MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

Zvonimir Bašić
Parampreet C. Bindra
Daniela Glätzle-Rützler
Angelo Romano
Matthias Sutter
Claudia Zoller

Discussion Paper
2024/2

THE ROOTS OF COOPERATION

The roots of cooperation *

Zvonimir Bašić, Parampreet C. Bindra, Daniela Glätzle-Rützler, Angelo Romano, Matthias Sutter, and Claudia Zoller†

Abstract. We study the developmental roots of cooperation in 929 young children, aged 3 to 6. In a unified experimental framework, we examine pre-registered hypotheses about which of three fundamental pillars of human cooperation – direct reciprocity, indirect reciprocity, and third-party punishment – emerges earliest and is more effective as a means to increase cooperation in a repeated prisoner's dilemma game. We find that already children aged 3 act in a conditionally cooperative way. Yet, direct and indirect reciprocity do not increase overall cooperation rates beyond a control condition. Compared to the latter, punishment more than doubles cooperation rates, making it the most effective mechanism to promote cooperation. We also find that children's cognitive skills and parents' socioeconomic background influence cooperation. We complement our experimental findings with a meta-analysis of studies on cooperation among adults and older children, confirming that punishment outperforms direct and indirect reciprocity.

JEL-Codes: C91, C93, D01, D91, H41

Keywords: Cooperation, reciprocity, third-party punishment, children, parents, prisoner's dilemma game, experiment, meta-analysis

This version: 20 December 2023

* We would like to thank Ingvild Almås, Silvia Angerer, Michal Bauer, Isabelle Brocas, Alexander Cappelen, Juan Carillo, Gary Charness, Julie Chytilova, Stefano DellaVigna, Armin Falk, Ernst Fehr, Simon Gächter, Henning Hermes, Felix Kölle, Fabian Kosse, Philipp Lergetporer, Hannah Schildberg-Hörisch, Daniel Schunk, Bertil Tungodden, and seminar participants at CESifo Munich, UC Santa Barbara, NHH Bergen, University of Mainz, University of Innsbruck, and University of Queensland for helpful comments. We are grateful to all parents of the participating children, the preschool teachers, and kindergartens involved for making this study possible. We thank also Brian Cooper, Alexandra Zwankhuizen, Anna-Elena Pinggera, and Jessica Loevenich for research assistance. Funding from the University of Cologne (through the Hans Kelsen Prize) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged. This study was approved by the IRB at the University of Innsbruck (application number 09/19) and pre-registered in the Open Science Framework under https://osf.io/th25y/?view_only=6ff7e93a9b4c481d963dae17cec3e9d9.

† Bašić: Adam Smith Business School, University of Glasgow, UK. Bindra and Glätzle-Rützler: University of Innsbruck, Austria. Romano: Leiden University, Netherlands. Sutter: Max Planck Institute for Research on Collective Goods Bonn, Germany, University of Cologne, Germany, University of Innsbruck, Austria, IZA Bonn, Germany, and CESifo Munich, Germany. Zoller: Management Center Innsbruck, Austria. Matthias Sutter is corresponding author. Email: matthias.sutter@coll.mpg.de.

1. Introduction

Cooperation is key to the success of many crucial human challenges. Among the mechanisms that promote and support cooperation, three have been identified as fundamental: direct reciprocity, indirect reciprocity, and third-party punishment (see, e.g., Trivers, 1971, Axelrod and Hamilton, 1981; Nowak and Sigmund, 1998, 2005; Fehr and Fischbacher, 2004). Previous studies have studied single mechanisms and their influence on cooperation in isolation, and mostly studied these mechanisms with adult subjects. Yet, understanding the relative importance of mechanisms that promote cooperation can be crucial in tailoring interventions to increase cooperative behavior, and this can be particularly effective if such interventions are implemented at a young age.

In this study, we examine the relative effects of these three fundamental mechanisms of cooperation at a very young age. To this end, we design and conduct a novel lab-in-the-field experiment with 929 children, aged 3 to 6 years. This allows examining whether and which of these mechanisms develops its potential to increase cooperation in young children, at which age this development occurs, and whether and how cognitive and non-cognitive skills, as well as socioeconomic determinants and parenting style, moderate the effects of these different mechanisms.

We then put the results from our study into perspective by complementing it with *i)* a novel meta-analysis (covering 105 publications) that estimates the treatment effects of direct and indirect reciprocity, and of punishment on cooperation of adults, and *ii)* a literature survey on related studies with older children. Together with our main experiment, this allows us to identify systematic patterns of cooperative behavior from early childhood into adult age.

We focus on three fundamental mechanisms behind cooperation. The first one, direct reciprocity, refers to the ability to engage in mutually beneficial interactions with the same partner by reacting cooperatively to a partner's cooperation and less cooperatively to defection (Axelrod and Hamilton, 1981). Direct reciprocity has been shown to yield higher cooperation rates than in one-shot situations or repeated interactions where subjects never meet anybody twice (e.g., Keser and van Winden, 2000; Clark and Sefton, 2001; Brandts and Rivas, 2009; Duffy and Ochs, 2009). The second mechanism, indirect reciprocity, promotes cooperation through the spread of reputational information, even if someone has not interacted with a particular partner in the past. By being cooperative towards others, one gains a positive reputation that is observable by future interaction partners. Compared to situations without any information about a subject's past behavior, and thus no reputational information, indirect reciprocity can increase cooperation, albeit typically less than direct reciprocity (e.g., Bolton et

al., 2005; Seinen and Schram, 2006). Previous research has claimed that direct and indirect reciprocity are particularly effective in small groups (Boyd et al., 2003; Boyd and Richerson, 2009). This is because direct and indirect reciprocity require either experiencing repeated interactions with the same partner or knowing an interaction partner's past behavior, which may be challenging in larger groups. Therefore, the gene-culture coevolution approach (Boyd et al., 2003; Boyd and Richerson, 2009) has proposed that the two reciprocity mechanisms do not suffice to sustain cooperation, but that a third mechanism, altruistic third-party punishment, is needed. In this case, an unaffected bystander may punish (at own costs) other subjects who defect in the cooperation game, while the bystander is not directly affected by the level of cooperation. Third-party punishment has typically been found to raise cooperation (see, e.g., Fehr and Fischbacher, 2004; Carpenter and Matthews, 2012; Leibbrandt and Lopez-Perez, 2012).

We present an experiment with 3- to 6-years old children who play a repeated prisoner's dilemma game. We compare a control condition with treatments that allow assessing the effects of direct and indirect reciprocity, and third-party punishment. Pursuing our research question specifically with young children helps us understand how and under which conditions cooperation can be expected to develop and flourish. Since early childhood is formative for children's skills and behavior and their lifetime outcomes (Heckman, 2006; Fehr et al., 2008; Almås et al., 2010; Heckman et al., 2013; Berger et al., 2020; Cappelen et al., 2020; García et al., 2020; Kosse et al., 2020), understanding how cooperation evolves in young children may help identify sources of economic efficiency in groups or society in general, and may inform practitioners regarding which mechanisms and institutions to target in order to facilitate cooperation. Indeed, there has been a push for research on how different institutions shape children's cooperative behavior across the development (House et al., 2013). Finally, promoting cooperative behavior may also have long-term beneficial consequences on the individual level, since recent work has shown that cooperative behavior – as an important aspect of prosociality – is linked to labor market success (Kosse and Tincani, 2020).

There has been little research on children's willingness to cooperate in strategic games in general. This dearth of evidence is slightly surprising since research on the formation of economic behavior at an early age has gained strong momentum and lots of attention in recent years (see Sutter et al., 2019, and List et al., 2023, for surveys). For instance, Cappelen et al. (2020) and Kosse et al. (2020) have reported how educational interventions and mentoring programs, respectively, can promote prosociality in disadvantaged children (in simple sharing tasks). Almås et al. (2010) have studied the development of meritocratic principles in allocation

tasks performed by adolescents. Fehr et al. (2008, 2013) and Bauer et al. (2014) have examined fairness concerns and the importance of egalitarian allocations for children. Contrary to this evidence on non-strategic sharing and allocation tasks, strategic interaction games like the prisoner's dilemma have rarely been studied with young children. To date, studies investigating social dilemmas primarily involve older children, i.e., school-age children above the age of 6 (e.g., Harbaugh et al., 2000; Peters et al., 2004; Lergetporer et al., 2014; Blake et al., 2015; Hermes et al., 2020), are often focused on one-shot interactions (e.g., Sutter and Untertrifaller, 2020), which makes it impossible to study whether (direct or indirect) reciprocity can improve cooperation, or do not have a control treatment to compare the relative effects of the mechanism they employ (Vogelsang et al., 2014).

Our experimental design enables us to investigate the comparative effects of direct and indirect reciprocity, as well as third-party punishment, and to reveal potential moderating effects of a child's and the parents' characteristics. While the socioeconomic status (SES) of parents has been shown to be strongly related to children's risk and time preferences and prosociality in non-interactive tasks (Dohmen et al., 2012; Bauer et al., 2014; Falk et al., 2021), we can examine its relation to behavior in a strategic game, and investigate how it interacts with the institutional environment. We can similarly test the relevance of parental warmth which has been found to be related to non-strategic prosociality (Falk et al., 2021). Finally, we can investigate the importance of several traits which might be decisive for the development of cooperative behavior: children's cognitive abilities (which may help to navigate strategic interactions; Proto et al., 2019), theory of mind (which may facilitate understanding a partner's or a third party's intentions and thus make cooperation easier; Brüne and Brüne-Cohrs, 2006; Fe et al., 2022), and the ability to delay gratification (which may help to resist the temptation of defection in a cooperation game; Kölle and Wenner, 2023).

In our experiment with more than 900 children, aged 3 to 6, we find a very strong effect of third-party punishment on cooperation rates. On average, and across all rounds, 68% of children cooperate when third-party punishment is possible. This is more than double the rate in all other conditions, where cooperation rates are in the range of 24% to 29%. Already the mere presence of third parties increases children's cooperation rates in the first round, i.e., before any punishment could have taken place. Experiencing actual punishment has a further effect by turning past defectors with a high probability into future cooperators.

In comparison to a control condition (with perfect stranger matching and without any information on a partner's previous behavior), the treatments that allow for direct or indirect reciprocity do not improve overall cooperation rates. This does not imply, however, that

children do not behave reciprocally in these treatments. In fact, we show that they do engage in reciprocating the past behavior of others, thus providing the first evidence that conditional cooperation (Fischbacher et al., 2001) exists already at age 3. Yet, given a low base rate of cooperation in the first round, reciprocity keeps the cooperation rates at a low level, since children typically respond to a partner's past defection with own defection. However, those pairs of children who start with mutual cooperation are significantly more likely to cooperate later on as well.

The age of children has a moderate relation to cooperation, yet it interacts with the institutional environment. Older children increase their cooperation in the third-party punishment treatment, but they tend to decrease it in the other treatments. We see a similar relationship for children's cognitive abilities. Children with higher abilities increase their cooperation more in the third-party punishment treatment than in the control treatment. Children's theory of mind and patience are most of the time unrelated to cooperation rates. The same holds true for parental warmth. Yet, the education of parents matters. Children of better-educated parents cooperate more often in the third-party punishment treatment in comparison to the control treatment. Together, our results suggest that age, cognitive abilities, and socioeconomic background are relevant factors for children's decision-making in strategic cooperation games, but they also interact with the institutional environment. Finally, we show that all our experimental results are highly robust to a variety of potential concerns – to more or less stringent comprehension inclusion criteria, to potential selection effects concerning passing the comprehension checks and the survey response among parents, and to different ways of calculating children's reputation.

The key finding of our experiment is that punishment clearly outperforms direct and indirect reciprocity in the young age group of 3- to 6-year olds. This finding on the relative effects of punishment, direct and indirect reciprocity raises the question whether similar relative effects prevail also after early childhood. No study has ever compared those three pillars of cooperation in a unified framework, yet it is possible to estimate effect sizes from a meta-analysis about existing studies that compare a single pillar to a control condition. We conclude our paper by complementing our experimental results with: *i)* such a novel meta-analysis on studies with adult subjects using 105 publications and 264 distinct effect sizes, and *ii)* a literature survey on related studies with older children and adolescents. This allows us to bridge the gap from childhood to adulthood and embed our results into the larger picture about which pillars promote cooperation. Based on our meta-analysis, we observe for adults that the introduction of all three mechanisms – punishment, direct reciprocity, and indirect reciprocity

– induces a positive effect on cooperation. Yet, in line with our findings for children, punishment shows the largest effect. This suggests that the reason why we observe a strong effect of third-party punishment, but no effect of reciprocity in young children, is not solely due to reciprocity mechanisms developing later, but they are also weaker in their effect on behavior, even once they are fully developed. Moreover, we find that punishment always causes a large effect on behavior, regardless if we look at 3-6 year-old children in our experiment, 6-11 year-old children in a related study, or the meta-analysis with adults. In contrast, by comparing the related literature, we reveal suggestive evidence that reciprocity mechanisms start supporting cooperative behavior roughly in the period within primary school age (6-10 years of age), and the effects always remain small to moderate, both when looking at related studies with older children, or at the meta-analysis with adult subjects. In summation, our experimental study, the literature survey on related studies with older children, and the meta-analysis with adults, together show that the main pillar to support high levels of cooperation – punishment – works from very early on in life, while the reciprocity pillars become more effective slightly later and do not catch up with punishment even in adulthood.

The rest of the paper is structured as follows. We will present our experimental design and procedure in Section 2. The experimental results are shown in Section 3. Our meta-analysis is reported in Section 4, and Section 5 discusses and concludes the paper.

2. Experimental design and procedures

2.1 Subject pool and procedures

We recruited 19 public kindergartens in Tyrol, Austria, and ran our study in spring 2019. We informed the children's parents about running a research project for two consecutive days. We received parental consent for 1,231 children, which represents 88% of all children in these kindergartens. Participation of children was voluntary. Since the project ran for two days, not all children could participate on both days, mainly due to illnesses or families taking a day off from kindergarten. For these reasons, 168 children were absent on the second day (when the cooperation game was played). We also had 86 children with very poor language skills (mostly due to German being their second language), which led to insurmountable comprehension difficulties.[1] The cooperation game was computerized and run on tablets (see below for details). We had a technical error that made the data of 29 children unusable. 19 children dropped out

[1] We identified the children who had insurmountable language difficulties on day one (with the assistance of kindergarten teachers and while playing with them the tasks on day one), and on day two these children only played non-interactive trial rounds and hence were not paired with other children.

during the study. In total, this leaves us with 929 children who played and finished the cooperation game. Among them, we had an equal distribution of gender (50% female). Austrian kindergartens have three age cohorts (3/4-year-olds, 4/5-year-olds, and 5/6-year-olds).[2] Table 1 summarizes the number of children in each age cohort, overall and separately for each treatment introduced below. Our pre-registered sample size of 180 subjects per treatment was chosen to be able to detect small-to-medium treatment effects.[3]

Table 1 about here

The experiment took place in separate rooms of each kindergarten, keeping children in a familiar and natural environment where they regularly spend their weekdays. To ensure the anonymity of decisions and to adhere to data-protection regulations, children were assigned a unique code, stuck onto a child's shirt by the kindergarten teachers. Research assistants (RAs) – who had been trained extensively – were only familiar with the first name of a child and removed the sticker after completion of the experiment.

The experimental instructions used only vocabulary that was familiar to children. We presented the cooperation game in a game-like, animated manner on tablets with touchscreens. The graphic design and animations were developed in cooperation with a professional company (see https://uxkids.com), which specializes in digital products for young children and their user experience. In addition to the animated, child-oriented interface, we used headphones with prerecorded texts. Using tablets with a child-friendly interface and supporting the interface with automated texts over headphones provides a very large degree of standardization in the experimental procedures. In addition to these technical tools, before the game started, the experiment was first explained in a one-to-one setting of one well-trained RA and one child, which allowed us to keep the explanation standardized while at the same time adjusting the explanation to each child's comprehension speed. When going through the instructions, children had to answer follow-up questions to make sure they paid attention and could follow

[2] The age cohort a child belongs to is determined by his or her birthday, with the cutoff on 31 August. Therefore, our youngest cohort included children who turned 4 sometime between 1 September 2018 and 31 August 2019.

[3] Our core goal in determining the sample size was to detect small-to-medium treatment effects for reciprocity and third-party punishment ($V = 0.15$). Cramer's V is a standardized measure of effect sizes for binary data and it can be considered as the strength of association between two variables (Mai and Zhang, 2016). A prior power analysis to detect a between-subjects effect at statistical power $(1-\beta) = 0.80$ and $\alpha = 0.05$ indicated a required sample size of 180 subjects per treatment (Mai and Zhang, 2016). This is a very conservative power estimation since each of these 180 subjects per treatment will decide on cooperation or defection for 5 times. In fact, we gathered more than 4,000 cooperation decisions across our four treatments.

the instructions. If children were unable to answer a question correctly, the RA explained the corresponding paragraph of the instructions again before moving on.

As incentives, children could earn tokens that could be exchanged for small presents. Each token was worth one present. We used a large variety of presents (e.g., balloons, toys, candy) that also differed between day 1 and day 2 in order to avoid satiation. The approach allowed each child to find presents they valued.

