strategic features of the situation, rendering cognitive development a potential confound. Of the few studies that attempt to translate the PGG to a more concrete child-appropriate context (Alencar, De Oliveira Siqueira, & Yamamoto, 2008; Vogelsang et al., 2014), set-ups have proven resource-intensive and none have adopted a computerized methodology which vastly simplifies data-collection.

Aims and Hypotheses

To address some of the limitations of previous work, the present study introduces a newly developed, age-appropriate computer task based on a concrete real-life situation, called the *Pizzagame*. We report data collected with this task in a sample of children and adolescents aged 9 to 16. In the *Pizzagame* children are led to believe they are connected over the Internet with three different sets of two same-sex co-players. All of the players receive a fixed set of resources (i.e., slices of pizza) which they can decide to pool towards the public good (i.e., take to school) or not (i.e., leave at home). At school, the "virtual teacher" adds 50% of all pooled slices which are then equally redistributed among all players.

In total, the *Pizzagame* progresses through three conditions in a predetermined sequence, each condition consisting of four rounds each. In each condition, the co-players, who are in fact computer-generated, pursue fixed strategies. During the first condition, subjects face cooperative co-players who contribute high quantities of their resource to the public good. In the second condition, they interact with selfish co-players who only contribute very little, while, in the final third condition, co-players strategies diverge, with one co-player exhibiting a cooperative and the other exhibiting a selfish strategy.

The present study pursued three main aims. The first and primary aim was to introduce and show the feasibility and reliability of a new life-like PGG that controls for the factor of strategic interdependency to assess cooperation and defection among children and adolescents. However, a mere description of the *Pizzagame* along with the claim that it is feasible and reliable would have begged the question as to whether the task is in fact a valid measure of cooperative behavior for children and adolescents.

As a second aim, we therefore sought to demonstrate the validity of this new measure. Given that prior findings indicate that individuals predominantly adopt a conditionally cooperative strategy, we expected that subjects would be cooperative toward cooperative, selfish toward selfish and show a medium level of contributions toward divergent co-players (Hypothesis 1).

Third, we aimed to shed further light on the role of age and gender effects in cooperative behavior to enrich our understanding of the developmental roots of cooperative behavior in children and adolescents. Therefore, we examined whether age and gender had a significant impact on two dimensions within the PGG, namely on contributions in each of the three different conditions and on behavioral change between conditions. Hence, as most studies report that older children show more cooperative behavior than younger children, we predicted that older children would contribute more across all three conditions (Hypothesis 2a). At the same time, based on the finding that freeriding spreads more readily among older children (Harbaugh & Krause, 2000), we predicted that older children would adapt more readily to the strategies of their co-players. That is, with increasing age children would show more pronounced lowering of contributions towards selfish co-players (compared to cooperative co-players), but also more pronounced elevation of contributions towards the divergent co-players, who were moderately cooperative (Hypothesis 2b). With respect to possible gender effects on cooperative behavior it is not clear if they exist, but if they do, girls seem to be more cooperative than boys. Thus, we predicted that girls would show more cooperative behavior across all three conditions (Hypothesis 3). Regarding gender effects on behavioral change between conditions there is no empirical evidence we could derive a proper hypothesis from. Thus, we will explore the impact of gender on behavioral change between conditions.

Methods

Sample

We recruited 216 children and adolescents aged 9 to 16 years as part of a general population sample of an ongoing large-scale study in a medium-sized German city (for detailed information see White et al., 2015). Institutional review board (IRB) approval was obtained from the university ethics committee. Parents or legal guardians consented and youth assented after being informed about the study prior to participation.

To rule out that children misunderstood the strategic set-up of the PGG we asked comprehension questions (see below) during the training phase of the procedure. Accordingly, in our analyses, we excluded 23 subjects because they erred on two or more out of nine comprehension questions. Finally, data from 2 subjects were not saved due to a technical error, yielding a final sample of 191 subjects (57.1 % girls, M_{age} = 12.03 years, *SD* = 1.92). With the exception of the 16-year-olds, age was spread relatively evenly across the full age-range (see Figure 1).

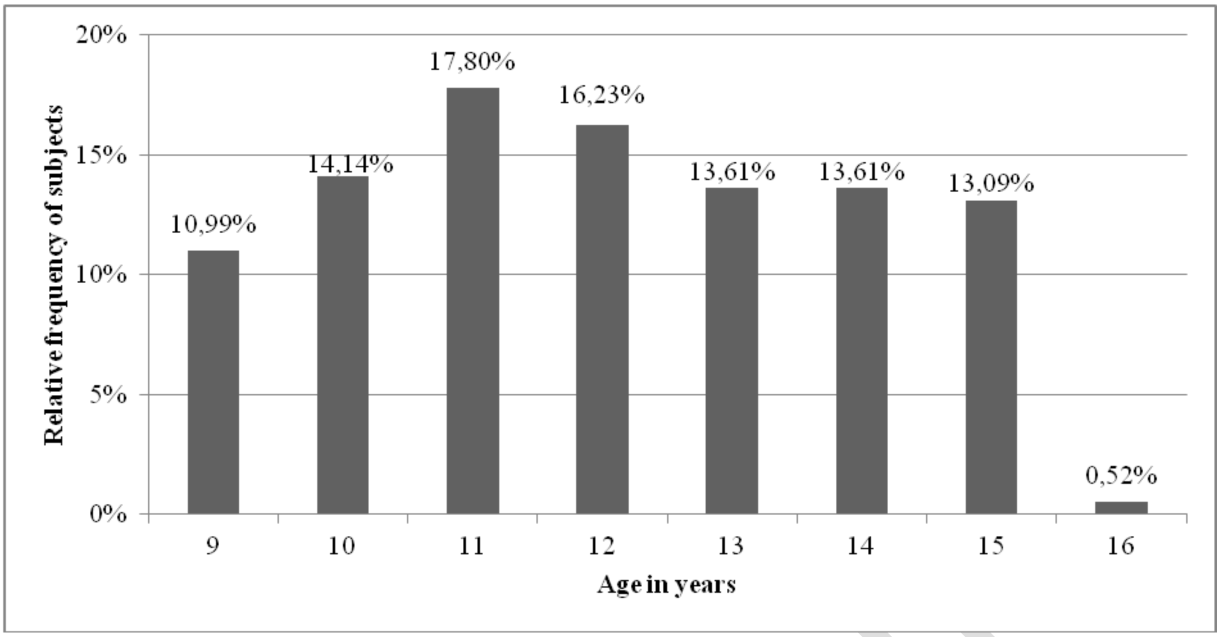

Figure 1 Age distribution of participants

There were no significant age differences between girls and boys ($t_{gender}(189)= 0.03$, $p = .97$). Additionally, the parental education and monthly household income did not differ as a function of age (parent education: $r_{age}(179) = -.07$, $p = .33$; monthly household income: $r_{age}(181) = .001$, $p = .99$) or gender (parent education: $t_{gender}(179)= 0.20$, $p = .84$; monthly household income: $t_{gender}(182)= -0.22$, $p = .83$).