2.2 Measuring cognitive abilities, theory of mind, and patience as potential prerequisites for cooperation (Day 1)

On day 1, we measured cognitive abilities, theory of mind, and patience (see the instructions in the Appendix).[4] To measure cognitive abilities, we used 12 puzzles of Raven's Colored Progressive Matrices, designed for implementation with young children. The matrices provide an estimate of fluid intelligence. We printed and presented them as a booklet, so that children could progress at their own speed. Children were instructed to find the missing puzzle piece from a set of 6 multiple-choice answers by marking the correct picture. The puzzles were ordered by increasing difficulty. The RA and the child solved the first puzzle together to make sure the child had understood the task. The remaining 11 puzzles were solved by the child independently without receiving any feedback on performance. On average, children gave correct answers in 6.96 out of 11 cases, with meaningful variation in correct answers across children (standard deviation 1.48). We report a balance table concerning cognitive abilities (as well as other characteristics and background variables) across treatments in Table A1 in the Appendix. As would be expected in light of children's cognitive development, we observe that cognitive abilities increase with the age cohort (Pearson correlation coefficient $r = 0.39$; $p < 0.01$). Each child received 1 token for this task, irrespective of performance.

To measure theory of mind, we used the standard change-of-location task, which is an established and validated measure for eliciting theory of mind in young children (Wimmer and Perner, 1983; Cowell et al., 2015). The RA reenacted a story of two dolls (matched to the gender of the child – called either "Sarah and Anna" or "Stefan and Adam", reflecting common Austrian names). Together, the two dolls hide a ball in location 1. Then Anna (Adam) leaves and Sarah (Stefan) takes the ball and hides it in location 2. Anna (Adam) returns and the child is asked where Anna (Adam) will look for the ball. A child is classified as possessing theory of

[4] We deliberately began with non-interactive tasks on day 1 (rather than with the cooperation game), because they are simpler to understand for children of this age, which helped them to get engaged in our project. Moreover, day 1 was intended to build up trust in children when working with us, in particular that they experienced that we followed up on what we said and that they could exchange the earned tokens into presents.

mind if he/she is able to answer this question correctly.[5] This was the case for 72.8% of children. This overall fraction hides a clear age pattern. Of the 3/4-year-olds, 56% of children solve the task correctly, and this fraction increases to 71% for 4/5-year-olds, and 82% for 5/6-year olds (Pearson correlation coefficient $r = 0.22$; $p < 0.01$), revealing a familiar pattern of increasing theory of mind in the kindergarten years. Each child received 1 token for participation in this task, irrespective of performance.

To measure patience, we gave children 2 tokens. They could exchange them for presents on the same day or could save one or two tokens. Saving meant that a token was doubled for the next day. In other words, children could choose between *(i)* 2 tokens today and 0 tokens tomorrow*, (ii)* 1 token today and 2 tokens tomorrow, or *(iii)* 0 tokens today and 4 tokens tomorrow. The number of tokens saved for the next day serves as our measure of patience. The average across all children was 0.92 tokens (out of 2 tokens) saved for the next day (standard deviation 0.85; see Table A1 in the Appendix). We observe no correlation between patience and age cohort (Pearson correlation coefficient $r = 0.04$; $p = 0.21$).

2.3 The cooperation game (Day 2)

On day 2, children played a prisoner's dilemma (PD) game for 5 rounds. In each round, children had to make a single choice – either keep 1 token for themselves (defect) or give 2 tokens to their partner (cooperate). We present the game in such binary choices to keep the game simple and appealing to children, while at the same time preserving the key payoff properties of a PD game. We report the stage-game payoff matrix in Table 2. Keeping the 1 token maximizes a child's earning in a given round, which makes this the dominant strategy under standard assumptions, yet mutual cooperation would lead to a Pareto improvement, yielding 2 tokens for both players. Under standard assumptions and by applying backward induction, a player should defect in all rounds.

The payoff parameters and the number of rounds were chosen to give sufficient room for all treatments to increase cooperation,[6] while still maintaining the payoff simplicity and an acceptable duration of the study for our very young subjects. In particular, the average session

[5] After this question, we also asked where the toy really was and where the dolls had hidden the toy in the beginning. 94.4% of children could answer these two questions correctly.

[6] In particular, when looking at finitely repeated PD games, Embrey et al. (2018) establish that the key parameters which matter for establishing cooperation in direct reciprocity settings are the stage-game payoff and the horizon of the game. When the game is standardized, the payoff parameters can be captured through the one-shot gain from defection and the one-shot loss from being defected on. Looking at previous studies, they find that the two payoff variables range from 0.44 to 4, and 0.78 to 4, respectively, with lower numbers being more supportive of cooperation. In our PD game, these parameters are 1 and 1, respectively. Moreover, Embrey et al. (2018) report that the number of rounds ranges from 2 to 10 in the surveyed studies, while we use 5 rounds. This means that our parameterization provides an environment that can be expected to foster cooperation.

lasted around 30 minutes, which is leaning towards the maximum length of how long young children are able to focus on such a task. We did not elicit beliefs about the other child's decision in the PD game. Kindergarten teachers recommended against it because running a repeated game and asking for beliefs after each round would have likely exceeded the children's attention span and mental capacity. The lack of eliciting beliefs in our study also seems justified for the following reasons. First, it has been shown that adult behavior does not differ between situations where either no beliefs are elicited or asking for beliefs is unincentivized; when beliefs are incentivized, the pattern of behavior still remains the same (compared to the other two conditions), but cooperation rates increase slightly (Gächter and Renner, 2010). Second, it is also known that asking for beliefs in repeated PD games yields a very high correlation between a partner's behavior in the previous round and a subject's belief about the partner's cooperation in the current round, thus limiting the (independent) informational value of elicited beliefs (Gächter et al., 2017).[7]

Table 2 about here

The options – of cooperation or defection – were indicated on the tablet screen as a closed hand (keeping the one token) or an open hand (handing over the two tokens to the partner), as can be seen in Panel A of Figure A1 in the Appendix. Players were always illustrated by an avatar, with each avatar having a different color and representing one other (but unknown) child in the room.

In each session, we randomly matched children into groups of 6, and (except for a few cases where it was logistically not feasible) all children within a group were: *(i)* within the same age cohort (young, middle, or old)[8], *(ii)* from at least two different classrooms, and including both boys and girls, and *(iii)* had no siblings within the groups. An experimental session was randomly assigned to one of the following treatments, ensuring balanced age across treatments (see Table 1):

1. *Control (CTR)*: Each child was matched with a different child in each round by implementing a perfect stranger matching so that none of the six children per session interacted with any other child more than once. The partner's avatar at the top of the

[7] From a methodological point of view, Costa-Gomes and Weizsäcker (2008) provide evidence that beliefs and behavior (in normal-form games) often do not match, which also limits the informational value from elicited beliefs considerably.

[8] Note that this is an established standard in the literature. While one could also match children across age cohorts, this is not a research question we pursue.

screen changed in each round to highlight the change of the partner. A child received no information on the current partner's behavior in the past rounds.

2. *Indirect Reciprocity (IR)*: Each child was matched with a different child in each round as in CTR. Yet, before the round started, a child learned the behavior of the current partner in each of all previous rounds (from the second round onwards). This was illustrated like in area A1 of Figure A1, and the current partner's previous behavior was also communicated via headphones. The child was also aware that the current partner was informed of the child's own behavior in each of the previous rounds.
3. *Direct Reciprocity (DR)*: Here, a child was matched with the same partner for all five rounds (with the partner's avatar also staying the same in all rounds). Like in IR, the child was reminded of the partner's behavior in each of the previous rounds through visual representation and an audio message. Moreover, a child was informed that the partner was reminded of the child's own behavior in each of the previous rounds.
4. *Third-Party Punishment (TPP)*: As in CTR, children were matched with a different child in each round by implementing a perfect stranger matching. Children received no information about their current partner's past actions. Here, a third, anonymous, character was introduced, as illustrated in area A2 of Figure A1. This character had one token available to either keep it or throw it into a box (but then lose it). Throwing the token in the box was only possible if at least one of the children defected. This represented third-party punishment, and it resulted in the loss of all tokens that a defecting child earned in that round (this rule applied both when a single child defected, but also when both children defected, in which case both children lost all tokens). This was accompanied by an animation where the tokens of a defecting child visually broke in half and vanished without entering the child's wallet. Hence, the stage-game payoff also depended on the punisher, and in case of punishment, it was always zero for a child who defected. Note that children playing the PD game were informed that third parties could keep their token or throw it into the red box, the latter meaning that defecting children lost their earnings. Since third-party punishment was costly, selfish third parties should not implement punishment.

   We decided to restrict punishment to only defection, in line with Lergetporer et al. (2014), for several reasons. First, it represents what young children experience in this stage of their life, meaning that they are hardly ever exposed to punishment when cooperating. Second, studies with adults reveal that cooperative behavior, in contrast to defection, is rarely punished (Fehr and Gächter, 2002, Fehr and Fischbacher, 2004,

Herrmann et al., 2008), implying that such punishment is generally neither taught nor encouraged through the upbringing. Third, allowing for children of such delicate age to punish cooperation (and cooperation to be punished) would have raised ethical concerns, as it might have taught children to be less cooperative in the future.

The decisions of third parties were collected at the end of previously run sessions of the other treatments. A subset of children who had finished playing in the CTR, DR, or IR treatment were asked to make one final decision. This was a surprise, meaning that these children did not know about this final decision when they played the PD game themselves. The RA explained that the child would receive one additional token which he/she could either keep or throw into a box. The screen then showed two new avatars playing the PD as previously experienced by the child (so the child knew exactly the rules of the game). He/she was made aware that these were two children playing the same game he/she had just played, and that they were from a different kindergarten. Then, the RA explained that throwing the token into the box meant losing it, but also that defecting children in the PD lost their tokens from the respective round. After two trials, the child was asked to make the actual decision. These decisions were collected, and we randomly assigned them to pairs of children playing in the TPP game, where one decision was matched to one round (see Appendix A.1.5 for more details). This also means that each pair of children had a different third party in each round. They were aware of this, and they saw the third-party avatar change each round which highlighted this aspect. This was a deliberate design choice in order to avoid repeated interaction of children with the same third party. Only in this way we can avoid any learning effects, such when, for example, players in the cooperation game adapt their behavior when they experience that the third party with whom they would be matched repeatedly would never exert any punishment.

The cooperation game was run as follows: Six children were accompanied by six RAs to a separate room where one of the RAs in the room asked the children to form a circle. Children were then shown the first screen of the program installed on the tablet. It showed six different avatars, and the RA explained that each child in the room would be represented by an avatar in the game. Moreover, the children were aware that they would play with all the other children (or one other child in treatment DR) from the circle, but that the true identity would remain disguised as an avatar.

Explaining this while being in a circle provided a natural learning experience with which all children were familiar. The instructions stressed *(i)* the matching mechanism, *(ii)* the fact that children would play anonymously with other children in the game, and *(iii)* that all children were participating in the same game. After this part, each of the six RAs introduced him- or herself to one child, and each pair sat down at a certain spot within the room (illustrated in Panel B of Figure A1). The spots were sufficiently distanced between the pairs and were additionally separated by partitions to minimize distractions. Each RA continued with the instructions of the game by explaining each screen and action. Children navigated the buttons throughout the instructions to familiarize themselves with the tablet. The positioning (left or right) of the two main decision buttons – cooperate and defect – was randomized across subjects. To make sure that children paid attention and could follow the instructions, RAs asked them follow-up questions to repeat parts of the instructions in their own words or answer questions for certain scenarios.

Before starting the game, children played four pre-specified trial rounds with four different animal avatars (or one animal avatar in treatment DR) as practice partner(s). During these trial rounds, children received audio messages through headphones, which instructed them on the action they should take (either keep one token or give two tokens to their partner). After each round, the program stopped and children physically had to allocate the tokens earned in that round to their own wallet and the wallet of a fictitious partner. The payoff scenarios consisted of *(i)* defect-defect in trial round 1, *(ii)* cooperate-defect in trial round 2, *(iii)* defect-cooperate in trial round 3, and *(iv)* cooperate-cooperate in trial round 4. In treatment TPP, four different monkey avatars served as third parties across the four trial rounds. They punished defection in 2 out of 4 rounds. The tokens earned in the trial rounds were exchanged for small stickers to mimic the exchange of tokens into (more valuable) presents after the real game.

Trial rounds had multiple purposes. First, they served as a key element in learning the game. Second, children experienced all scenarios and tried out all buttons of the game, deterring them from taking an action out of sheer curiosity about the associated animation. Third, and most importantly, the trial rounds served as a check for comprehension of the game. If the children allocated all tokens in trial rounds 3 and 4 correctly, we classified this as a successful comprehension check, i.e., as successfully passing the control questions. Note that if children did not correctly distribute tokens in the first two trial rounds, the RA corrected them and explained why their allocation was not correct. If children failed to distribute tokens correctly in trial rounds 3 or 4, they were asked to repeat the token allocations to their own and their fictitious partner's wallet once more and to explain their reasoning. Incorrect allocations were

then classified as failing the comprehension check. In that case, they continued with the game but were excluded from the main analyses. Subjects who passed the comprehension check form our main sample for the analyses. Altogether 814 children (88%) passed it successfully. While this constitutes a very large fraction of our participants, it nevertheless raises the question of potential selection effects. Hence, we additionally engage in several robustness checks where we *(i)* correct for selection concerning passing the control questions by reweighting our dataset, and *(ii)* employ a less stringent comprehension criteria by removing any exclusion criteria based on comprehension. Moreover, to ensure that our comprehension checks were not too lenient, *(iii)* we employ a more stringent comprehension criterion in which we utilize children's answers to an RA's follow-up questions to build a stricter variable of comprehension. We repeat our entire analysis for each of these three approaches and find that the results presented in the next section remain robust. We report the results of these checks in Appendix A.1.

Overall, our highly standardized approach with a child-friendly and animated computer interface on tablets, the automated messages over headphones, and the well-trained experimental helpers (who had two weeks of training before going to the field) ensures that children understood the experiment very well and that they kept their attention span. Experimental instructions, a video sequence of the animated implementation of the game for all treatments and a translated transcription of the (German) audio instructions is available on https://osf.io/y7h4v/?view_only=f4627bf3ffe347e085f48144b19451e5.

When designing our experiment, we pre-registered several hypotheses on the Open Science Framework (https://osf.io/th25y/?view_only=6ff7e93a9b4c481d963dae17cec3e9d9). We hypothesized positive effects of IR, DR, and TPP on cooperation rates, in comparison to CTR. We also hypothesized that cognitive abilities, theory of mind, and patience will be positively related to cooperation. With respect to age, we expected either a stable or increasing level of cooperation, but not a decreasing one. Moreover, we expected that age will either positively interact with the three treatment conditions (such that treatment effects become larger with age) or that we will observe no interaction effects, but not that we will observe negative interaction effects. While we indicated to also collect information from parents, we did not pre-register any hypotheses about the influence of parental background on children's cooperation.

2.4 Parental questionnaire

We complement our experimental data of children with a survey among their parents. At the end of day 1, each child received a printed questionnaire to take home to his/her parents. The questionnaire (fully reprinted in the Appendix) already contained a child's anonymous

code, so that the parents' answers could be matched to the child's decision, while all answers remained anonymous. The questionnaire included items on sociodemographics, most importantly the highest parental educational level for each parent, and a question on parenting style by asking about parental warmth (degree of agreement on a scale from 1 to 5 with the statements "I show my child with words and gestures that I like him/her" and "I praise my child."; see Falk et al., 2021). Out of our main sample, 80% of parents returned the questionnaire with completed information on educational attainment and parental warmth.[9] We construct the socioeconomic status variable as the highest educational attainment between the two parents. Here, we follow the European Qualifications Framework and assign values from 1 to 8, with 1 being the lowest educational level (compulsory schooling) and 8 the highest (PhD). The average highest parental education across our main sample is coded as 5.66 (which is roughly the level of a Bachelor's degree; standard deviation 1.54; see Table A1 in the Appendix). To calculate the variable of parental warmth, we sum up the two answers on the parenting style questions, yielding an average of 9.64 (standard deviation 0.74). Note that this variable predominantly captures the mother's parental warmth, as mothers answered the questionnaire in 93% of cases.

3. Results

3.1 Cooperation across treatments

Figure 1 presents the average cooperation rates across all five rounds. For CTR, IR, and DR, we note very similar rates, ranging from 24% in IR to 29% in CTR, suggesting that the possibility of reputation formation in IR and DR is not sufficient to increase cooperation rates in the aggregate. This is in stark contrast to the effects of third-party punishment. In TPP, the overall average cooperation rate of 68% is more than double the rate of any other treatment. We compare the cooperation rates across treatments in the regressions shown in Table 3. Column 1 is the most basic regression where we regress a subject's cooperation decision on three treatment dummies, taking CTR as the omitted category. Column 2 applies kindergarten fixed effects, and columns 3 and 4 add additional control variables. Across all columns, it stands out that cooperation rates in TPP are estimated to be about 40 percentage points higher than in CTR. For IR and DR, we mainly see insignificant coefficients in comparison to CTR, and the coefficients are significantly smaller than in TPP.

[9] As 80% of parents returned the questionnaire, we additionally implemented a robustness check where we reweight our dataset to correct for observed imbalances regarding who returned the questionnaire. Our results remain robust in this check.

Figure 1 and Table 3 about here

Result 1: *In comparison to CTR, third-party punishment (TPP) increases cooperation rates significantly (by more than doubling them). This is in line with our expectations. The two treatments with an opportunity for reputation building (DR and IR) fail to yield cooperation rates above the level in CTR, which we did not expect.*

3.2 Cooperation across rounds

Figure 2 breaks down cooperation rates by round and treatment. Again, we see that CTR, IR, and DR lie in a narrow range, while cooperation rates in TPP are much higher in every single round. We first focus on the very first round, which is stripped from any experience from previous rounds. For TPP, round 1 cooperation rates indicate whether the mere existence of the punishment threat – *before* ever experiencing actual punishment – helps to increase cooperation. This is obviously the case, given a cooperation rate of 64% in the first round in TPP. Similarly, for IR and DR, round 1 data tell us whether the introduction of the reputation-building opportunity – without yet observing any signal about a partner's reputation – is sufficient to affect children's behavior. This is not the case, as Figure 2 shows, since cooperation rates in the first round are practically identical in CTR, IR, and DR (ranging from 21% to 23%). In Table A2 in the Appendix, we show a regression that uses only first-round data, confirming that TPP, but not DR or IR, increases cooperation rates significantly. This means that changes in the information structure (in DR and IR) are not sufficient to change the cooperation rates, but changes in the game (by adding a third party in TPP) are (despite the fact that under standard assumptions behavior should not differ between these different conditions).

Figure 2 and Table 4 about here

Next, we look more closely at the dynamics of cooperation across rounds. Figure 2 reveals a slightly decreasing trend for CTR, IR, and DR, and a modest upward trend for TPP. Table 4 shows that these trends are significantly negative for CTR and IR (see columns 1-4) and positive for TPP (columns 7-8). In columns 9 and 10 of Table 4, we can examine how the development of cooperation across rounds (which is generally negative, see the main effect of the variable Round) interacts with the three treatment conditions. To calculate the interaction effects in non-linear models, however, one needs to be careful, as calculating only the marginal effect of the interaction term can yield the wrong magnitude, sign, and significance level (Ai

and Norton, 2003). Hence, to calculate the interaction effects in our non-linear models, we apply the methodology of Norton et al. (2004), which corrects for the issue using cross-partial derivatives and calculating the interaction effect and the significance level at each observation. Following this approach, we inspect the interaction effect and the significance level across all observations, and we summarize the findings in the notes to the tables.[10]

The results in columns 9 and 10 of Table 4 reveal that the difference in cooperation rates between IR and CTR does not change across rounds. The DR treatment effect reveals a small and significant increase across rounds (at the 10% level in column 9); however, this is not robust to including control variables (in column 10). Turning to TPP, we observe that, with each increasing round, the treatment effect of TPP is significantly increasing by 3.3% (or 3.5%, with controls), on average. Taken together, this means that TPP is successful in reversing the decreasing pattern of cooperation in CTR.

Result 2: *Looking at behavior in the first round, we observe that the anticipation of potential punishment increases cooperation rates strongly, while the anticipation of a reputation-building opportunity – both in DR and IR – has no significant effect in comparison to CTR. Across rounds, cooperation rates are generally slightly downward-trending. Yet, this decline is reversed successfully in TPP, where cooperation rates increase with repetition.*

3.3 Sources for the dynamics of cooperation: reaction to punishment and a partner's reputation

The development of cooperation across rounds differs between DR and IR on the one hand, and TPP on the other hand. In order to understand these differences in dynamics better, we continue by examining how children react to third-party punishment or a partner's reputation.

From the observation that cooperation rates are not higher in DR and IR than in CTR, one could conclude that the possibility of reputation building and the information about a partner's past behavior in DR and IR do not have any effect on children's behavior. However, such a conclusion would be wrong. To show this, we first calculate the image score of a child (Nowak and Sigmund, 1998). This score sums up the actions of all previous rounds in the following way: cooperation increases the score by 1 point, whereas defection reduces it by 1 point. Thus, the image score represents a measure of one's reputation as being more or less cooperative. We use the image score as the main measure for analyzing the reaction to the

[10] Note, however, that our reported results are also robust to using a linear probability model.

partner's reputation, yet we also show the robustness of our findings using a different measure in Appendix A.1.4.

Figure 3 and Table 5 about here

In Figure 3, we show the predictions of a regression model where we regress a child's likelihood to cooperate on the current partner's image score, separately for IR (left panel) and DR (right panel). The regression estimates are presented in Table 5. We observe a very clear pattern. A child's likelihood to cooperate is significantly positively related to the current partner's image score. This holds true for all specifications in DR, and in 3 out of 4 specifications in IR. Comparing the coefficients across the two treatments, we observe that the slope of the relation is slightly steeper in DR (with estimated coefficients around 5% in Table 5) than in IR (with estimated coefficients between 1.7% and 3.8%), but this difference is not significant most of the time.

Interestingly, controlling for a partner's behavior from the last round *only* is not significant in DR (columns 7 and 8 of Table 5), nor does adding it affect the size of the partner's image score coefficient. This finding suggests that children do indeed care about the entire history of their partner, and not just about how a partner acted in the round before.[11]

Taken together, these results provide the first evidence that the reputation mechanism already affects the behavior of 3- to 6-year-olds by generating patterns of conditional cooperation. Table A3 in the Appendix provides additional support for this finding. It shows that pairs of children who cooperate in round 1 have significantly higher cooperation rates in subsequent rounds than pairs of children who start with defection. This holds at the 5% level in DR and at the 10% level in IR. Not surprisingly, it does not apply to CTR, where information about a partner's past behavior is missing.

Next, we investigate the dynamic pattern caused by the punishment tool in TPP. In Figure 4, we show the average cooperation rate in round t for subjects who defected in round $t-1$, where $t \in \{2, 3, 4, 5\}$. The left bar represents those defectors who did not get punished by the third party in round $t-1$, while the right bar represents those who experienced punishment. We see a clear difference between the two groups. While only 34.9% of unpunished defectors switch to cooperation in the next round, 60.7% of them do so if they got

[11] As a kind of placebo test, we can show that neither a partner's image score (over all previous rounds) nor a partner's last round behavior have any significant effect on cooperation rates in the control treatment (CTR) ($p > 0.2$ in all regressions). This was to be expected since subjects are not informed about their partner's previous behavior in CTR.

punished. The observed difference is statistically significant and it is robust to controls (see Table A4 in the Appendix). Thus, the overall evidence indicates that the punishment mechanism not only functions through the anticipation of potential punishment (see *Result 2*), but also through the execution of punishment itself.

Figure 4 about here

Result 3: *Although DR and IR do not increase average cooperation rates, 3- to 6-year-old children do already condition their behavior systematically on the reputation of their partners. The more cooperative a child's partner has been in previous rounds, the higher the probability that a child will cooperate, indicating conditional cooperation already at this young age. In TPP, experiencing punishment increases the probability of cooperation in the next round, thus improving the chances that defectors turn into cooperators.*

Before proceeding with further analyses of the determinants of cooperation, we briefly address the behavior of third parties. Overall, we see a relatively high rate of punishment of defection (note that each third party made only one punishment decision). In fact, 55% of children in the punisher's role decided to punish defectors (at their own cost of losing 1 token). Recall that the role of the third party was played by children who had played the PD game themselves before in one of the treatments CTR, DR, or IR. We can show that there is a significantly positive raw correlation between the likelihood to punish defection as a third party and the number of times a child cooperated in the PD game (Pearson's $r = 0.12$; $p = 0.045$). Similarly, we find that experiencing more cooperative behavior from one's partner(s) across the 5 rounds is correlated with a higher likelihood to punish defection (Pearson's $r = 0.14$; $p = 0.026$).

3.4 Payoffs across treatments

After having examined how our treatments affect cooperation rates, we ask the question whether cooperation is profitable on an individual basis. To this end, we first regress a child's round payoff on cooperation, as done in Table 6. For CTR, IR, and DR, we see that defection predicts higher profits. This is not very surprising, since defection strictly dominates cooperation in terms of payoffs. Yet, in TPP, the relation changes. There we observe that cooperation is significantly more profitable than defection, and given these relations, children clearly earn much more when switching from defection to cooperation in TPP than in CTR (see

column 5 in Table 6). While purely money-maximizing third parties would not want to spend their token on punishing defectors, the relatively high frequency of punishment of 55% makes defection unattractive and unprofitable in TPP.

Table 6 about here

Result 4: *With the exception of TPP, children who cooperate earn fewer tokens than children who defect. In TPP, this relationship is reversed, because the relatively high likelihood of punishment makes defection less profitable.*

3.5 Determinants of cooperation

The role of age. After analyzing cooperation rates across treatments, we now investigate how children's characteristics relate to cooperation. We start by analyzing the role of age in the regressions in Table 7. In columns 1 to 8, we look at the four different treatments. In CTR, IR, and DR, age (in months) is negatively related to the likelihood of cooperation (contrary to our expectations), although the coefficients are not always significant. In TPP, age is positively and significantly related to cooperation, showing that older children cooperate more than younger children.

In columns 9 and 10 of Table 7, we look in more detail at the interaction effects of age with our treatments. For DR, we note a significant positive interaction in column 9, which suggests that older children have an easier time realizing that cooperation can pay off when direct reciprocity is possible; however, this finding is not robust to adding control variables (in column 10). In TPP, this interaction is much more pronounced and statistically robust, indicating that as children get older they are more likely to cooperate in TPP than in CTR.

Table 7, Table 8, and Table 9 about here

The role of cognitive abilities, time preferences, and theory of mind. Next, we examine the relation between children's traits elicited on day 1 – cognitive abilities, theory of mind, and patience – and cooperation on day 2. We expected positive effects in our pre-registration. Across columns 1 to 8 in Table 8, we note that higher cognitive abilities let children cooperate less in our control condition CTR, and more in the TPP treatment, although none of these effects are robust to controls. For theory of mind, we find a weakly significant positive effect on

cooperation in TPP; yet, this effect also vanishes with the addition of further controls. Finally, we observe that patience is positively related to cooperation in TPP.

When we look at interaction effects of cognitive abilities, theory of mind, and patience with our treatment conditions (in columns 9 and 10), we observe that for all three of our treatments – IR, DR, and TPP – higher cognitive abilities imply a significantly higher difference in cooperation between the respective treatment and CTR (in column 9). The interaction effects remain significant in DR and TPP when adding further controls (in column 10). A one-standard-deviation increase in cognitive abilities is estimated to increase the difference in cooperation rates between CTR and DR by 5.3%, and between CTR and TPP by even 8.8% (column 10). Turning to patience, we observe no significant interaction effects between patience and any of our treatments. Concerning theory of mind, only TPP shows indications of positively interacting with theory of mind, although the interaction effect lacks robustness.

Socio-economic status of parents and parental warmth. Finally, we look at how the family environment of a child is related to cooperative behavior. Table 9 presents regression results for how a child's cooperation relates to parental warmth and parents' socioeconomic status (SES; as measured by their highest level of education). From columns 1 to 8, which consider each treatment separately, we note that parental warmth is hardly ever significantly related to cooperation rates. For SES, we see a clearly positive effect on the cooperation rates in TPP, but a negative one in the other treatments, although the latter is typically not significant.

In columns 9 and 10, we look again at the interaction of our treatments with the variables on which we focus here. We observe that parental warmth does not interact with any of the treatments. In contrast, we find evidence that SES interacts with TPP. In particular, a one-step increase in the highest education level of parents (e.g., from high-school certificate to a Bachelor's degree) is associated with a 6.8% higher cooperation rate in TPP in comparison to CTR (column 10). These results also hold when we take into account potential selection issues regarding obtaining the SES and parental warmth variables from parents (See Appendix A.1.2.)

Result 5: *Cooperation decreases with age in IR, while it increases with age in TPP. Moreover, the TPP treatment effect is increasing with age. Children with higher cognitive abilities have significantly higher cooperation rates when exposed to DR and TPP (in comparison to CTR). More patient children cooperate more in TPP. Theory of mind hardly ever matters. Children from families with higher education are more likely to cooperate in TPP. Parental warmth does not play a noticeable role in our sample.*

4. Putting our findings into perspective beyond childhood – Insights from a meta-analysis

Our results reveal that punishment strongly increases cooperation in 3-6 year-old children, however no such effect is observed when a reciprocity mechanism is introduced. While it is known that reciprocity is effective in adults, it is unknown for the adult age whether punishment still has a relatively larger effect on cooperation, compared to direct and indirect reciprocity. In this section, we address this question, because it can help uncover whether the patterns we found for young children persist later on as well. Additionally, there is a second question regarding the developmental stage in life when reciprocity mechanisms become effective in fostering cooperation. Our experiment has not found an effect in the aggregate, while for adults it seems that reciprocity can increase cooperation rates, yielding the speculation that older childhood and adolescence might be candidate periods for when reciprocity starts to be effective.

To answer our questions, we take two steps. First, we conduct a large meta-analysis about whether punishment, direct and indirect reciprocity affect cooperation behavior in adults. Second, to better understand the transition from childhood to adulthood, and how this interacts with the effectiveness of the different pillars, we conduct an additional literature survey where we focus on cooperation studies with older children and adolescents.

4.1 Search for studies, inclusion criteria, and coding of cooperation for the meta-analysis on adult behavior

Studies for the meta-analysis were selected from all the records included in the Cooperation Databank (CoDa; Spadaro et al., 2022). CoDa contains an archive of all studies on human cooperation reported in published articles, working papers, dissertations, theses, and book chapters written in English, Japanese, and Chinese until 2017.

To be included in our meta-analysis, a study had to meet the following criteria: *(i)* the study employed either a PD game or a public-goods game, *(ii)* participants were adults, *(iii)* no deception was used in the study, and *(iv)* the study manipulated (at least) one of the three mechanisms of our interest. Regarding *(iv)*, we incorporate studies which correspond to broad definitions of our three mechanisms of interest. In particular, the *goal of our meta-analysi*s is not to narrow down on studies that use identical design choices as in our experiment with children, but to get a comprehensive and generalizable overview of how punishment and reciprocity mechanisms affect cooperative behavior based on the findings of the existing

literature. Thus, we take full advantage of the large and comprehensive dataset of studies recorded in CoDa.

To study the effects of punishment on cooperation, we include all studies which, in various designs, introduce a punishment mechanism (in terms of the possibility of imposing negative payoffs on a specific participant after the regular round of play) and compare it to behavior when no such mechanism exists.[12] To study the effects of indirect reciprocity on cooperation, we include all studies which manipulate indirect reciprocity in one of the following ways: *(i)* manipulating subjects' anonymity in front of others that are not paired with the subject in the game, *(ii)* introducing the possibility to select partners based on reputation or identifiable information, or *(iii)* introducing the possibility to gossip. To study the effects of direct reciprocity on cooperation, we analogously include all studies which, in various designs, manipulate direct reciprocity mechanisms in one of the following ways: *(i)* manipulating the extent of one's interaction with the partner (e.g., interacting before the game), *(ii)* manipulating the extent of one's anonymity towards the partner(s), and *(iii)* manipulating the visibility of one's contribution towards the partner(s).

Altogether, this results in 105 studies with 264 distinct effect sizes, i.e., 264 distinct comparisons between a treatment where a mechanism of our interest is present and a treatment without it. Out of those, 187 effect sizes refer to the punishment mechanism, 19 effect sizes to the indirect reciprocity mechanism, and 58 effect sizes to the direct reciprocity mechanism.

The outcome variable we analyze is the effect size in cooperation between a control treatment and a treatment where the particular mechanism was introduced. We capture this with Cohen's *d*, which is a standardized measure of the difference in the outcome variable between two groups. It is calculated by $(\bar{X}_1 - \bar{X}_2)/s_{pooled}$, where $\bar{X}_1$and $\bar{X}_2$ are the means of the two treatments, and s_{pooled} their pooled standard deviation. If the measurement of cooperation is binary, we calculate Cohen's *d* by using the approach of Chinn (2000). In particular, we convert the odds ratio $\frac{p_1 n_1 (1-p_2) n_2}{p_2 n_2 (1-p_1) n_1}$ to Cohen's *d* using the following formula: $\frac{\text{LN(odds ratio)}}{\pi/\sqrt{3}}$, where p_i captures the cooperation proportion of treatment i, and n_i the sample size.

4.2 Results of the meta-analysis

For each mechanism – punishment, direct reciprocity, and indirect reciprocity – we meta-analyze the mean differences in cooperative behavior between the treatment where the

[12] Here we do not distinguish between second-party punishment (where players in the game can punish each other) and third-party punishment (where only external third parties can punish the players in the game). See Leibbrandt and López-Pérez (2012) for a comparison of both types of punishment.

mechanism is introduced and the corresponding control treatment. To take into account that more effect-sizes can come from a single study, we run a multilevel meta-analysis with study as a random intercept. We first meta-analyze the 187 treatment effects of the punishment mechanism. Our results reveal a Cohen's *d* of 0.62, i.e., a treatment effect of 0.62 standard deviations. The effect is significantly different from 0 ($p<0.01$), with a confidence interval of [0.48, 0.75]. We report forest plots showing the estimates of each individual effect size and the overall effect size in Figures A2 to A5 in the Appendix. Next we look at the meta-analysis of 58 treatment effects of the direct reciprocity mechanism. We observe a Cohen's *d* of 0.40, i.e., a treatment effect of 0.40 standard deviations. The finding is again significantly different from 0 ($p<0.01$), with a confidence interval of [0.19, 0.61]. We report forest plots of the individual effect sizes in Figures A6 and A7. Finally, we look at the meta-analysis of the 19 treatment effects of the indirect reciprocity mechanism. We observe a Cohen's *d* of 0.18. The finding is significantly different from 0 ($p<0.01$), with a confidence interval of [0.03, 0.33]. We report a forest plot of the individual effect sizes in Figure A8.

Taken together, the results indicate that when looking at the 105 studies used in our meta-analysis, all three mechanisms cause a positive effect on the level of cooperation. However, we observe a very clear ordering where the punishment mechanism causes the largest effect ($d = 0.62$), followed by direct ($d = 0.40$) and indirect reciprocity ($d = 0.18$). This finding indicates that reciprocity mechanisms are not as strong in affecting cooperative behavior as punishment, which is therefore largely in line with the findings from our experiment with children. In our experiment, we observe that the TPP treatment causes a large positive effect: Cohen's $d = 0.99$. In contrast, the two reciprocity treatments cause no significant shift in behavior (and in terms of size, the effects are negligible or small: Cohen's $d = -0{,}06$ for DR and -0.19 for IR). The findings further suggest that the stark difference in effects we observe between our punishment treatment and the reciprocity treatments cannot be explained solely by reciprocity mechanisms developing later in life, as the gap in the punishment and reciprocity mechanisms does not fully close with age. Instead, it appears that the reciprocity mechanisms affect cooperative behavior to a lesser extent, even after they do develop. In the next subsection we therefore deal with the question when reciprocity mechanisms become effective in raising cooperation, but first we summarize the insights from our meta-analysis.