Instructions and set-up

As part of the large-scale study, children were invited for one appointment that lasted approximately three hours. They received a battery of measures (e.g., storytelling task, verbal skills test, several questionnaires and interviews). The *Pizzagame* was the penultimate procedure of the appointment. Before starting the *Pizzagame* participants received thorough information about the rules and set-up (i.e. number of trials, number of players etc.) of the game via a slide show. They were also informed that the value of the gift they could choose at the end of the appointment would increase with the number of slices of virtual pizza they managed to retrieve throughout the course of the game. Such incentivization is a common feature of economic games to ensure a basic degree of motivation among subjects. Specifically, we used three boxes of different sizes (small, medium, big) each of which contained a different set of gifts. We informed children that the biggest box contained the most attractive presents, the medium box contained moderately attractive presents, and the smallest box contained the least attractive presents. We also told them that the more slices they collected during the game, the bigger the box would be that they could choose a gift from (incentivization). This procedure aimed to prevent specific gift preferences from affecting results. Unbeknownst to participants, everyone was offered the same selection of presents from the big box to not place anyone at a disadvantage due to their game behavior.

Instructions were followed by three illustrative scenarios (i.e., non-cooperative, exploitative, and cooperative) including different potential outcomes of the game. Multiple scenarios were used to safeguard against biasing children in their decision-making. To check subjects' comprehension of the strategic configuration of the PGG, they were asked the following questions regarding each scenario: "Which players have more pizza slices than at the start of the round?" "Which players have the same number of pizza slices as at the start of the round?" and "Which players have fewer pizza slices than at the start of the round?"

Afterwards, the experimenters ran a test version of the game, to explain and familiarize participants with the game interface. To further enhance the cover story, subjects were asked if they, just like their co-players, would be comfortable with a picture being taken of them via the webcam. In the absence of their child, parents were informed about the deception used in the PGG, and consented to it and all parts of the game before the procedure was started (see Discussion).

After starting the game, the experimenter claimed to have something else to do, took a seat at another table, and asked the child to continue playing the game. This aimed to minimize socially desirable response patterns due to the presence of the experimenter. The appointment was videotaped to check and ensure a high level of standardization throughout the period of data collection. Participant choices were recorded directly by E-Prime software suite (Schneider & Zuccoloto, 2007). After the *Pizzagame,* subjects evaluated the appointment including an open question asking what part of the appointment they liked best (the question allowed children to name more than one measure as their favorite).

Design of the *Pizzagame*

The *Pizzagame* implements the structural features of a PGG. In the course of developing the *Pizzagame*, we aimed to find a scenario that was as close to a life-like situation as possible. In line with other peer-based paradigms (Crowley, Wu, Molfese, & Mayes, 2010;Gunther Moor, Bos, Crone, & van der Molen, 2014;Guyer, Choate, Pine, & Nelson, 2012;Reijntjes, Stegge, Terwogt, Kamphuis, & Telch, 2006), we presumed that a situation where boys and girls ostensibly interact with same-age, same-sex peers in a school setting would act as a familiar and ecologically valid cue to trigger children's everyday social behavior. The importance of these aspects is cast into relief by ample data showing that concrete, familiar, and relevant scenarios improve performance on a variety of cognitive tasks even among adults (e.g., Wason & Shapiro, 1971;Sperber & Girotto, 2002) and facilitate earlier understanding among children (e.g., Donaldson, 1978;Doherty, 2009). Moreover, the use of pictures of other players, sound features, and background music for transitional slides, builds on previous adaptations of popular paradigms for children (e.g.,Crowley et al., 2010) and were designed to intuitively appeal to children and adolescents. The *Pizzagame* was programmed and presented using E-Prime® software suite (Schneider, Eschman, & Zuccolotto, 2002). Before starting the game, children were informed about the rules of the game and led to believe they were playing with two other children over the Internet by incorporating a fake website link. Actually, they played against computer-generated co-players with fixed strategies. We used this procedure to enhance the deception in light of empirical evidence showing that people behave differently when they know they are interacting with computer agents compared to humans (Krach et al., 2008; Shechtman & Horowitz, 2003). This deception procedure draws on previous work using a similar strategy for the well-validated ball-game Cyberball (Crowley et al., 2010). To enhance the credibility of the cover story and to minimize the impact of subjects' inferences based on different facial expressions of co-players, we used pictures from two emotional faces databases of children with facial expressions confirmed as neutral (Egger et al., 2011; Langner et al., 2010). To consolidate the impression that subjects were playing with three sets of two same-age, same-sex co-players we used facial portraits of boys or girls aged 9 to 12 for the younger subjects and pictures of boys or girls aged 13 to 16 for the older subjects.

Each of the four rounds of the different conditions of the *Pizzagame* begins by endowing the three players with nine virtual pizza slices. Without learning of one another's decisions, subjects (i for the participant, j and k for the pre-programmed co-players) then decide how many slices (0, 3, 6 or 9) they would prefer to leave at home (or keep for themselves) and how many they would like to bring to school and contribute to the public good (g_i, g_j, g_k) (see Figure 2A).

At school all slices are placed on a "communal plate" (without showing which player contributed how much) before the virtual teacher adds 50% to whatever number of slices are on the plate (see Figure 2B). Afterwards, all slices on the plate are divided equally among players, regardless of what each player contributed initially (see Figure 2C). At the end of the round, the slices obtained at school and those left at home are added up to display the individual outcome of the round for subject i (see Figure 2D). The payoff per round and the overall payoff were displayed after each round (i.e., not permanently) in order to reduce the amount of information per screen and to minimize potential sources of distraction and confusion.

The payoff function which operationalizes the gains from each round for player i is thus specified by the following equation: $\prod_i = 9 - g_i + 0.5(g_i + g_j + g_k)$. The program performs all the computations in full view of the players to minimize the influence of mathematical competencies on game behavior.

Original article: Keil, J. et al. (2017). The Pizzagame: A virtual public goods game to assess cooperative behavior in children and adolescents. *Behavior Research Methods*, *49*(4), 1432-1443.

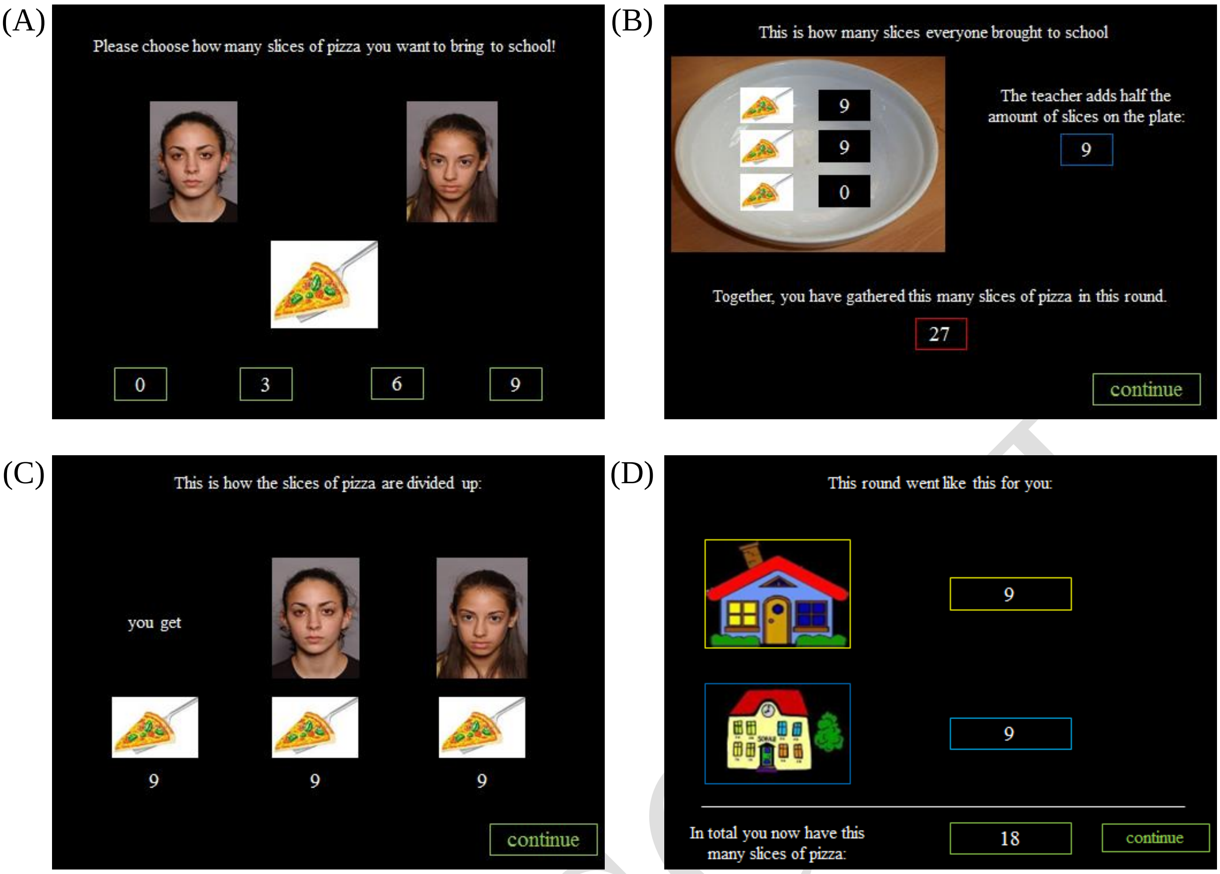

Figure 2 One hypothetical round of the *Pizzagame* illustrating the four key stages of the game: **(A)** Decision situation with pictures of the co-players and contribution options. **(B)** Display of anonymous individual contributions and subsidization by the teacher (50% of the sum of individual contributions). **(C)** Display of the redistribution of the public good to each individual player. **(D)** Display of the resource balance after one round of the PGG (NIMH- ChEFS; Egger et al., 2011).

In the first condition, subjects interact with highly cooperative co-players who both contribute all of their initial endowment of 9 slices of pizza in the first round. In the subsequent three rounds one co-player keeps on contributing 9 slices while the contributions of the other player slightly decline to 6 slices from the second round onwards. In the second condition, the co-players pursue a selfish strategy, commencing with 3 and 0 slices in the first round and then minimizing contributions to complete freeriding (0 slices) from the second round onwards. Finally, in the third condition, the co-players adopt a divergent strategy with first-round contributions of 9 and 3 slices from the cooperative and the selfish player, respectively. From the second round onwards, the selfish player decreases the contributions toward complete freeriding whereas the cooperative player carries on contributing all resources. Co-players' reductions of contributions from the first to the final round within each condition aimed to simulate the general behavioral pattern of decreasing contributions that was found in prior studies (e.g. Fehr & Gächter, 2000; Harbaugh & Krause, 2000).

Data Analysis

First, we report descriptive statistics for all study variables using SPSS statistical software, version 20.0 (SPSS Inc). In order to demonstrate the feasibility of the instrument we describe the number of errors on comprehension questions and the frequency with which children stated that the *Pizzagame* was their favorite part of the appointment for the whole sample of 216 children (prior to applying exclusion-criteria). Moreover, we report on the internal consistency of each condition by calculating Cronbach's alpha.

Subsequently, we applied structural equation modeling (SEM) using Mplus 7.11 (Muthén & Muthén, 2013) to test hypotheses 1, 2a, 2b and 3. As Aguirre-Urreta (2014) points out, these techniques bear several advantages as they better account for measurement error and offer more complex models than traditional techniques and might therefore strengthen data-analysis in experimental research. These analyses were carried out in three steps:

Step 1. To test the first hypothesis regarding conditional cooperation, we compared differences in latent mean contributions between conditions. To this end, a latent-state model was used to estimate latent means of cooperative behavior in each one of the three consecutive conditions. We specified a latent variable for each condition (i.e., cooperative, selfish, divergent). The first round of each condition was not included in the analyses. The rationale behind this decision was that subjects could only incorporate information about co-players' cooperative behavior into their decisions *after* playing the first round of each condition. Additionally, this minimized the potential carry-over of behavior from previous to subsequent conditions, reducing the risk of biasing the latent means. Accordingly, the last three rounds of each condition were used as indicators for the respective latent variable. No cross-loadings were specified. We also specified an autocorrelated residual structure between corresponding observed indicators from the three conditions (Sörbom, 1975). The latent mean scores were freely estimated using the effect coding method for identification of latent means (Little, 2013). To examine whether latent mean scores substantially differed between conditions, we specified three latent difference scores subtracting the latent mean score of the selfish from the cooperative condition (*d1*), the latent mean score of the divergent from the selfish condition (*d2*), and the latent mean score of the divergent from the cooperative condition, and checked if those differences in latent mean scores were significant.