Result 6: *Our meta-analysis shows for adult subjects that all three mechanisms – direct and indirect reciprocity and punishment – cause a significant and positive effect on cooperative behavior. Punishment causes the largest effect, which is followed in size by direct reciprocity,*

and then indirect reciprocity. We see from our experiment with 3-6 year-olds that punishment has the largest effect on cooperation, suggesting that this pillar is strongest across age and that the relative advantage of punishment for cooperation develops very early on.

4.3 Searching for the age when reciprocity mechanisms start improving cooperation

To gain a better understanding of the ontogeny regarding the effects of our three mechanisms of cooperation, we next turn our focus to related studies with children and adolecents. As argued before, the age of the children in our experiment is very young, which pushes the boundaries regarding age in current studies on cooperation. Looking at related studies that run experiments with older children can then inform us about the behavioral patterns when it comes to the gap between very young children (our main experiment) and adulthood (the meta-analysis). This allows us to study whether the effect of punishment on cooperation is always as strong as we observe it in our experiment and in our meta-analysis. Further, we can search for the period in life when reciprocity mechanisms become effective. Given that they are positive in adulthood, but ineffective in 3-6 year olds, it seems reasonable to expect reciprocity mechanisms to become effective in the age period in between young childhood and adult age.

To this end, we search the literature for studies that investigate cooperation and introduce one of the mechanisms of our interest. We focus on studies that *(i)* use either a PD or a public goods game, *(ii)* have underage participants playing with other underage participants, and *(iii)* use no deception. We find five studies matching these criteria.[13]

First we focus on the punishment mechanism. In a well-powered study, Lergetporer et al. (2014) investigate the cooperative behavior of 7-11 year old children using a PD game. They compare a treatment without any punishment to one that has third-party punishment. They observe that the existence of punishers significantly increases cooperative behavior. The Cohen's *d* of this change in behavior is 0.80, which constitutes a large effect. This observation fits nicely to the findings from our main experiment with very young children (d=0.99) as well as our meta-analysis for adults (d=0.62), suggesting that a punishment mechanism consistently causes a large positive effect on cooperative behavior, whether it is with 3-6 year-olds, 7-11 year-olds, or adults.

Next, we turn to the reciprocity mechanisms. We find altogether four studies that fit our criteria. In one study, the authors exogenously introduce a direct reciprocity mechanism and compare it to a control treatment (Blake et al., 2015). The authors show that 10-11 year-olds

[13] Where appropriate, effect sizes were calculated with information reported in the respective papers, or if some information was missing, we contacted the authors.

increase cooperative behavior when they are always paired with the same partner in contrast to switching partners in a PD game. They observe an effect size of 0.31 (Cohen's *d*). This size is smaller than the one punishment causes with similarly aged children (0.80) in Lergetporer at al. (2014), and is close to what we find for the direct reciprocity mechanism in our meta-analysis with adults ($d = 0.40$). Given that we had not found an effect of reciprocity for 3-6 year olds, but Blake et al. (2015) found one for 10-11 year olds (at least for direct reciprocity), both studies together suggest that it might be the age range between 6 and 10 where reciprocity mechanisms start improving cooperation levels.

Three other studies we identify do not have a comparison between a control treatment and a treatment with reciprocity, yet they are still informative when compared to our experiment. Harbaugh and Krause (2000), Vogelsang et al. (2014), and Hermes et al. (2020) all conduct public goods games with repetition and with fixed groups (thus, with direct reciprocity mechanisms involved). As there is no comparison group, we cannot calculate any effect size, yet we can look at the proportion of cooperation in these studies. Vogelsang et al. (2014) find that 5-6 year-olds cooperate in 24% of cases, which is very close to 28% that we observe in our DR treatment. Hermes et al. (2020) let children choose how many out of 5 coins they want to contribute to the public good, and find that 6-year old children on average contribute 38% of their endowment. Then moving to older children and adolescents, Harbaugh and Krause (2000) find that 6-14 year old children on average contribute between 42% and 64% (depending on the round), thus showing an increasing trend in the level of cooperation in comparison to studies with younger children. These observations are also in line with the conjecture that reciprocity mechanisms roughly start to improve cooperative behavior within primary school age.

Result 7: *Taking together the results of our experiment, our meta-analysis, and the survey about studies with children and adolescents, we find that introducing a punishment mechanism always causes a large effect on behavior, whether it is with children or with adults. Reciprocity mechanisms (at least those of direct reciprocity) have been found to increase cooperation levels around age 10, suggesting that they develop their effectiveness around primary school age (6-10 years of age).*

5. Conclusion

The ability of humans to cooperate with genetically unrelated strangers is remarkable. From an economic perspective, it allows for increasing the efficiency, and thus the welfare, created in human interactions in both small and large groups. While it is obvious that

cooperation is necessary to meet crucial human challenges, they will be met more easily the better we understand the roots of human cooperation, i.e., when cooperation emerges and under which conditions it can be expected to flourish. For these reasons, we studied cooperation of more than 900 children, aged 3 to 6 years. Based on earlier insights that childhood is the most formative period for a human being's skills and behavior (Heckman, 2006; Fehr et al., 2008, 2013; Almås et al., 2010; Bauer et al., 2014; Alan and Ertac, 2018; Sutter et al., 2019; Berger et al., 2020; Brocas and Carillo, 2020, 2021; Cappelen et al., 2020; Castillo et al., 2020; Hermes et al., 2020; Kosse et al., 2020; Alan et al., 2021; List et al., 2023), our goal was to improve our understanding of when cooperation evolves in young children and what the institutional as well as personal, respectively parental, prerequisites are for this.

In our experiment, we designed a unified framework to examine which of three fundamental pillars of human cooperation – direct and indirect reciprocity, as well as third-party punishment – emerges earliest and is most effective in increasing cooperation in a repeated PD game. These three pillars have never been compared to each other in a unified framework with respect to their potential to increase cooperation in comparison to a baseline condition, neither with adults nor with children. While each of them has been shown separately to affect cooperation (e.g., Axelrod and Hamilton, 1981; Fehr and Fischbacher, 2004; Bolton et al., 2005), the lack of a unified setting has made it difficult to compare their relative effects. The relative comparison of these mechanisms is particularly relevant in early developmental phases, as it can inform researchers about the developmental trajectory of each mechanism, and practitioners interested in shaping cooperative tendencies in young children.

We found that cooperation rates are modest (around one quarter) in a control condition and in the two treatments that allow for reputation building by informing children about their partner's past behavior (which makes indirect and direct reciprocity possible). Compared to the control condition, reciprocity does not increase cooperation rates. Yet, this implies by no means that children as young as the age of 3 would not be able to apply reciprocal strategies. On the contrary, we are the first to observe a pattern of conditional cooperation (Fischbacher et al., 2001) already at this young age, because we provide clear evidence that children react systematically, and positively, to a partner's past history of cooperative behavior. The likelihood of children to cooperate increases when they meet a partner who has been more cooperative in the past, either with them or with other (anonymous) children. Given the modest level of cooperation at the beginning of the repeated interaction, however, such reciprocal strategies are not capable of lifting cooperation rates beyond what we observe in the control condition.

Cooperation levels are only increasing in this young age group when a third party that is unaffected by the outcome of the PD game enters the stage. With potential third-party punishment, cooperation rates skyrocket, one could say, by almost tripling in comparison to the other conditions. This effect prevails from the very first round onward, even before any punishment could have been applied. This means that children are able to anticipate that in the presence of a third party it is better to cooperate rather than to defect; and in fact cooperation pays off monetarily in the third-party punishment condition (only). Yet, executing (costly) punishment also serves a purpose by making it more likely for former defectors to turn into cooperators. Thus, both the third party's existence as well as his/her actions matter for the higher cooperation rates in this condition.

These findings provide valuable insights for practitioners and policy makers that aim to promote cooperation among future generations. To increase cooperation, our results would support interventions that target children's willingness to object to selfish actions – an altruistic type of behavior children are undoubtedly familiar with (Gummerum and Chu, 2014; Mcauliffe et al., 2015; Bašić et al., 2020). Given the substantial effect on cooperation we observe in our study, such interventions, if successful, might have a remarkable potential for increasing the levels of cooperation. Our study does not offer the same support for targeting the mechanisms of reciprocity at that young age. As we show that children do reciprocate others yet fail to increase cooperation early on, it remains a question whether initial cooperation can be supported in some way so that it can be sustained through reciprocity. One potential way of tackling this could be an integration of both punishment and reciprocity mechanisms, where the early threat of punishment would elevate cooperation levels such that reciprocity mechanisms could sustain it in future encouters. We leave this question open for future studies.

In addition to studying and comparing the effects of direct and indirect reciprocity and of third-party punishment, we also look at how children's individual characteristics and their parents' background may affect cooperation rates. We found that the positive effects of third-party punishment on cooperation become even stronger with increasing age and also with increasing cognitive abilities. Presumably, children with higher cognitive abilities and older children might be better at understanding how to navigate in complex strategic environments and what to anticipate from third parties. Moreover, we find a similar relation of cooperation rates to parents' socioeconomic status. Children of better-educated parents, when exposed to the third-party punishment condition, react with a stronger increase in cooperation.

Importantly, we complemented our main experimental study with a literature survey on related studies with older children, and a large meta-analysis covering 105 distinct studies on

how punishment and reciprocity mechanisms affect cooperative behavior in adults. This helps in understanding whether our findings with young children also prevail in older age groups. We observe that the introduction of a punishment mechanism always causes a large positive effect on cooperative behavior, regardless if it is done with very young children, older children, or adults. The meta-analysis also reveals that reciprocity mechanisms cause an increase in cooperative behavior in adults. While we had seen patterns of conditional cooperation in 3-6 year olds in our study, both direct and indirect reciprocity had not increased cooperation rates in the aggregate in comparison to a control group. Screening the literature for children who are older than 6 years, but who have not reached adult age, we found suggestive evidence that reciprocity mechanisms start being supportive of cooperation around primary school age.

Our meta-analysis reveals a weaker effect of reciprocity mechanisms in comparison to punishment in adults. This mirrors the observation we see also with our young children, namely that punishment is more effective than reciprocity in promoting cooperation. It further suggests that the reason why we observe this finding with young children is not only because the effect of reciprocity mechanisms develop later than the punishment mechanism, but even when they do develop, they are weaker in the extent to which they affect behavior. So, the mechanisms which are at the roots of cooperation develop early in life, and their relative effectiveness seems to prevail into adulthood.

References

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80 (1): 123-129.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective-Taking." *Quarterly Journal of Economics* 136 (4): 2147-2194.

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865-1911.

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2010. "Fairness and the Development of Inequality Acceptance." *Science* 328 (5982): 1176-1178.

Axelrod, Robert, and William D. Hamilton. 1981. "The Evolution of Cooperation." *Science* 211 (4489): 1390-1396.

Bašić, Zvonimir, Armin Falk, and Fabian Kosse. 2020. "The Development of Egalitarian Norm Enforcement in Childhood and Adolescence." *Journal of Economic Behavior and Organization* 179: 667-680.

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka. 2014. "Parental Background and Other-Regarding Preferences in Children." *Experimental Economics* 17 (1): 24-46.

Berger, Eva, Ernst Fehr, Henning Hermes, Daniel Schunk, and Kirsten Winkel. 2020. "The Impact of Working Memory Training on Children's Cognitive and Noncognitive Skills." IZA Discussion Paper No. 13338.

Blake, Peter R., David G. Rand, Dustin Tingley, and Felix Warneken. 2015. "The Shadow of the Future Promotes Cooperation in a Repeated Prisoner's Dilemma for Children." *Scientific Reports* 5 (14559): 1-9.

Bolton, Gary E., Elena Katok, and Axel Ockenfels. 2005. "Cooperation Among Strangers with Limited Information About Reputation." *Journal of Public Economics* 89 (8): 1457-1468.

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences of the United States of America* 100 (6): 3531-3535.

Boyd, Robert, and Peter J. Richerson. 2009. "Culture and the Evolution of Human Cooperation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1533): 3281-3288.

Brandts, Jordi, and María Fernanda Rivas. 2009. "On Punishment and Well-Being." *Journal of Economic Behavior and Organization* 72 (3): 823-834.

Brocas, Isabelle, and Juan D. Carillo. 2020. “The Evolution of Choice and Learning in the Two-Person Beauty Contest Game from Kindergarten to Adulthood.” *Games and Economic Behavior* 120: 132-143.

Brocas, Isabelle, and Juan D. Carillo. 2021. “Steps of Reasoning in Children and Adolescents.” *Journal of Political Economy* 129 (7): 2067-2111.

Brüne, Martin, and Ute Brüne-Cohrs. 2006. “Theory of Mind-Evolution, Ontogeny, Brain Mechanisms and Psychopathology.” *Neuroscience and Biobehavioral Reviews* 30 (4): 437-455.

Cappelen, Alexander W., John List, Anya Samek, and Bertil Tungodden. 2020. “The Effect of Early-Childhood Education on Social Preferences.” *Journal of Political Economy* 128 (7): 2739-2758.

Castillo, Marco, John A. List, Ragan Petrie, and Anya Samek. 2020. “Detecting Drivers of Behavior at an Early Age: Evidence from a Longitudinal Field Experiment.” NBER Working Paper 28288.

Carpenter, Jeffrey P., and Peter H. Matthews. 2012. “Norm Enforcement: Anger, Indignation, or Reciprocity?” *Journal of the European Economic Association* 10 (3): 555-572.

Chinn, Susan. 2000. “A Simple Method for Converting an Odds Ratio to Effect Size for Use in Meta-Analysis.” *Statistics in Medicine* 19 (22): 3127-3131.

Clark, Kenneth, and Martin Sefton. 2001. “The Sequential Prisoner's Dilemma: Evidence on Reciprocation.” *Economic Journal* 111 (468): 51-68.

Costa-Gomes, Miguel, and Georg Weizsäcker. 2008. “Stated Beliefs and Play in Normal-Form Games.” *The Review of Economic Studies* 75: 729-762.

Cowell, Jason M., Anya Samek, John A. List, and Jean Decety. 2015. The Curious Relation between Theory of Mind and Sharing in Preschool Age Children. *PLoS ONE* 10(2): e0117947.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2012. “The Intergenerational Transmission of Risk and Trust Attitudes.” *The Review of Economic Studies* 79 (2): 645-677.

Duffy, John, and Jack Ochs. 2009. “Cooperative Behavior and the Frequency of Social Interaction.” *Games and Economic Behavior* 66 (2): 785-812.

Embrey, Matthew, Guillaume R. Fréchette, and Sevgi Yuksel. 2018. “Cooperation in the Finitely Repeated Prisoner’s Dilemma.” *The Quarterly Journal of Economics* 133 (1): 509-551.

Falk, Armin, Fabin Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021.

"Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences." *Journal of Political Economy* 129 (9): 2504-2545.

Fe, Eduardo, David Gill, and Victoria Prowse. 2022. "Cognitive Skills, Strategic Sophistication, and Life Outcomes." *Journal of Political Economy* 130 (10): 2643-2704

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach. 2008. "Egalitarianism in Young Children." *Nature* 454 (7208): 1079-1083.

Fehr, Ernst, and Urs Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25 (2): 63-87.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868): 137-140.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The Development of Egalitarianism, Altruism, Spite and Parochialism in Childhood and Adolescence." *European Economic Review* 64: 369-383.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence From a Public Goods Experiment." *Economics Letters* 71 (3): 397-404.

Gächter, Simon, Felix Kölle, and Simone Quercia. 2017. " Reciprocity and the Tragedies of Maintaining and Providing the Commons." *Nature Human Behaviour* 1(9): 650-656.

Gächter, Simon, and Elke Renner. 2010. "The Effects of (Incentivized) Belief Elicitation in Public Goods Experiments." *Experimental Economics* 13(3): 364-377.

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José Prados. 2020. "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program." *Journal of Political Economy* 128(7): 2502-2541

Gummerum, Michaela, and Maria T. Chu. 2014. "Outcomes and Intentions in Children's, Adolescents', and Adults' Second-and Third-party Punishment Behavior." *Cognition* 133(1): 97-103.

Harbaugh, William T., and Kate Krause. 2000. "Children's Altruism in Public Good and Dictator Experiments." *Economic Inquiry* 38 (1): 95-109.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900-1902.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6): 2052-2086.

Hermes, Henning, Florian Hett, Mario Mechtel, Felix Schmidt, Daniel Schunk, and Valentin

Wagner. 2020. “Do Children Cooperate Conditionally? Adapting the Strategy Method for First-Graders.” *Journal of Economic Behavior and Organization*, 179: 638-652.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. “Antisocial Punishment Across Societies.” *Science* 319 (5868): 1362-1367.

House, Bailey R., Joan B. Silk, Joseph Henrich, H. Clark Barrett, Brooke A. Scelza, Adam H. Boyette, Barry S. Hewlett, Richard McElreath, and Stephen Laurence. 2013. “Ontogeny of Prosocial Behavior Across Diverse Societies.” *Proceedings of the National Academy of Sciences* 110 (36): 14586-14591.

Keser, Claudia, and Frans Van Winden. 2000. “Conditional Cooperation and Voluntary Contributions to Public Goods.” *Scandinavian Journal of Economics* 102 (1): 23-39.