Step 2. Age and gender were included in the latent state model as time-invariant predictor variables to test our second hypothesis. All three latent variables were regressed on both age and gender and we analyzed their impact on latent mean scores in all three conditions.

Step 3. As a third step we expanded the latent-state model to an autoregressive model in order to analyze the potential change in conditional cooperativeness as a function of age and gender. Thus, autoregressive paths were specified from the latent mean scores of the cooperative to the selfish and from the selfish to the divergent condition. In this autoregressive model we also analyzed the impact of age and gender on the latent mean of the selfish and divergent conditions as those latent variables reflect the behavioral change. Model fit was evaluated using (a) the chi-square statistic, (b) the comparative fit index (CFI), (c) the root mean squared error of approximation (RMSEA), and (d) the standardized root mean square residual (SRMR). According to Hu and Bentler (1999) a RMSEA of $\leq .05$ (.08) a CFI $\geq .95$ (.90) and a SRMR $\leq .05$ (.08) indicate a good (respective adequate) model fit.

Results

Results showed that the majority of children answered all nine comprehension questions correctly or committed only a single error (see Figure 3). The number of errors on the comprehension questions negatively correlated with age ($r_{age}(214) = -.22$, $p < .001$) but did not differ between boys and girls ($M_{male} = 0.4$, $SD = 0.76$; $M_{female} = 0.56$, $SD = 1.07$; $t(214)= -1.31$, $p = .19$). Excluded children were significantly younger than included children ($M_{age\_excluded}=11.00$, $SD = 1.96$; $M_{age\_included} = 12.03$, $SD = 1.92$; $t_{incl\_excl}(214)= 2.58$, $p < .05$), but there was no gender difference between excluded and included children ($\chi^2(1, N=216) = 1.09$, $p = .30$). Furthermore, 66.3% of the subjects stated that the *Pizzagame* was their favorite part of the appointment, followed by the verbal skills test (19.3%) and the storytelling task (14.4%).

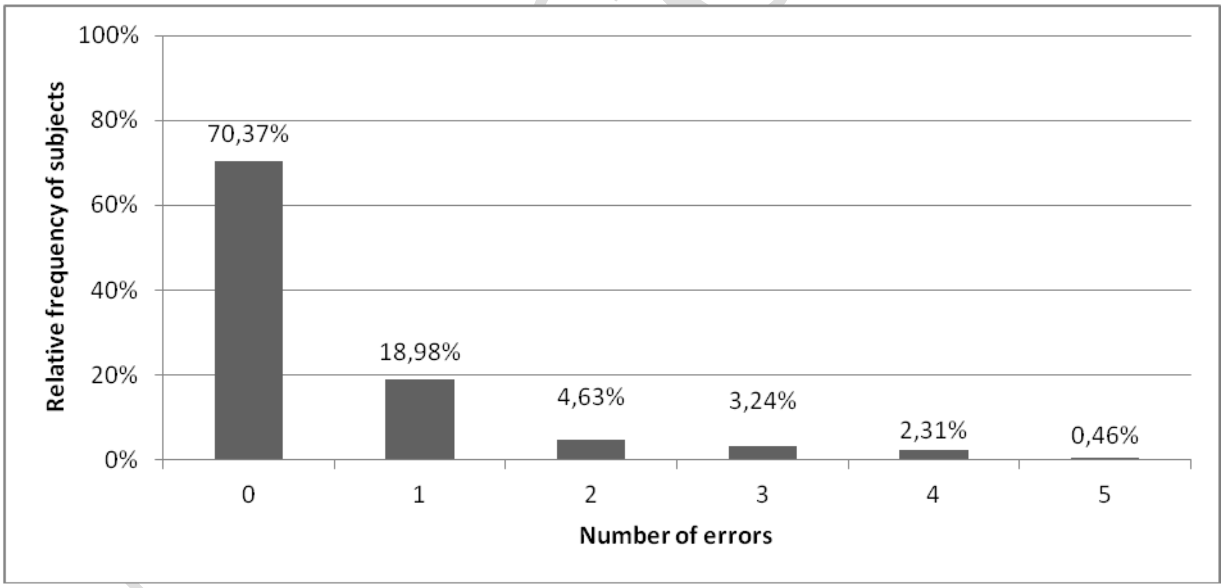

Figure 3 Distribution of committed number of errors on the nine comprehension questions (no children committed more than five errors)

In the first round of the cooperative condition subjects contributed 3.09 (34.3 %) pizza slices. First round contributions of all conditions were not associated with age ($r_{coop}(191) = .064$, $p = .381$; $r_{self}(191) = .116$, p = . 111; $r_{diverg}(191) = -.010$, $p = .889$) and did not differ as a function of gender ($t_{coop}(189)= -.330$, $p = .742$; $t_{self}(156.949) = .084$, $p = .993$;

$t_{diverg}(189) = -.581$, $p = .562$). The distributions of contributions initially increased in the cooperative condition, followed by a decrease in the selfish condition and a rise back to intermediate levels in the divergent condition (see Figure 4).