Kölle, Felix, and Lukas Wenner. 2023. “Time-Inconsistent Generosity: Present Bias Across Individual and Social Contexts.” *Review of Economics and Statistics* 105 (3): 683-699.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. “The Formation of Prosociality: Causal Evidence on the Role of Social Environment.” *Journal of Political Economy* 128 (2): 434-467.

Kosse, Fabian, and Michela M. Tincani. 2020. “Prosociality Predicts Labor Market Success Around the World.” *Nature Communications* 11, 5298.

Leibbrandt, Andreas, and Raúl López-Pérez. 2012. “An Exploration of Third and Second Party Punishment in Ten Simple Games.” *Journal of Economic Behavior and Organization* 84 (3): 753-766.

Lergetporer, Philipp, Silvia Angerer, Daniela Glätzle-Rützler, and Matthias Sutter. 2014. “Third-Party Punishment Increases Cooperation in Children through (Misaligned) Expectations and Conditional Cooperation.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (19): 6916-6921.

List, John A., Ragan Petrie, and Anya Samek. 2023. “How Experiments with Children Inform Economics.” *Journal of Economic Literature* 61 (2): 504-564.

Mai, Yujiao, and Zhiyong Zhang. 2016. “Statistical Power Analysis for Comparing Means with Binary or Count Data Based on Analogous ANOVA.” In: van der Ark L.A., Wiberg M., Culpepper S.A., Douglas J.A., Wang WC. (eds). *Quantitative Psychology*. IMPS 2016. Springer Proceedings in Mathematics & Statistics 196: 381-393. Springer, Cham.

McAuliffe, Katherine, Jillian J. Jordan, and Felix Warneken. 2015. “Costly Third-party Punishment in Young Children.” *Cognition* 134: 1-10.

Norton, Edward C., Hua Wang, and Chunrong Ai. 2004. “Computing Interaction Effects and Standard Errors in Logit and Probit Models.” *The Stata Journal* 4 (2): 154-167.

Nowak, Martin A., and Karl Sigmund. 1998. “Evolution of Indirect Reciprocity by Image Scoring.” *Nature* 393 (6685): 573-577.

Nowak, Martin A., and Karl Sigmund. 2005. “Evolution of Indirect Reciprocity.” *Journal of Mathematical Biology* 34: 183-231.

Peters, H. Elizabeth, A. Sinan Ünür, Jeremy Clark, and William D. Schulze. 2004. “Free-Riding and the Provision of Public Goods in the Family: A Laboratory Experiment.” *International Economic Review* 45 (1): 283-299.

Proto, Eugenio, Aldo Rustichini, and Andis Sofianos. 2019. “Intelligence, Personality, and Gains from Cooperation in Repeated Interactions.” *Journal of Political Economy* 127: 1351-1390.

Seinen, Ingrid, and Arthur Schram. 2006. “Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment.” *European Economic Review* 50 (3): 581-602.

Spadaro, Giuliana, Ilaria Tiddi, Simon Columbus, Shuxian Jin, Annette ten Teije, CoDa Team, and Daniel Balliet. 2022: “The Cooperation Databank: Machine-readable Science Accelerates Research Synthesis.” *Perspectives on Psychological Science* 17 (5): 1472-1489.

Sutter, Matthias, and Anna Untertrifaller. 2020. “Children's Heterogeneity in Cooperation and Parental Background: An Experimental Study.” *Journal of Economic Behavior and Organization* 171: 286-296.

Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler. 2019. “Economic Behavior of Children and Adolescents – A First Survey of Experimental Economics Results.” *European Economic Review* 111: 98-121.

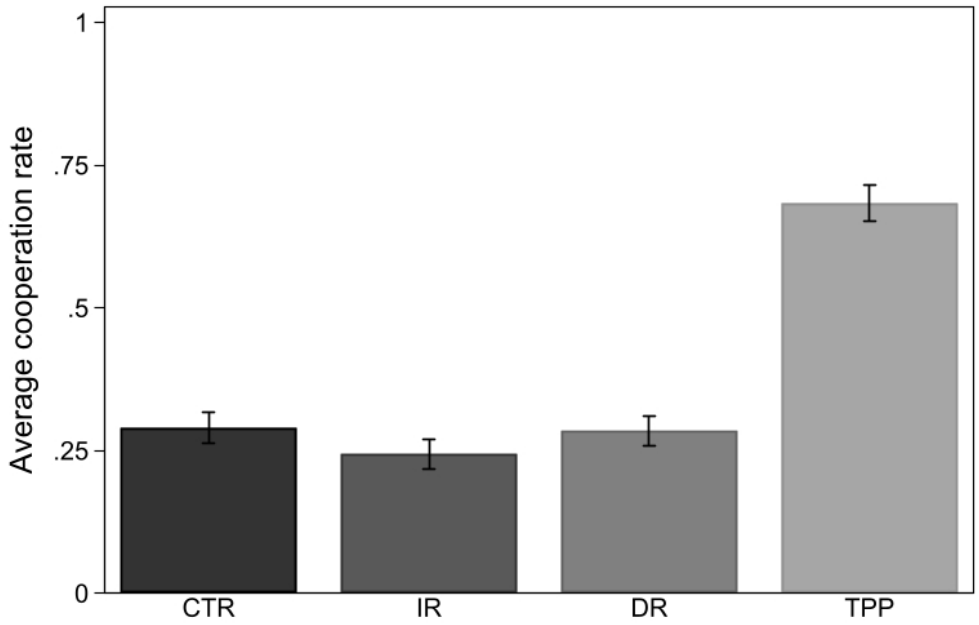
Trivers, Robert L. 1971. “The Evolution of Reciprocal Altruism.” *The Quarterly Review of Biology* 46 (1): 35-57.

Vogelsang, Martina, Keith Jensen, Sebastian Kirschner, Claudio Tennie, and Michael Tomasello. 2014. “Preschoolers Are Sensitive to Free Riding in a Public Goods Game.” *Frontiers in Psychology* 5: 729.

Wimmer, Heinz, and Josef Perner. 1983. “Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children’s Understanding of Deception.” *Cognition* 13 (1): 103-128.

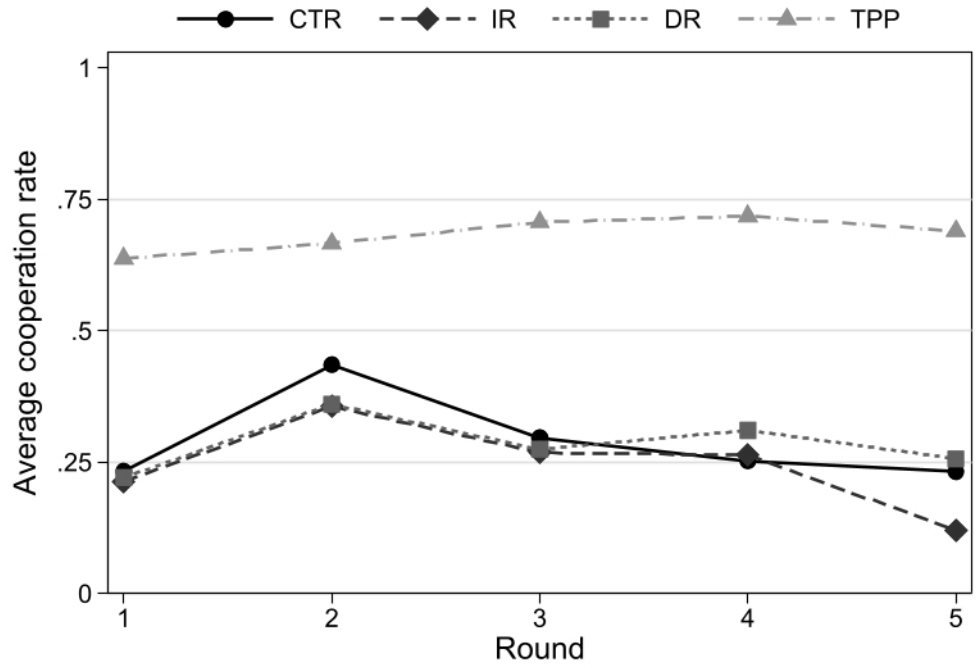
Figures and tables

Figure 1. Average cooperation rate across treatments



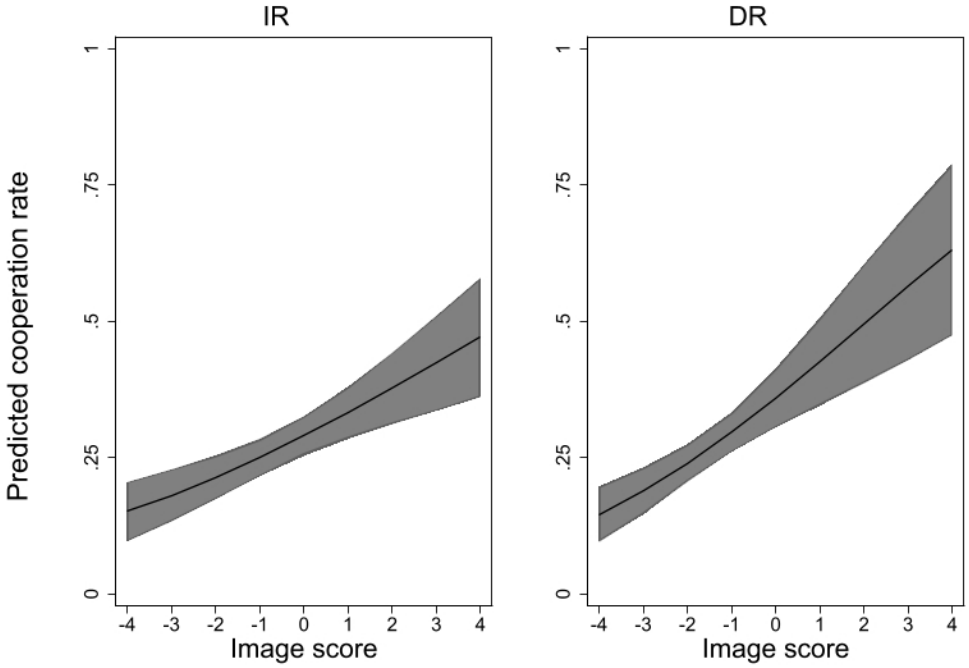
The figure shows average cooperation rates across all rounds for each treatment. Error bars indicate 95% CI. Treatment abbreviations: CTR (Control), IR (Indirect reciprocity), DR (Direct reciprocity), and TPP (Third-party punishment).

Figure 2. Average cooperation rate across rounds and treatments



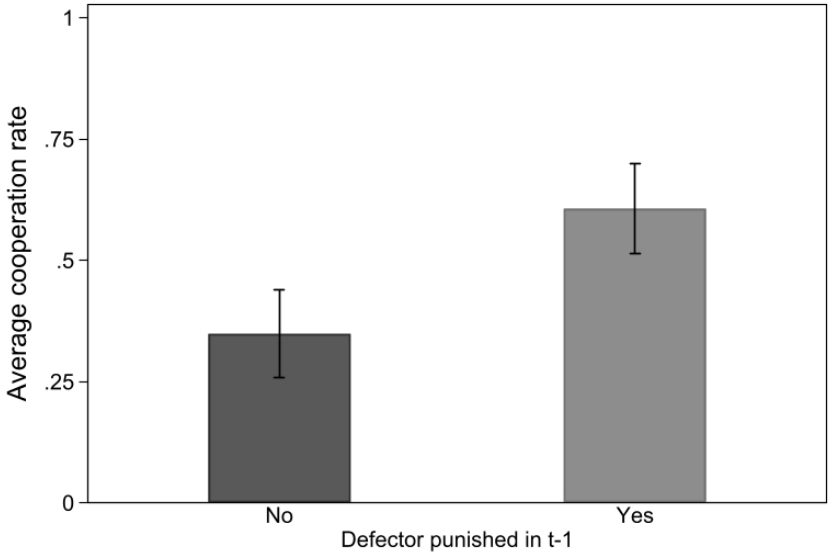
The figure shows average cooperation rates for each round and treatment.

Figure 3. Predicted cooperation rate conditional on the partner's image score



The figure shows the predicted probability of cooperation in IR and DR, conditional on the current partner's image score. The image score is calculated as the sum of a subject's actions in previous rounds, where cooperation increases the score by one, and defection decreases it by one. Shaded areas indicate 95% CI.

Figure 4. Average cooperation rate in round t of subjects who defected in round t – 1, conditional on punishment in round t – 1



The figure shows the average cooperation rate in TPP in round t for subjects who defected in round t – 1, conditional on punishment in round $t-1$ (left bar: no punishment; right bar: punishment), where $t \in \{2, 3, 4, 5\}$. Error bars indicate 95% CI.

Table 1. Distribution of subjects per age cohort and treatment

	All	3/4 years old	4/5 years old	5/6 years old
Overall	929 (814)	224 (171)	323 (282)	382 (361)
CTR	241 (202)	54 (36)	87 (74)	100 (92)
IR	236 (216)	58 (51)	81 (73)	97 (92)
DR	249 (222)	55 (41)	92 (82)	102 (99)
TPP	203 (174)	57 (43)	63 (53)	83 (78)

The table shows the number of children participating in the study across treatments (CTR, IR, DR, TPP; see below) and age cohorts. In brackets we show the number of children who passed the control questions. The latter set constitutes the basis for our analysis.

Table 2. Stage-game payoff matrix

	Cooperate	Defect
Cooperate	2,2	0,3
Defect	3,0	1,1

The table represents the payoff matrix for the PD stage-game (payoff matrix within one round).

Table 3. Probit regression estimates of main treatment effects

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
IR	-0.046*	-0.041	-0.047*	-0.046
	(0.025)	(0.028)	(0.027)	(0.030)
DR	-0.006	-0.002	-0.006	-0.024
	(0.027)	(0.028)	(0.028)	(0.029)
TPP	0.393***	0.405***	0.410***	0.421***
	(0.040)	(0.040)	(0.040)	(0.039)
Round			-0.009**	-0.008*
			(0.004)	(0.005)
Age [§]			-0.001	-0.000
			(0.001)	(0.001)
Girl (= 1)			0.011	0.001
			(0.018)	(0.020)
Number of siblings			0.004	-0.004
			(0.011)	(0.012)
Cognitive abilities			-0.001	0.002
(standardized)			(0.011)	(0.011)
Patience			0.016	0.022**
			(0.010)	(0.011)
Theory of mind (= 1)			0.026	0.034
			(0.022)	(0.024)
SES of parents				0.001
				(0.007)
Parental warmth				0.008
				(0.016)
Kindergarten FE	No	Yes	Yes	Yes
p-value: IR = DR	0.124	0.172	0.146	0.459
p-value: IR = TPP	<0.001	<0.001	<0.001	<0.001
p-value: DR = TPP	<0.001	<0.001	<0.001	<0.001
Observations	4,062	4,062	4,007	3,217

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and treatment dummy variables as the main independent variables. The reported coefficients represent average marginal effects. The omitted treatment category is the CTR treatment. Additional variables include the round, age (i.e., number of months), gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; numbers represent the number of tokens saved for the next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; a higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

[§] The observed age patterns stay the same when looking at the effect of age cohorts. In particular, we regress age cohort dummy variables with the youngest age cohort (3-4 year-olds) as the omitted category, on cooperation. The positive effect on cooperation in the TPP treatment becomes progressively larger with increasing age cohort, while for the IR treatment, the negative age effect seems to be mostly driven by a pronounced decrease of cooperation of the oldest age cohort (5-6 year-olds) in comparison to the middle and the youngest age cohort.

Table 4. Probit regression estimates of the effects of repeated play on cooperation

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Round	-0.019**	-0.018*	-0.029***	-0.028***	0.002	-0.002	0.015*	0.018*	-0.018**	-0.018*
	(0.008)	(0.010)	(0.009)	(0.009)	(0.007)	(0.008)	(0.008)	(0.010)	(0.008)	(0.009)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.011	-0.011
									(0.041)	(0.047)
DR									-0.066*	-0.072**
									(0.036)	(0.037)
TPP									0.284***	0.307***
									(0.049)	(0.053)
IR × round									-0.013	-0.011
									(0.013)	(0.013)
DR × round									0.020[§]	0.017
									(0.011)	(0.012)
TPP × round									0.033°	0.035°
									(0.011)	(0.014)
Observations	1,002	787	1,080	850	1,110	840	870	740	4,062	3,217

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable. Columns 1 to 8 use round as the main independent variable, while columns 9 and 10 use round, treatment dummy variables, and their interaction terms. Reported coefficients represent average marginal effects. The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

[#] Control variables include age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

[§] DR × round interaction effect is positive and significant at the 10% level for all observations in column 9, while it is not significant for any observation in column 10.

° TPP × round interaction effect is positive and significant at the 5% level for all observations, both in columns 9 and 10.