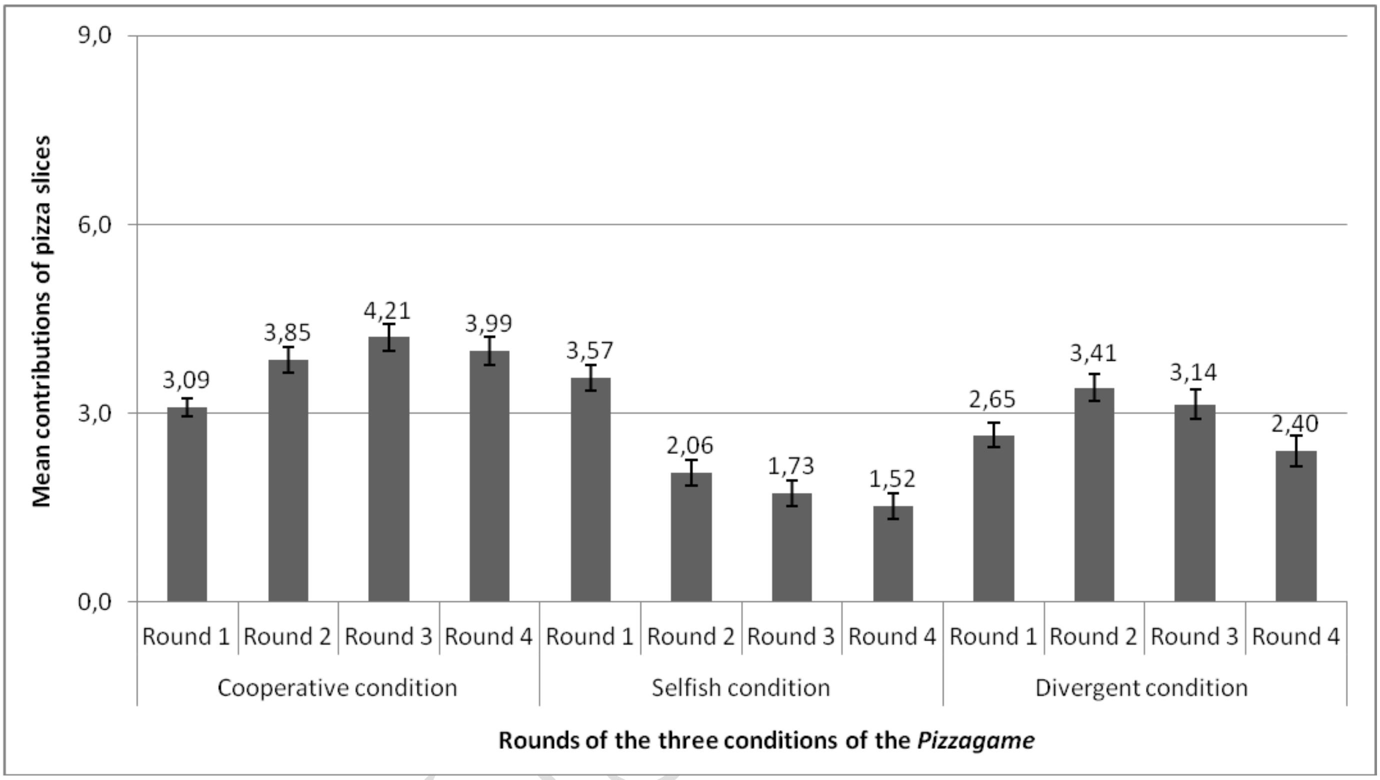

Figure 4 Mean contributions and corresponding standard error bars of each round of the cooperative, selfish and divergent condition.

Contributions in each condition yielded acceptable internal consistency (cooperative condition: $\alpha = .762$; selfish condition: $\alpha = .663$; divergent condition: $\alpha = .889$).

The latent-state model testing the first hypothesis regarding conditional cooperation showed an adequate model fit, $\chi^2(15) = 30.05$, $p < .05$, CFI = .97, RMSEA = .07, SRMR = .04. Estimated means of the latent variables indicated that the level of cooperative behavior was highest toward cooperative co-players ($M_{coop} = 4.02$; $SD = 2.39$), lowest toward selfish co-players ($M_{self} = 1.77$; $SD = 2.05$), and at a medium level toward divergent co-players ($M_{diverg} = 2.98$; $SD = 2.71$). Significant differences emerged between the latent means of the cooperative and the selfish condition (d1 = 2.25, $p < .001$) between the selfish and the divergent condition (d2 = -1.22, $p < .001$), and between the cooperative and the divergent

Original article: Keil, J. et al. (2017). The Pizzagame: A virtual public goods game to assess cooperative behavior in children and adolescents. *Behavior Research Methods*, *49*(4), 1432-1443.

condition (d3 = 1.03, $p < .001$), indicating decreased contributions when co-players were selfish or divergent rather than cooperative, and increased contributions when co-players were divergent rather than selfish.

In a second step, age and gender were included in the latent state model as time-invariant covariates to test their impact on contributions (Hypotheses 2a and 3). The resulting model showed an adequate model fit, $\chi^2(27) = 41.694$, $p < .05$, CFI = .975, RMSEA = .053, SRMR = .037. Results revealed that neither age (cooperative condition: $\beta = .028$, $p = .741$; selfish condition: $\beta = -.156$, $p = .091$; divergent condition: $\beta = .064$, $p = .434$) nor gender (cooperative condition: $\beta = .087$, $p = .298$; selfish condition: $\beta = .081$, $p = .389$; divergent condition: $\beta = -.040$, $p = .622$) had an impact on the level of contributions across all three conditions.

An autoregressive model (see Figure 6) was used to test whether age and gender predicted the change in contributions between conditions (Hypotheses 2b). The model showed an adequate model fit, $\chi^2(28) = 41.700$, $p < .05$, CFI = .976, RMSEA = .051, SRMR = .037. Results indicated that children decreased their contributions from the cooperative to the selfish condition more readily with increasing age ($\beta = -.18$, $p < .05$). Furthermore, they also increased their contributions more substantially from the selfish to the divergent condition with increasing age ($\beta = 0.21$, $p < .01$; see Figure 5). In contrast, gender neither impacted behavioral change from the cooperative to the selfish condition ($\beta = .002$, $p = .977$), nor from the selfish to the divergent condition ($\beta = -.12$, $p = .107$).

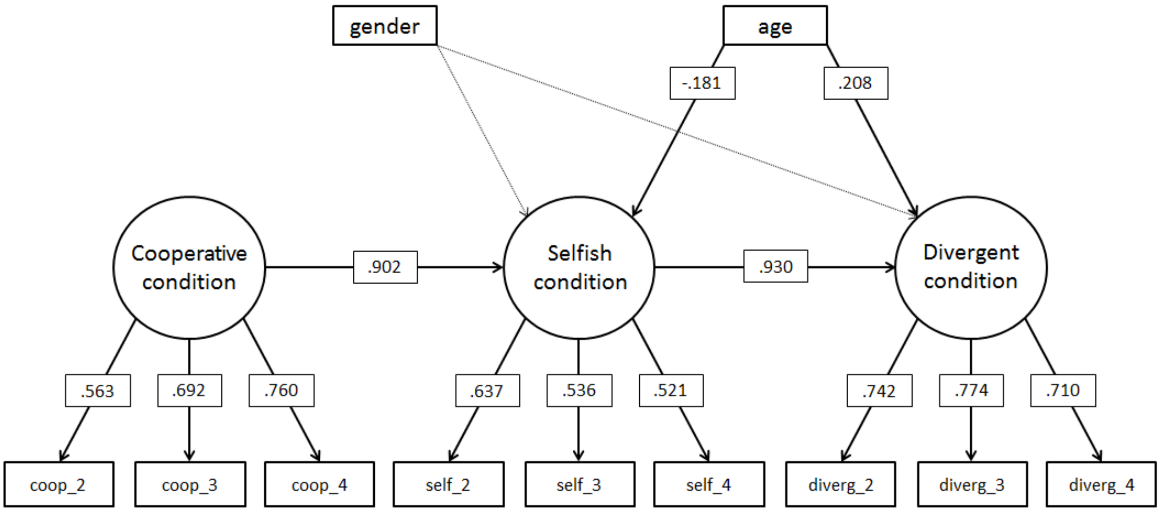

Figure 5Autoregressive model with standardized coefficients and paths between the three latent variables of the corresponding conditions (cooperative, selfish and divergent) and age and gender as time-invariant predictor variables (dashed lines indicate non-significant paths)