Table 5: Probit regression estimates of the effects of the current partner's image score on a child's cooperation

	Dependent variable: Cooperation (= 1)							
	IR				DR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Partner's image score	0.038***	0.034***	0.021**	0.017	0.059***	0.053***	0.052***	0.049***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(0.013)	(0.016)	(0.018)
Round			-0.037***	-0.048***			0.026	0.022
			(0.012)	(0.014)			(0.016)	(0.020)
Cooperation of partner in previous round (= 1)			-0.009	-0.012			-0.014	-0.025
			(0.029)	(0.038)			(0.047)	(0.055)
Subject's image score			0.077***	0.065***			0.091***	0.098***
			(0.011)	(0.012)			(0.016)	(0.018)
Subject's cooperation in previous round (=1)			-0.189***	-0.184***			-0.280***	-0.311***
			(0.036)	(0.041)			(0.042)	(0.053)
Age			-0.003*	-0.004**			0.001	0.003
			(0.002)	(0.002)			(0.002)	(0.002)
Girl (= 1)			0.048*	0.003			0.014	0.019
			(0.029)	(0.034)			(0.030)	(0.034)
Siblings			0.015	0.020			0.013	-0.016
			(0.015)	(0.021)			(0.022)	(0.021)
Std. cognitive abilities			-0.021	0.002			0.007	0.015
			(0.015)	(0.020)			(0.018)	(0.019)
Patience			0.001	-0.001			-0.013	-0.014
			(0.018)	(0.024)			(0.014)	(0.016)
Theory of mind (= 1)			0.042	0.060			0.032	0.047
			(0.031)	(0.040)			(0.041)	(0.040)
SES				-0.007				-0.005
				(0.013)				(0.011)
Parental warmth				0.019				-0.021
				(0.031)				(0.019)
Kindergarten FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	864	864	860	680	888	888	876	672

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and the partner's image score as the main independent variable. The sample consists of subjects in IR (columns 1-4) and DR (columns 5-8). The coefficients represent average marginal effects. Additional variables include the round, a dummy variable indicating whether a partner from the previous round cooperated (= 1), a subject's image score, a dummy variable indicating whether a subject cooperated in the previous round (= 1), age as the number of months, a gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; numbers represent the number of tokens saved for next day), a theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher numbers indicate higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table 6: OLS regression estimates of the effects of cooperation on a subject's payoffs

	Dependent variable: subject's round payoff				
	CTR	IR	DR	TPP	All treatments
	(1)	(2)	(3)	(4)	(5)
Cooperation (= 1)	-0.882***	-0.824***	-0.944***	0.462***	-0.882***
	(0.068)	(0.065)	(0.108)	(0.102)	(0.067)
IR					-0.094*
					(0.049)
DR					-0.012
					(0.058)
TPP					-0.617***
					(0.088)
IR × cooperation					0.059
					(0.093)
DR × cooperation					-0.061
					(0.126)
TPP × cooperation					1.344***
					(0.121)
Constant	1.577***	1.483***	1.564***	0.960***	1.577***
	(0.040)	(0.029)	(0.043)	(0.079)	(0.040)
Observations	1,002	1,080	1,110	870	4,062
R-squared	0.160	0.140	0.181	0.043	0.134

The table reports regression results from OLS models using a subject's round payoff as the dependent variable. Columns 1 to 4 use a dummy variable indicating whether the subject cooperated (= 1) as the main independent variable, while column 5 uses the dummy variable indicating whether the subject cooperated, the treatment dummy variables, and their interaction terms. The omitted treatment category in column 5 is the CTR treatment. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

Table 7. Probit regression estimates of the effects of age on cooperation

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age (in months)	-0.006***	-0.002	-0.003*	-0.005**	-0.001	-0.000	0.005*	0.007**	-0.006***	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.241	-0.132
									(0.167)	(0.193)
DR									-0.342**	-0.284
									(0.165)	(0.187)
TPP									-0.387*	-0.284
									(0.217)	(0.210)
Age × IR									0.003	0.001
									(0.002)	(0.003)
Age × DR									0.005[§]	0.004
									(0.002)	(0.003)
Age × TPP									0.011°	0.010°
									(0.003)	(0.003)
Observations	997	787	1,080	850	1,110	840	870	740	4,057	3,217

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable. Columns 1 to 8 use age (in months) as the main independent variable, while columns 9 and 10 use age (in months), treatment dummy variables, and their interaction terms. Reported coefficients represent average marginal effects. The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** $p < 0.01$. ** $p<0.05$. * $p<0.1$.

[#] Control variables include the round, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

[§] In column 9, the Age × DR interaction effect is positive and significant for all observations (predominantly at the 5% level). In column 10, it is positive, but insignificant, for all observations.

° The Age × TPP interaction effect is positive and significant at the 1% level across all observations, both in columns 9 and 10.

Table 8. Probit regression estimates of the effects of cognitive abilities, patience, and theory of mind (TOM) on cooperation

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Std. cognitive	-0.061***	-0.031	-0.019	0.019	-0.004	0.018	0.047**	-0.005	-0.060***	-0.044**
abilities	(0.020)	(0.021)	(0.014)	(0.020)	(0.017)	(0.018)	(0.023)	(0.029)	(0.020)	(0.021)
Patience	0.020	0.011	0.002	-0.003	-0.005	-0.002	0.049*	0.069***	0.020	0.020
	(0.017)	(0.021)	(0.016)	(0.022)	(0.015)	(0.015)	(0.029)	(0.023)	(0.017)	(0.021)
Theory of mind	-0.033	-0.020	0.003	0.012	0.025	0.007	0.100*	-0.002	-0.032	-0.008
(= 1)	(0.040)	(0.047)	(0.037)	(0.044)	(0.039)	(0.045)	(0.058)	(0.045)	(0.039)	(0.046)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.045	-0.045
									(0.068)	(0.084)
DR									-0.006	-0.020
									(0.070)	(0.081)
TPP									0.232**	0.252**
									(0.099)	(0.109)
Std. cognitive									0.043[§]	0.042
abilities × IR									(0.025)	(0.028)
Std. cognitive									0.056°	0.053°
abilities × DR									(0.026)	(0.027)
Std. cognitive									0.113˜	0.088˜
abilities × TPP									(0.032)	(0.035)
Patience × IR									-0.019	-0.009
									(0.024)	(0.030)
Patience × DR									-0.025	-0.022
									(0.022)	(0.026)
Patience × TPP									0.034	0.044
									(0.036)	(0.038)
TOM × IR									0.036	0.011
									(0.055)	(0.066)
TOM × DR									0.057	0.037
									(0.055)	(0.063)
TOM × TPP									0.142≈	0.114
									(0.075)	(0.075)
Observations	982	787	1,075	850	1,100	840	860	740	4,017	3,217

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable. Columns 1 to 8 use standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), and a theory of mind dummy variable (= 1) as the main independent variables, while columns 9 and 10 use the same three variables, but also treatment dummy variables and their interaction terms. Reported coefficients represent average marginal effects. The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their reported error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

[#] Control variables include age as the number of months, round, gender dummy variable (girl = 1), number of siblings, SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

[§] Std. cognitive abilities × IR is positive for all observations, both in columns 9 and 10. For the majority of observations, the interaction effect is significant at the 10% level in column 9 and insignificant (at any conventional level) in column 10.

° Std. cognitive abilities × DR is positive and significant for all observations, predominantly at the 5% level in column 9, and at the 5% and the 10% level in column 10.

˜Std. cognitive abilities × TPP is positive and significant at the 5% level for all observations, both in columns 9 and 10.

≈ TOM × TPP is positive for all observations, both in columns 9 and 10. It is significant for the majority of observations (predominantly at the 10% level) in column 9, and for the minority of observations (at the 10% level) in column 10.

Table 9. Probit regression estimates of the effects of socioeconomic status (SES) and parental warmth on cooperation

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SES	-0.025**	-0.018	-0.011	-0.014	-0.004	-0.010	0.053***	0.046***	-0.024**	-0.019
	(0.010)	(0.013)	(0.011)	(0.013)	(0.011)	(0.012)	(0.019)	(0.017)	(0.010)	(0.012)
Parental warmth	0.010	-0.003	0.031	0.025	-0.016	-0.042**	0.053	0.045	0.010	0.016
	(0.026)	(0.028)	(0.030)	(0.033)	(0.014)	(0.020)	(0.044)	(0.041)	(0.025)	(0.027)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.233	-0.171
									(0.257)	(0.277)
DR									0.089	0.284
									(0.245)	(0.232)
TPP									-0.308	-0.202
									(0.193)	(0.242)
IR × SES									0.017	0.010
									(0.017)	(0.019)
DR × SES									0.017	0.009
									(0.016)	(0.014)
TPP × SES									0.070[§]	0.068[§]
									(0.042)	(0.031)
IR × parental warmth									0.019	0.014
									(0.038)	(0.040)
DR × parental warmth									-0.025	-0.049
									(0.029)	(0.035)
TPP × parental warmth									0.035	0.030
									(0.045)	(0.050)
Observations	807	787	850	850	850	840	745	740	3,252	3,217

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable. Columns 1 to 8 use SES as the highest education level of parents (from 1 to 8; higher number indicates higher education) and parents' self-reported parental warmth as the main independent variables, while columns 9 and 10 use SES, parental warmth, treatment dummy variables, and their interaction terms. Reported coefficients represent average marginal effects. The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their reported error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

[#] Control variables include age as the number of months, round, a gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), and kindergarten fixed effects.

[§] Both in columns 9 and 10, the TPP × round interaction effect is positive for all observations, and it is significant at the 5% level for all observations whose predicted probability of cooperation is more than 0.5. For observations whose predicted probability of cooperation is less than 0.5, the interaction effect is not significant for the majority of observations in column 9, and significant (predominantly at the 1% or the 5% level) for the majority in column 10.

Online Appendix

A.1. Supporting information on robustness checks and additional information on TPP procedures

A.1.1. Selection and control questions

Inverse probability weighting. A small fraction of children failed to pass the control questions (12%). This opens a door to potential selection issues, as passing the control questions can depend on certain characteristics of the child. To account for this, we first test which observable characteristics are predictive of passing the control questions (see Table A5). We look at the effect of standardized cognitive abilities, age, gender, the number of siblings, and also a proxy for a general willingness to cooperate, captured by the total amount of cooperation across the 5 rounds. We find that older children and those with higher cognitive abilities are more likely to succeed in correctly answering the questions. We then construct individual weights as the inverse probabilities of passing the control questions, where the probabilities result from a probit model of a binary indicator (showing whether the subject has passed the control questions) as a function of standardized cognitive abilities and age. After obtaining the weights, we then reweight our dataset with the constructed weights and repeat the entire analysis (see Tables A6 to A14; see Kosse et al., 2020, Bašić et al., 2020, and Falk et al, 2021, for a similar approach). Our results stay robust to reweighting of the dataset.

Including children who failed the control questions. As an alternative approach to accounting for potential selection due to passing control questions, we also rerun our entire analysis by also adding to the sample those who failed the control questions. Thus, we repeat all regressions from Tables 3 to 9, and A2 to A4, with 934 subjects altogether. Due to space considerations, we merely summarize here the findings from this robustness check (the complete results are available on request). The robustness check induces a certain level of noise in the data by design; hence, one might expect subtle changes in the results. However, again we observe that our results remain highly robust overall. We notice a marginal change when looking at the treatment effects. In the robustness check, the negative effect of the IR treatment also remains significant when including controls. Also, when looking at the effect of age or the effect of cognitive abilities on cooperation in CTR, both variables also remain a significant negative predictor (at the 10% level) when we include control variables. Together, the inverse probability weighting robustness check and the robustness check with subjects who failed the control questions provide strong evidence that our findings do not suffer from selection issues due to passing the control questions.

A.1.2. Selection, parents' SES, and parental warmth

Inverse probability weighting. We elicited the education of parents and information on parental warmth through a take-home survey. With this approach, we obtained information for SES (highest educational level of parents) and parental warmth for 80% of our main sample.[14] As this sample is not randomly determined, there is a possibility that selection might affect our findings concerning SES and parental warmth, in particular their relation to cooperation (see Table 9). To account for this concern, we first test which observable characteristics are predictive of passing the control questions (see Table A16). Here, we look at standardized cognitive abilities of the child, gender, age, number of siblings, a proxy variable for general cooperativeness captured by the total amount of cooperation across the 5 rounds, and kindergarten fixed effects. We find that age and gender (at the 10% level), as well as kindergarten fixed effects are significantly predictive of obtaining information on SES and parental warmth. To account for these imbalances, we construct individual weights as the inverse probabilities of obtaining information on both SES and parental warmth, where the probabilities result from a probit model of a binary indicator (indicating whether both SES and parental warmth were obtained for a child)

[14] Not obtaining information for the remaining 20% of parents is almost exclusively due to parents not completing the survey at all. For those who did return a completed survey, only 2% did not provide an answer on questions regarding education or parental warmth.

as a function of gender, age, and kindergarten fixed effects. After constructing the weights, we reweight our dataset and repeat the analysis for SES and parental warmth (see Table A17; see Kosse et al., 2020, Bašić et al., 2020, and Falk et al, 2021, for a similar approach). We find that our results concerning the relation between SES, parental warmth, and cooperative behavior of children stay robust, suggesting that selection concerning survey responses does not affect our findings.

A.1.3. Comprehension of the game

A key aspect of studies with very young children is ensuring that subjects understand the game. To provide an additional robustness check and ensure that our results are not affected by comprehension issues, we have checked what would happen to our results if we took a more conservative approach in classifying who understood the game. To this end, we utilize an RA's follow-up questions to build a stricter comprehension criterion. Of the 814 children who passed our main comprehension checks, we additionally exclude those who exhibited any difficulty in verbalizing with whom they were playing, and explain the introduced mechanisms in the particular treatments, i.e., the meaning of icons on the tablet screen representing the current partner's previous round behavior in IR and DR, or the punisher's role, and the options in TPP. This set comprises the most important follow-up questions, which were asked directly before the game began. We find that with this approach we would exclude another 134 subjects from our main sample, leaving us with 680 subjects altogether. We rerun our entire analysis with this restricted sample (results are available upon request). Overall, our results stay highly robust and consistent with our main analysis.

A.1.4. Varying the calculation of the image score

To assess one's reputation for being cooperative in DR and IR treatments we rely on the image score as proposed in Nowak and Sigmund (1998). An aspect one could be concerned about when using this score is the fact that the image score can fail in certain cases when differentiating between those who only showed cooperation so far but did not play many rounds, and those who played many rounds but had played different actions. For example, a subject who cooperated once will have the same image score as a subject who cooperated twice and defected once. To differentiate between such subjects, we employ an alternative measure, namely the percentage of previous cooperation (e.g., for a subject who cooperated twice and defected once, the measure would yield a value of 0.67, and for a subject who cooperated in the first round, it would yield a value of 1 for that round). Our results remain highly similar to our main analysis (Table A15 in the Appendix).

A.1.5. TPP procedures

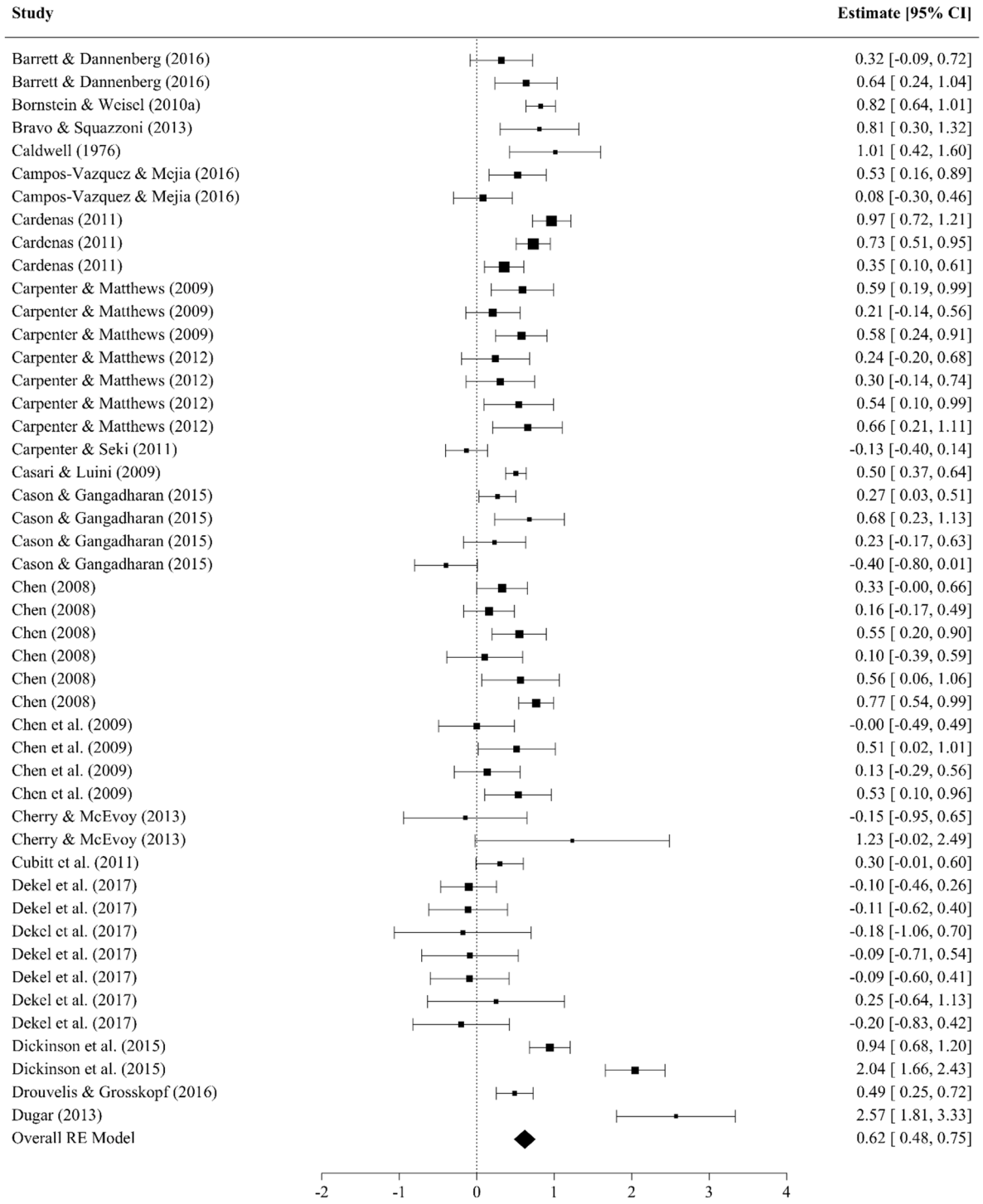
To collect punishment decisions for the TPP treatment, at the end of sessions in the other treatments CTR, IR, and DR, children were invited (as a surprise) to play a final task on an additional screen where they acted as the observers of how two other children played the prisoner's dilemma game (see Section 2.3). All children accepted the invitation and were then confronted with one of the two punishment decision scenarios: they were either asked to make a decision on whether to punish defection (scenario 1) or were shown a setting of mutual cooperation where no punishment was possible (scenario 2; Experimental instructions can be found on https://osf.io/y7h4v/?view_only=f4627bf3ffe347e085f48144b19451e5). Thus, the punisher always got to keep the token in scenario 2. The collected punishment decisions were always matched to one of the five rounds of children playing the PD game in a TPP session in a different kindergarten, where scenario 1 decision was used when at least 1 child within the pair defected, and scenario 2 when both children cooperated. Punishment decisions for the initial TPP sessions were collected in prior sessions of CTR, IR, and DR. We then recorded (and followed) the actual defection rate in TPP during the experiment to estimate the number of punishment decisions (both scenario 1 and scenario 2) required for the remaining sessions, and collected them accordingly. The punishment decisions were applied by a random draw in the TPP treatment.