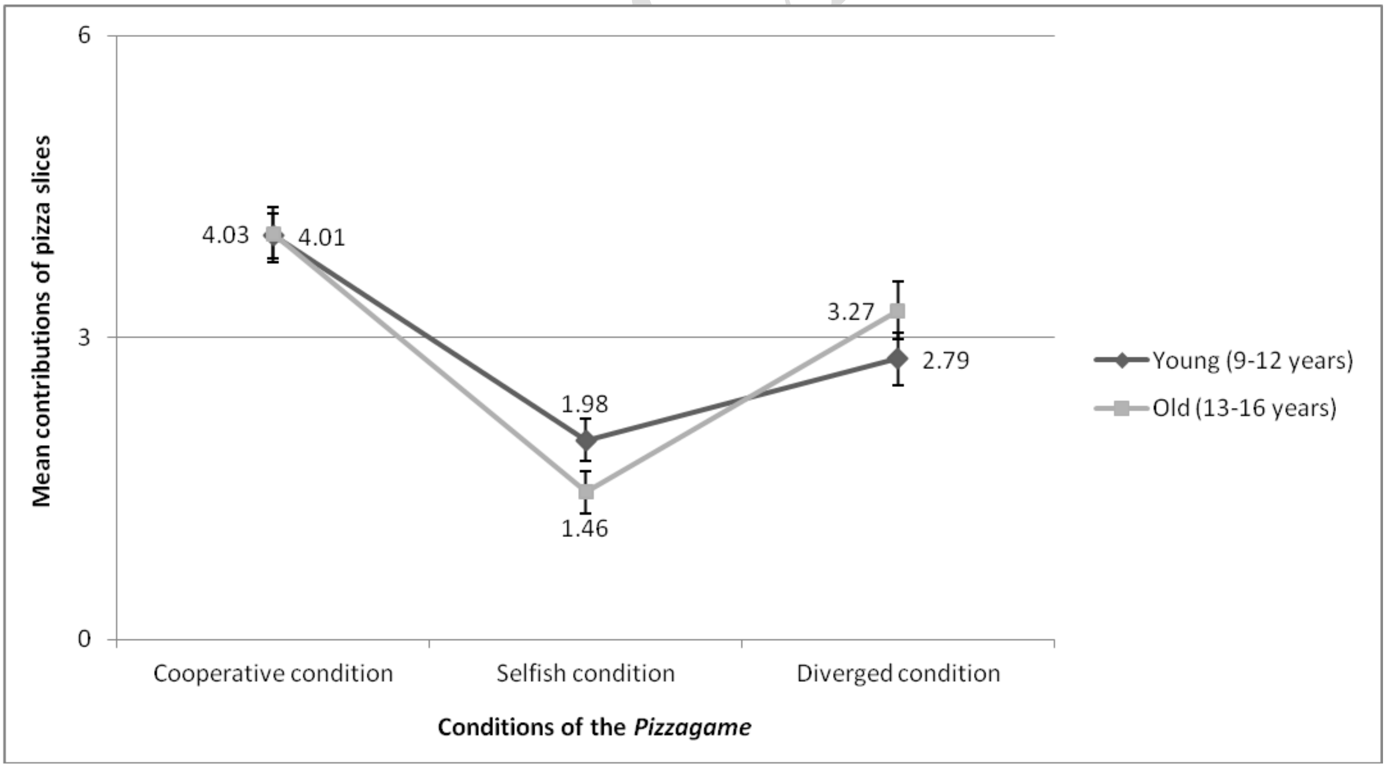

Figure 6 Illustrative graph displaying the latent mean contributions and standard error bars of the cooperative, selfish and divergent cooperation for younger (9-12 years) and older (13-16 years) children (Please note: Subjects were grouped into younger and older subjects to illustrate the age-effect for the purposes of this graph. All analyses were based on age as a continuous variable).

Discussion

Our study suggests that the *Pizzagame* – a newly developed computerized, child-friendly PGG based on a concrete real-life scenario – is an engaging instrument to feasibly and reliably assess cooperative behavior of children and adolescents. Lending support to the validity of the *Pizzagame*, our pattern of results confirmed that children exhibit a conditionally cooperative strategy, i.e., cooperate when others do, but also act selfishly when others do. Intriguingly, we also found that this conditionally cooperative strategy varies as a function of age, such that cooperation becomes more conditional on other players' strategies as children transition to adolescence.

Regarding the feasibility of our task, the *Pizzagame* was by far the most engaging procedure of the appointment. This is especially remarkable considering that other procedures (e.g. storytelling task) which are often considered highly engaging (Emde, Wolf, & Oppenheim, 2003), were also part of the appointment. Besides this, our impression was that children often reacted quite emotionally when their co-players shifted their strategies, which additionally underscores the engaging nature of the *Pizzagame*.

Supporting the *Pizzagame* as a valid measure for children and adolescents, results confirmed that we effectively altered cooperative behavior of our subjects. The *Pizzagame* was therefore successful in evoking a behavioral pattern which follows a conditionally cooperative strategy, thus falling into line with other findings among children (Vogelsang et al., 2014) and adults (Fischbacher et al., 2001).

The first round contributions in our study (i.e., 34.3 %) just fell below the lower end of the typical range detected by other studies (40-60%). In general, it is difficult to know why this minor discrepancy occurred and future research might examine whether this result is inherent to the specifics of the *Pizzagame* or simply due to error variance. Potentially, the 33%-choice may have struck the best balance between initial caution and still signaling a willingness to cooperate in the first round (58,6 % of children opted for the 33% choice in the initial round) giving rise to lower initial round contributions.