A.2. Supporting figure and tables

Figure A1. Interface of the game and the experimental setting

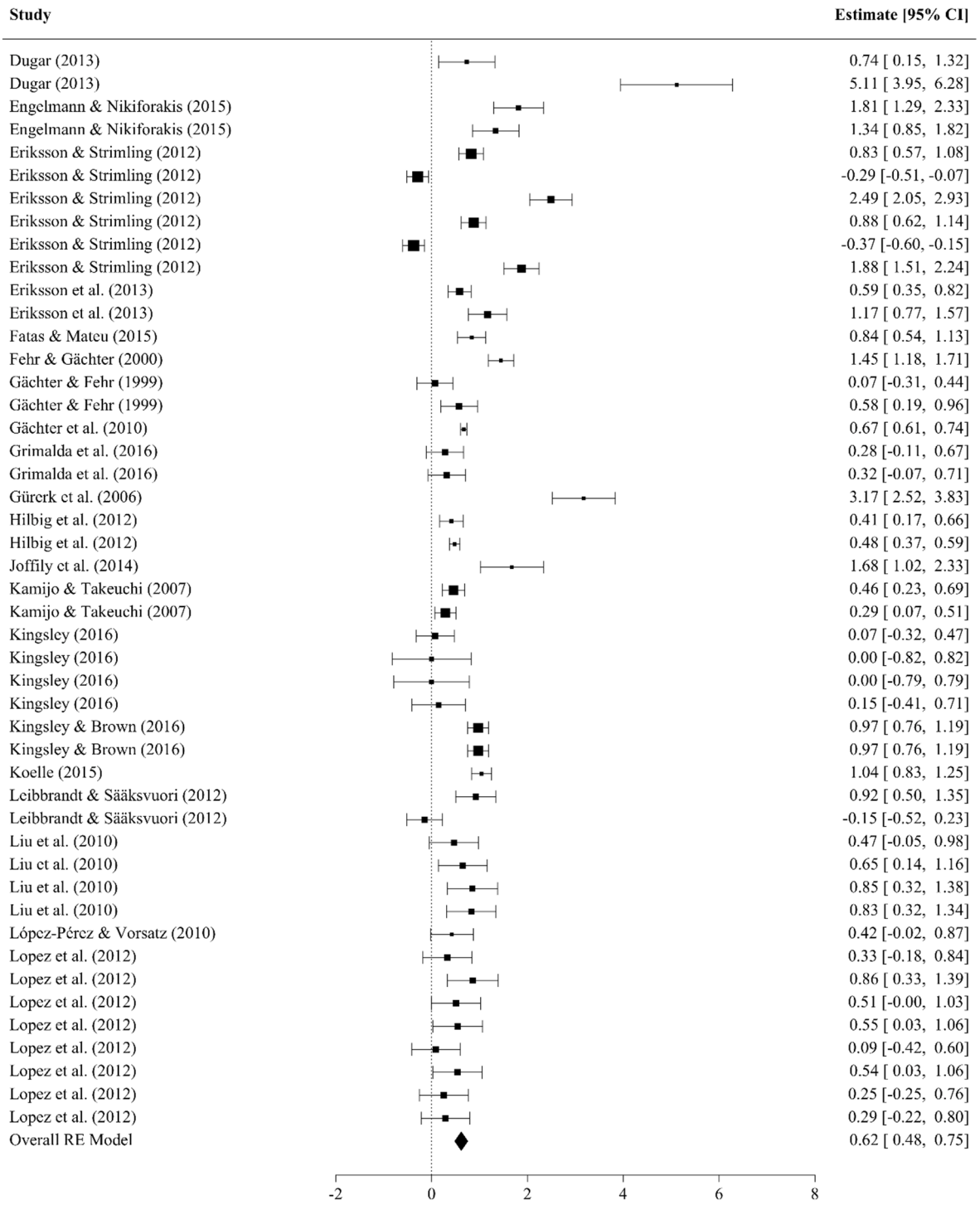
Panel A shows the interface used to make decisions in the prisoner's dilemma game. Decision-makers were represented by the lower avatar and could either give 2 tokens to their partner (represented by the upper avatar) by pressing the button with the open hand, or keep 1 token for themselves by pressing the button with the closed hand. Panel A1 was only shown in the direct (DR) and indirect reciprocity (IR) treatments and represents all of the actions the partner chose in the previous round(s). Panel A2 was only shown in the third-party punishment (TPP) treatment and shows the third-party punisher who could use one token to punish if at least one of the two players defected. Panel B shows a graphic representation of our experimental setting in kindergartens. One session consisted of six children making simultaneous decisions. Decisions were made on tablet computers, and children listened to audio instructions during the trial rounds and all rounds of the actual game. Each child was accompanied by one helper.

Figure A2. Meta-analysis on the effect of punishment (part 1 of 4)



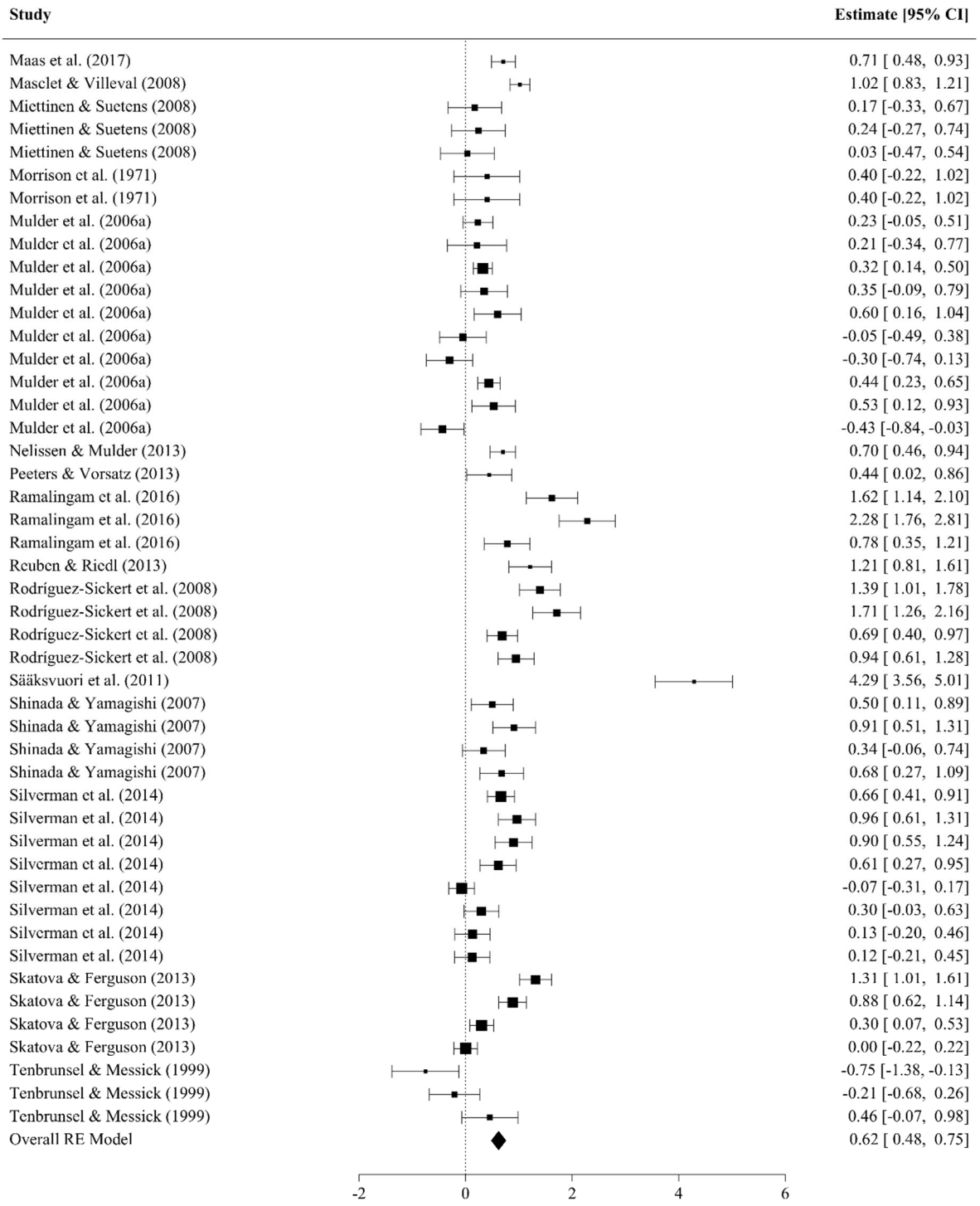
The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how punishment affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a punishment mechanism is introduced. The figure is the first of four figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the four figures). 95% CI in brackets.

Figure A3. Meta-analysis on the effect of punishment (part 2 of 4)



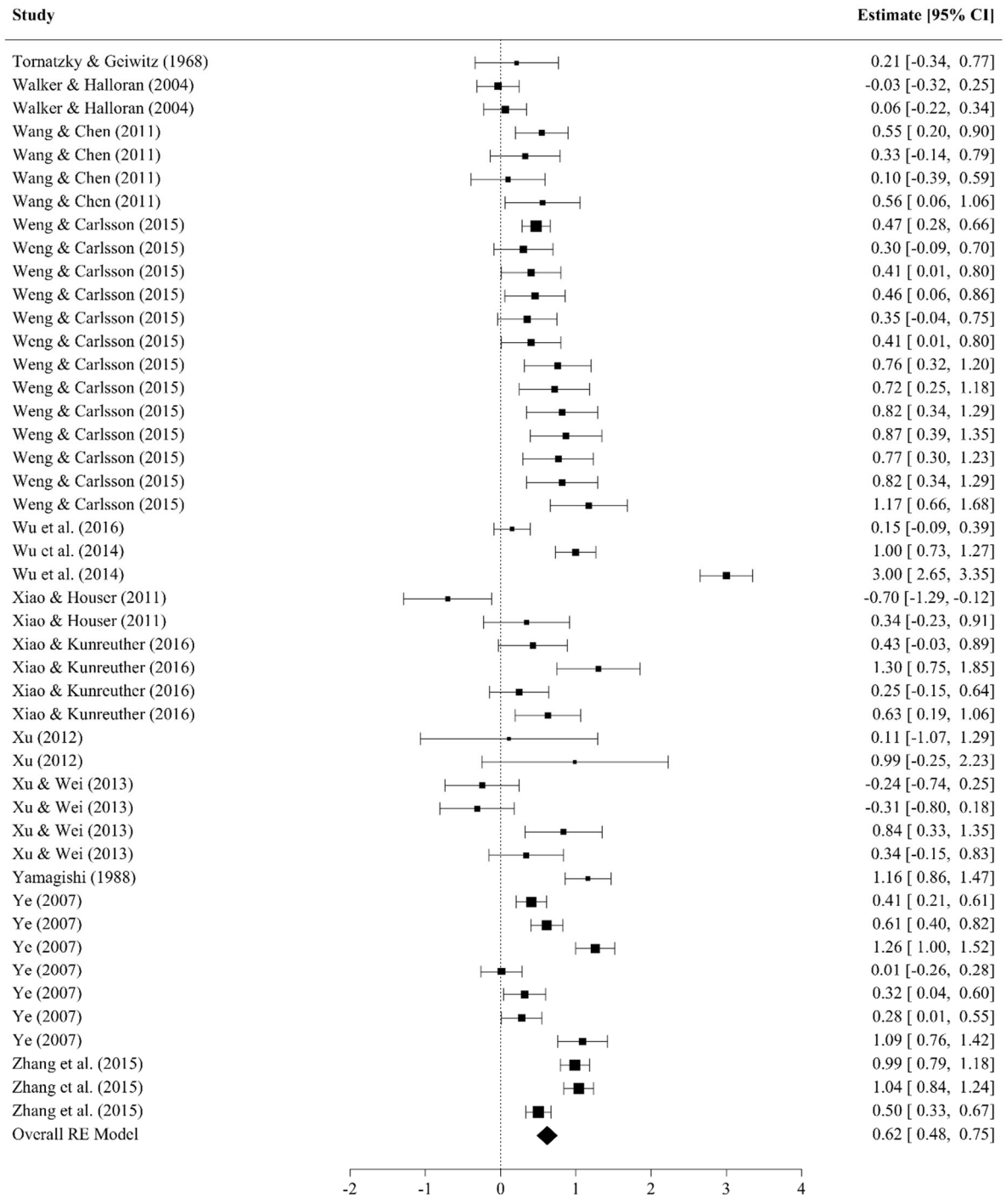
The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how punishment affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a punishment mechanism is introduced. The figure is the second of four figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the four figures). 95% CI in brackets.

Figure A4. Meta-analysis on the effect of punishment (part 3 of 4)



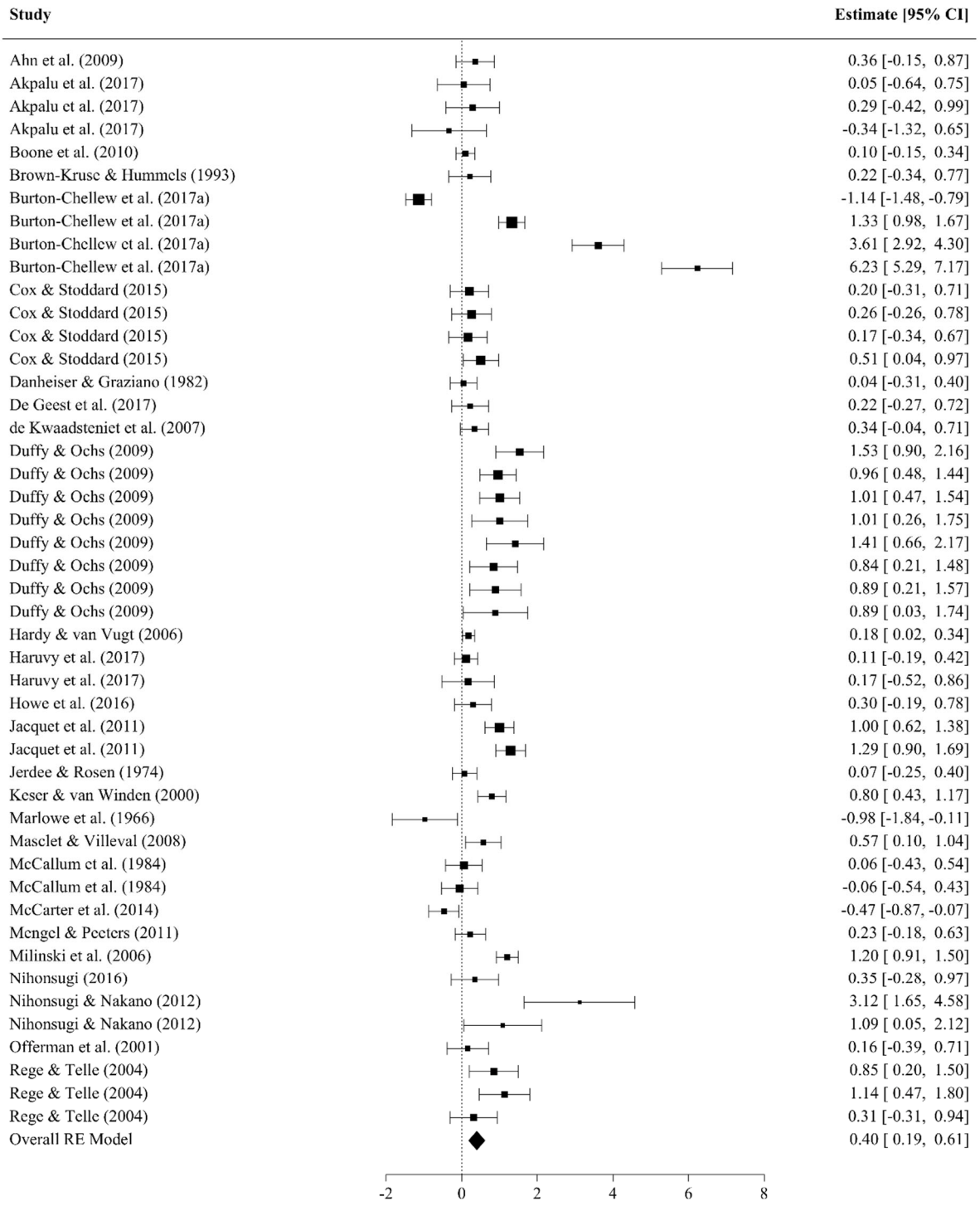
The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how punishment affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a punishment mechanism is introduced. The figure is the third of four figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the four figures). 95% CI in brackets.