In contrast to our hypotheses, neither age nor gender had a significant direct impact on how much, on average, subjects contributed towards cooperative, selfish or divergent co-players. However, our data yield support for an increasingly conditional strategy of cooperation as children transition towards adolescence. This raises an interesting alternative interpretation of previous findings that have mainly reported linear associations between age and cooperativeness. Rather than contributing more or less than younger children, older children may adapt more readily to both cooperative and selfish strategies of their co-players. Thus, developmental differences may emerge primarily when children face variations in strategic behavior of their co-players which compel them to flexibly tailor their behavior accordingly. This could reflect a heightened sensitivity regarding the meaning of social behaviors and cooperative strategies among older children (e.g., due to better perspective-taking skills).

The null effects regarding gender concord with Harbaugh and Krause's (2000) study, but contrast with other work on children (Cipriani et al., 2007; Vogelsang et al., 2014). Given that gender effects have emerged among younger and older children (Cipriani et al., 2007; Vogelsang et al., 2014), a gender effect that diminishes with age is not likely. Another possible explanation for the absence of gender effects comes from a study on adults (Espinosa & Kovářík, 2015). The results of this study imply that when the context of the experiment is neutral, men and women do not behave differently. This may also apply here, as we devised the *Pizzagame* interface to be as gender-neutral as possible. However, further research is needed to clarify the role of gender in cooperative behavior in PGGs.

Limitations

Some limitations deserve mentioning. First and most importantly, conditions were not counterbalanced to control for any order effects. For example, it is conceivable that individuals were influenced by the initial experience of cooperative co-players for the remainder of the game. For the present purposes, we opted against counterbalancing due to the demands this would have placed on the number of different sequential arrangements as well as a balanced distribution of young and old girls and boys. Furthermore, because all participants were exposed to all conditions, individual differences are still meaningful, even though mean contributions might differ when re-ordering the sequence of conditions. As a first step towards validating the *Pizzagame* and in line with other paradigms that use similar set-ups (e.g., trust-rupture-trust, inclusion-exclusion-inclusion; King-Casas et al., 2008; White, Wu, Borelli, Mayes, & Crowley, 2013) we assumed that the cooperative-exploitative-divergent order would yield the most informative results. Specifically, we did not want children to face selfish co-players at the outset because this might have had a frustrating effect on participants at an early stage as well as repercussions on engagement and carry-over effects on later conditions. Moreover, we assumed that the best way to establish baseline cooperative behaviors was to program co-players to begin cooperatively. By placing the exploitative strategy second, we aimed to induce a large behavioral change from the cooperative condition, while the divergent condition would tap into a potential recovery of cooperative behavior and offer participants a choice between cooperative and exploitative strategies. Future research should certainly explore the order of the conditions.

Second, we used a forced choice design to make the game easier for children and adolescents. To be sure, a forced choice design makes comparisons with other studies on adults using open-choice paradigms tentative. At the same time, we faced a trade-off between comparability with previous work and age-appropriateness of the paradigm, and fell on the side of the latter given the main aim of this study.

Besides forced choice, our paradigm may raise concerns whether children (especially the older ones) actually believed they were connected to real children over the Internet. Our impression from the video recordings, however, was that most children were very engaged with the task, for example, responding with negative affect to the shift from cooperative to selfish co-players. In addition, previous work using other social interaction paradigms (e.g., Crowley et al., 2010; White et al., 2013) have also implemented similar measures while yielding valid and reliable results.

It is noteworthy that the *Pizzagame* involves deception. We believe that deception critically enhances credibility and ecological validity of the *Pizzagame* and it is therefore the preferred mode of administration (Bonetti, 1998). However, deception may raise ethical concerns (e.g., debriefing participants). Importantly, debriefing may be less effective for younger children (e.g., due to difficulties in reappraising experience in light of new information) and may even induce children to distrust experimenters and lead to negative affect (see Thompson, 1990, pp. 11-12). We therefore informed caregivers about the deception and then asked them (1.) whether they consent to their child playing the *Pizzagame* and (2.) whether they would like the experimenter to fully debrief their child afterwards regarding the deception. Notably, all children were exposed to an uplifting closing experience in our procedure (moderately cooperative condition followed by receiving a gift from the biggest box) to defuse potential negative affect. Also, they were given contact information of clinically trained personnel which they were encouraged to contact in the event of further distress. Overall, it is important to weigh the pros and cons of deception in the *Pizzagame* and to obtain ethical approval from the IRB before using deception.

Conclusion

We consider the *Pizzagame* to be a highly valuable tool in future research on cooperative behavior in children and adolescents for a number of reasons. First, the paradigm has proven highly engaging for children and adolescents. Second, compared to social games with multiple real-life subjects requiring coordination, the *Pizzagame* greatly simplifies data-collection and measurement of individual cooperative strategies while controlling for strategic interdependence. Third, using this instrument permits flexible manipulation of strategies of co-players in various ways, thus allowing investigation of individual differences in cooperation in different contexts. Finally, and perhaps most importantly, the instrument enables objective assessment of developmental and individual differences in cooperative *behavior* of children and adolescents. In so doing, the *Pizzagame* complements the vast number of subjective self-report measures for children, parents, and teachers to assess cooperative behavior of children (and similar constructs), commonly used in developmental science. Here, we have presented first evidence for a developmental shift towards more conditional cooperation as children move from middle childhood to adolescence. Given the burgeoning literature showing that peer problems figure prominently in the formation and maintenance of maladaptive behavior (Parker, Rubin, Stephen, Wojslawowicz, & Buskirk, 2006), behavioral assessments of cooperative strategies applicable to large samples may also add an important layer of understanding to the field of developmental psychopathology.

Acknowledgements

This research was part of the project *Analyzing pathways from childhood maltreatment to internalizing symptoms and disorders in children and adolescents* (AMIS; principal investigator Kai von Klitzing) and was funded by the Federal Ministry of Education and Research.

Conflict of Interest

The authors have declared that no conflict of interest exists.

References

Aguirre-Urreta, M. (2014). Improving Between-group Comparisons in IS Research Through the Use of SEM and Latent Variables: An Introduction and Some Examples. *Communications of the Association for Information Systems*, *34*(1), 30.

Alencar, A. I., De Oliveira Siqueira, J., & Yamamoto, M. E. (2008). Does group size matter? Cheating and cooperation in Brazilian school children. *Evolution and Human Behavior*, *29*(1), 42–48. doi:10.1016/j.evolhumbehav.2007.09.001

Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, *19*(3), 377–395. doi:10.1016/S0167-4870(98)00012-9

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. doi:10.1007/s10683-010-9257-1

Cipriani, M., Giuliano, P., & Jeanne, O. (2007). Like Mother Like Son? Experimental Evidence on the Transmission of Values from Parents to Children. Retrieved from http://hdl.handle.net/10419/34481

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and brain sciences*, *26*, 139–198.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448–474.

Crowley, M. J., Wu, J., Molfese, P. J., & Mayes, L. C. (2010). Social exclusion in middle childhood: rejection events, slow-wave neural activity, and ostracism distress. *Social Neuroscience*, *5*(5-6), 483–495.

Dawes, R. M. (1980). Social Dilemmas. *Annual Review of Psychology*, *31*(1), 169–193. doi:10.1146/annurev.ps.31.020180.001125

De Dreu, C. K. (2013). Human Cooperation: Challenges for Science and Practice. *Psychological Science in the Public Interest*, *14*(3), 117–118. doi:10.1177/1529100613496959

Doherty, M. (2009). *Theory of mind*: Philadelphia, PA: Psychology Press.

Donaldson, M. (1978). Children\'s minds.

Egger, H. L., Pine, D. S., Nelson, E. E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS). A new set of childrens facial emotion stimuli. *International journal of methods in psychiatric research*, *20*(3), 145–156.

Emde, R. N., Wolf, D. P., & Oppenheim, D. (2003). *Revealing the inner worlds of young children: The MacArthur story stem battery and parent-child narratives*: Oxford University Press.

Espinosa, M. P., & Kovářík, J. (2015). Prosocial behavior and gender. *Frontiers in behavioral neuroscience*, *9*, 88. doi:10.3389/fnbeh.2015.00088

Fan, C.-P. (2000). Teaching children cooperation — An application of experimental game theory. *Journal of Economic Behavior & Organization*, *41*(3), 191–209. doi:10.1016/S0167-2681(99)00072-4

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review*, *90*(4), 980–994.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. doi:10.1016/S0165-1765(01)00394-9

Gross, M., & Heinrichs, H. (2010). *Environmental sociology: European perspectives and interdisciplinary challenges*. Dordrecht, New York: Springer.

Gummerum, M., Hanoch, Y., & Keller, M. (2008). When Child Development Meets Economic Game Theory: An Interdisciplinary Approach to Investigating Social Development. *Human Development, 51*(4), 235–261. doi:10.1159/000151494

Gunther Moor, B., Bos, M. G. N., Crone, E. A., & van der Molen, M. W. (2014). Peer rejection cues induce cardiac slowing after transition into adolescence. *Developmental Psychology*, *50*(3), 947.

Guyer, A. E., Choate, V. R., Pine, D. S., & Nelson, E. E. (2012). Neural circuitry underlying affective response to peer feedback in adolescence. *Social Cognitive and Affective Neuroscience*, *7*(1), 81–92.

Harbaugh, W. T., & Krause, K. (2000). Children's Altruism In Public Good And Dictator Experiments. *Economic Inquiry*, *38*(1), 95–109.

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248. doi:10.1126/science.162.3859.1243

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *science*, *321*(5890), 806–810.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE, 3*(7), e2597. doi:10.1371/journal.pone.0002597

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, Daniel H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388. doi:10.1080/02699930903485076

Ledyard, J. O. (1995). Public Goods. A Survey of Experimental Research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (Chap. 2). Princeton, N.J.: Princeton University Press.

Leipold, K., Vetter, N. C., Dittrich, M., Lehmann-Waffenschmidt, M., & Kliegel, M. (2013). Individual and developmental differences in the relationship between preferences and theory of mind. *Journal of Neuroscience, Psychology, and Economics*, *6*(4), 236–251. doi:10.1037/npe0000008

Little, T. D. (2013). *Longitudinal structural equation modeling*. *Methodology in the social sciences*. New York: The Guilford Press.

McClure, E. B., Parrish, J. M., Nelson, E. E., Easter, J., Thorne, J. F., Rilling, J. K.,. . . Pine, D. S. (2007). Responses to Conflict and Cooperation in Adolescents with Anxiety and Mood Disorders. *Journal of Abnormal Child Psychology*, *35*(4), 567–577. doi:10.1007/s10802-007-9113-8

Muthén, B. & Muthén, L. (2013). *Mplus statistical software version 7.11.*

Olson, M. (1971 (1965)). *Logic of collective action: Public goods and the theory of groups* (10.print). Cambridge, mass.

Parker, J. G., Rubin, K. H., Stephen, E. A., Wojslawowicz, J. C., & Buskirk, A. A. (2006). Peer Relationships, Child Development, and Adjustment. A developmental psychopathology perspective. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental Psychopathology. Theory and Method* (pp. 419–493). John Wiley & Sons, Inc.

Parks, C. D., Joireman, J., & Van Lange, P. A. (2013). Cooperation, Trust, and Antagonism: How Public Goods Are Promoted. *Psychological Science in the Public Interest*, *14*(3), 119–165. doi:10.1177/1529100612474436

Reijntjes, A., Stegge, H., Terwogt, M. M., Kamphuis, J. H., & Telch, M. J. (2006). Emotion regulation and its effects on mood improvement in response to an in vivo peer rejection challenge. *Emotion*, *6*(4), 543.

Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology*, *27*(1), 73–97. doi:10.1016/j.joep.2005.06.015

Schneider, E., & Zuccoloto, A. (2007). E-prime 2.0 [Computer software]: Pittsburg. *PA: Psychological Software Tools*.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime computer software and manual. *Pittsburgh, PA: Psychology Software Tools Inc*.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281–288). ACM.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, *28*(2), 138–151.

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition*, *85*(3), 277–290.

Thompson, R. A. (1990). Vulnerability in research: A developmental perspective on research risk. *Child Development*, *61*(1), 1–16.

Van Lange, P. A., Joireman, J., Parks, C. D., & van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125–141.

Vogelsang, M., Jensen, K., Kirschner, S., Tennie, C., & Tomasello, M. (2014). Preschoolers are sensitive to free riding in a public goods game. *Frontiers in Psychology*, *5*.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology*, *23*(1), 63–71.

White, L. O., Klein, A. M., Kirschbaum, C., Kurz-Adam, M., Uhr, M., Müller-Myhsok, B.,. . . Stalder, T. (2015). Analyzing pathways from childhood maltreatment to internalizing symptoms and disorders in children and adolescents (AMIS): a study protocol. *BMC Psychiatry*, *15*(1), 126.

White, L. O., Wu, J., Borelli, J. L., Mayes, L. C., & Crowley, M. J. (2013). Play it again: neural responses to reunion with excluders predicted by attachment patterns. *Developmental science*, *16*(6), 850–863.

Zelmer, J. (2003). Linear Public Goods Experiments A Meta-Analysis. *Experimental Economics*, *6*, 299–310.