Figure A5. Meta-analysis on the effect of punishment (part 4 of 4)



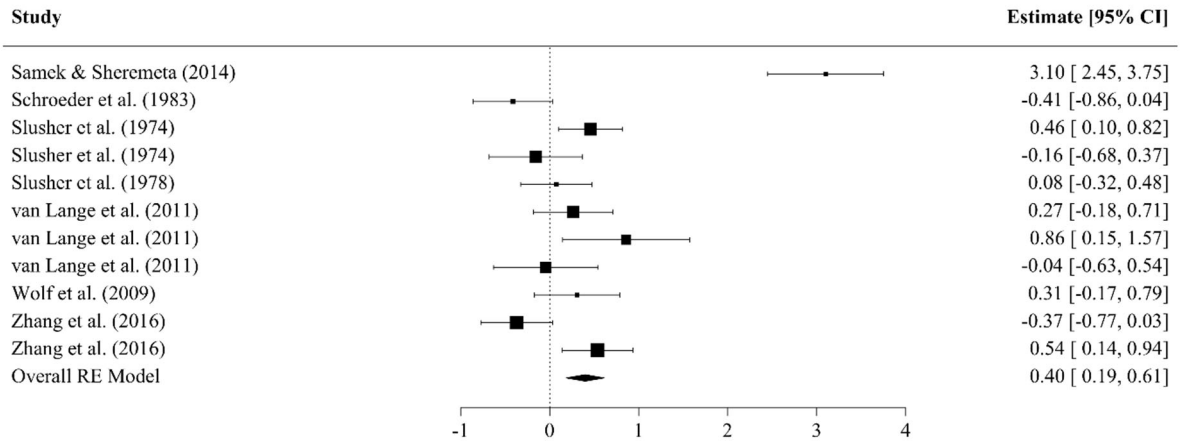
The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how punishment affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a punishment mechanism is introduced. The figure is the fourth of four figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the four figures). 95% CI in brackets.

Figure A6. Meta-analysis on the effect of direct reciprocity (part 1 of 2)



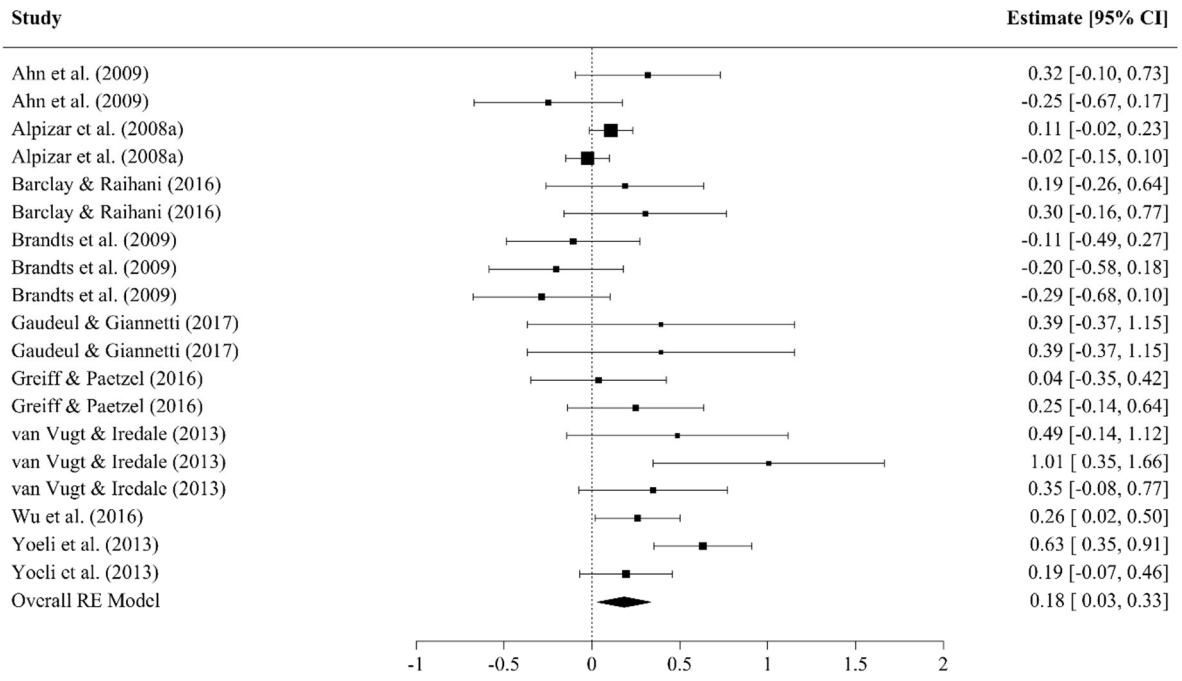
The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how direct reciprocity affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a direct reciprocity mechanism is introduced. The figure is the first of two figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the two figures). 95% CI in brackets.

Figure A7. Meta-analysis on the effect of direct reciprocity (part 2 of 2)



The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how direct reciprocity affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where a direct reciprocity mechanism is introduced. The figure is the second of two figures. The overall RE model captures the estimation of Cohen's *d* using all treatment effects (reported across the two figures). 95% CI in brackets.

Figure A8. Meta-analysis on the effect of indirect reciprocity



The figure shows estimations of effect sizes (Cohen's *d*) for studies that report findings on how indirect reciprocity affects cooperative behavior. Each reported effect size represents one comparison within a study between a control treatment and a treatment where an indirect reciprocity mechanism is introduced. The overall RE model captures the estimation of Cohen's *d* using all treatment effects. 95% CI in brackets.

Table A1. Descriptive statistics across treatments

	CTR	IR	DR	TPP	All subjects
	(1)	(2)	(3)	(4)	(5)
Girl (= 1)	0.490	0.486	0.455	0.523	0.486
	(0.501)	(0.501)	(0.499)	(0.501)	(0.500)
Siblings	1.129	1.162	1.104	1.006	1.105
	(0.775)	(0.811)	(0.770)	(0.701)	(0.769)
Age (in months)	65.95	65.46	65.65	64.87	65.51
	(9.827)	(9.892)	(9.300)	(10.31)	(9.799)
Cognitive abilities	7.010	6.851	7.045	6.920	6.958
	(1.476)	(1.701)	(1.755)	(1.708)	(1.664)
Theory of mind (= 1)	0.715	0.749	0.741	0.701	0.728
	(0.453)	(0.435)	(0.439)	(0.459)	(0.445)
Patience	0.965	0.916	0.914	0.866	0.917
	(0.873)	(0.827)	(0.856)	(0.844)	(0.849)
SES	5.745	5.399	5.669	5.853	5.657
	(1.505)	(1.531)	(1.541)	(1.539)	(1.535)
Parental Warmth	9.639	9.735	9.550	9.640	9.641
	(0.671)	(0.591)	(0.994)	(0.627)	(0.743)
Observations	202	216	222	174	814

The table reports means and standard deviations (in brackets) of variables across the four treatments (columns 1 to 4) and for the entire sample (column 5). Variables include a gender dummy variable (girl = 1), the number of siblings, age in months, cognitive abilities (as the number of correctly solved Raven Matrices), a dummy variable indicating whether a subject possesses theory of mind (= 1), patience (from 0 to 2; number represents the number of tokens saved for next day), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), and self-reported parental warmth.

We conduct an F-test for non-binary variables (χ^2 for binary variables) to jointly test whether the reported values in IR, DR, and TPP differ in comparison to CTR. For each of our 8 variables, we regress the variable on the three treatment dummy variables with CTR as the omitted category. Out of 8 regressions, only the regression with dependent variable SES exhibits a significant joint test at any conventional significance level ($p = 0.045$). The significance seems to be driven by the IR treatment (removing the IR treatment from the joint test yields a joint-test p-value of 0.556, while removing DR or TPP yields a significant result in both cases: $p < 0.087$ for both comparisons).

Table A2. Probit regression estimates of main treatment effects in the first round

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
IR	-0.020	-0.029	-0.035	-0.062
	(0.042)	(0.044)	(0.043)	(0.049)
DR	-0.012	-0.020	-0.024	-0.057
	(0.046)	(0.043)	(0.043)	(0.046)
TPP	0.405***	0.391***	0.396***	0.378***
	(0.056)	(0.056)	(0.056)	(0.056)
Age			-0.002	-0.001
			(0.002)	(0.002)
Girl (= 1)			-0.033	-0.041
			(0.033)	(0.035)
Siblings			-0.002	-0.024
			(0.019)	(0.021)
Std. cognitive abilities			0.016	0.015
			(0.017)	(0.020)
Patience			0.035**	0.039**
			(0.017)	(0.019)
Theory of mind (= 1)			0.003	-0.009
			(0.039)	(0.043)
SES				-0.009
				(0.013)
Parental warmth				-0.027
				(0.023)
Kindergarten FE	No	Yes	Yes	Yes
Observations	814	814	803	645

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and treatment dummy variables as the main independent variables. The sample includes only behavior from the first round. The reported coefficients represent average marginal effects. The omitted treatment category is the CTR treatment. Additional variables include age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A3. Probit regression estimates of the effect of being in a cooperative pair in round 1

	Dependent variable: Cooperation (= 1)					
	CTR		IR		DR	
	(1)	(2)	(3)	(4)	(5)	(6)
Both players cooperated	0.104	0.128	0.153*	0.129*	0.215**	0.273***
in round 1 (= 1)	(0.069)	(0.097)	(0.082)	(0.075)	(0.087)	(0.056)
Round		-0.096***		-0.089***		-0.054***
		(0.019)		(0.014)		(0.014)
Age		-0.004		-0.003*		0.008***
		(0.005)		(0.002)		(0.003)
Girl (= 1)		0.028		-0.035		0.075*
		(0.056)		(0.046)		(0.044)
Siblings		-0.003		0.016		-0.064*
		(0.035)		(0.029)		(0.033)
Std. cognitive abilities		-0.021		-0.037		0.016
		(0.035)		(0.035)		(0.022)
Patience		0.025		-0.006		0.001
		(0.034)		(0.032)		(0.021)
Theory of mind (= 1)		-0.024		0.085*		0.113**
		(0.076)		(0.050)		(0.054)
SES		-0.037**		-0.003		0.004
		(0.015)		(0.019)		(0.013)
Parental warmth		0.001		0.009		-0.012
		(0.045)		(0.042)		(0.037)
Kindergarten FE	No	Yes	No	Yes	No	Yes
Observations	492	360	572	472	580	444

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and dummy variable indicating the pair's cooperative behavior in round one (1 if both the subject and the partner in the first round cooperated in the first round, or 0 if both the subject and the partner in the first round did not cooperate in the first round) as the main independent variable. The sample includes only observations from the second to the fifth round from subjects where they, and their first-round partners, either both cooperated in the first round or both defected in the first round. The reported coefficients represent average marginal effects. Additional variables include the round, age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A4. Probit regression estimates of the effect of experiencing punishment on cooperation

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
Experienced punishment in previous round (= 1)	0.249***	0.289***	0.310***	0.333***
	(0.060)	(0.063)	(0.067)	(0.069)
Round			-0.005	0.009
			(0.022)	(0.023)
Age			0.005	0.006
			(0.004)	(0.005)
Girl (= 1)			-0.052	-0.035
			(0.084)	(0.086)
Siblings			-0.070	-0.060
			(0.071)	(0.078)
Cooperation of partner in previous round (= 1)			-0.152**	-0.139*
			(0.063)	(0.078)
Std. cognitive abilities			0.064	0.051
			(0.039)	(0.041)
Patience			-0.042	-0.008
			(0.052)	(0.049)
Theory of mind (= 1)			0.042	-0.068
			(0.112)	(0.125)
SES				0.027
				(0.033)
Parental warmth				0.000
				(0.072)
Kindegarten FE	No	Yes	Yes	Yes
Observations	221	214	206	177

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and a dummy variable indicating whether a subject was punished in the round before (= 1) as the main independent variable. The sample consists only of subjects from TPP who kept their token (were selfish) in the previous round. The reported coefficients represent average marginal effects. Additional variables include the round, age as the number of months, gender dummy variable (girl = 1), number of siblings, dummy variable indicating whether partner from the previous round cooperated, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** p<0.01. ** p<0.05. * p<0.1.

Table A5. Probit regression estimates of selection on observables concerning passing the control questions

	Dependent variable: Control questions passed (= 1)
	(1)
Total amount of cooperation	-0.002
	(0.007)
Siblings	0.003
	(0.013)
Girl (= 1)	-0.024
	(0.021)
Age	0.006***
	(0.001)
Std. cognitive abilities	0.027**
	(0.011)
Observations	917

The table reports regression results from a probit model using a dummy variable indicating whether the SES variable (highest educational level of the parents) and the parental warmth variable were obtained for the subject. Independent variables include the subject's total amount of cooperation in the repeated prisoner's dilemma game, number of siblings, gender dummy variable (girl= 1), age as the number of months, and standardized cognitive abilities. The reported coefficients represent average marginal effects. Standard errors in parentheses. *** p<0.01. ** p<0.05. * p<0.1.

Table A6. Probit regression estimates of main treatment effects using inverse probability weighting

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
IR	-0.050*	-0.045	-0.051*	-0.050*
	(0.026)	(0.028)	(0.028)	(0.030)
DR	-0.012	-0.007	-0.011	-0.031
	(0.028)	(0.028)	(0.028)	(0.029)
TPP	0.381***	0.394***	0.399***	0.410***
	(0.041)	(0.040)	(0.040)	(0.040)
Round			-0.009**	-0.008*
			(0.004)	(0.005)
Age			-0.001	-0.000
			(0.001)	(0.001)
Girl (= 1)			0.013	0.004
			(0.018)	(0.021)
Siblings			0.004	-0.003
			(0.012)	(0.013)
Std. cognitive abilities			0.000	0.005
			(0.011)	(0.012)
Patience			0.018*	0.025**
			(0.010)	(0.011)
Theory of mind (= 1)			0.025	0.033
			(0.022)	(0.025)
SES				-0.001
				(0.008)
Parental warmth				0.008
				(0.016)
Kindergarten FE	No	Yes	Yes	Yes
Observations	4,062	4,062	4,007	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable and treatment dummy variables are used as the main independent variables. The reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category is the CTR treatment. Additional variables include the round, age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A7. Probit regression estimates of main treatment effects in the first round using inverse probability weighting

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
IR	-0.031	-0.041	-0.047	-0.074
	(0.043)	(0.044)	(0.043)	(0.049)
DR	-0.015	-0.023	-0.027	-0.060
	(0.047)	(0.044)	(0.044)	(0.047)
TPP	0.392***	0.381***	0.385***	0.366***
	(0.056)	(0.056)	(0.055)	(0.057)
Age			-0.003	-0.001
			(0.002)	(0.002)
Girl (= 1)			-0.031	-0.038
			(0.033)	(0.035)
Siblings			-0.003	-0.025
			(0.019)	(0.021)
Std. cognitive abilities			0.017	0.016
			(0.017)	(0.020)
Patience			0.037**	0.042**
			(0.017)	(0.019)
Theory of mind (= 1)			0.002	-0.012
			(0.039)	(0.043)
SES				-0.010
				(0.013)
Parental warmth				-0.029
				(0.023)
Kindergarten FE	No	Yes	Yes	Yes
Observations	814	814	803	645

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable and treatment dummy variables are used as the main independent variables. The sample includes only behavior from the first round. The reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category is the CTR treatment. Additional variables include age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A8. Probit regression estimates of the effect of round on cooperation using inverse probability weighting

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Round	-0.019**	-0.018*	-0.027***	-0.027***	0.002	-0.003	0.015*	0.018*	-0.019**	-0.018*
	(0.008)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.011)	(0.008)	(0.009)
Control variables#	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.022	-0.020
									(0.041)	(0.047)
DR									-0.072*	-0.076**
									(0.037)	(0.038)
TPP									0.272***	0.299***
									(0.050)	(0.054)
IR × round									-0.009	-0.010
									(0.012)	(0.013)
DR × round									0.020§	0.015
									(0.011)	(0.013)
TPP × round									0.033°	0.035°
									(0.012)	(0.014)
Observations	1,002	787	1,080	850	1,110	840	870	740	4,062	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable. Columns 1 to 8 use the round as the main independent variable, while columns 9 and 10 use the round, treatment dummy variables, and their interaction terms. Reported coefficients represent (weighted) average marginal effects. Weights are predicted inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category in columns 9 and 10 is the CTR treatment. The interaction coefficients in columns 9 and 10 represent average interaction effects and their error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

Control variables include age as the number of months, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

§ DR × round interaction effect is positive and significant at the 10% level for all observations in column 9, while it is not significant for any observation in column 10.

° TPP × round interaction effect is positive and significant (at least) at the 5% level for all observations, both in columns 9 and 10.

Table A9. Probit regression estimates of the effect of the partner's image score on cooperation using inverse probability weighting

	Dependent variable: Cooperation (= 1)							
	IR				DR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Partner's image score	0.038***	0.034***	0.021**	0.016§	0.059***	0.053***	0.052***	0.050***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(0.013)	(0.016)	(0.017)
Round			-0.037***	-0.049***			0.027*	0.022
			(0.012)	(0.014)			(0.016)	(0.020)
Cooperation of partner in previous round (= 1)			-0.012	-0.016			-0.012	-0.022
			(0.028)	(0.038)			(0.048)	(0.056)
Player's image score			0.077***	0.063***			0.092***	0.098***
			(0.011)	(0.012)			(0.015)	(0.018)
Subject's cooperation in previous round (= 1)			-0.182***	-0.177***			-0.284***	-0.314***
			(0.036)	(0.041)			(0.042)	(0.053)
Age			-0.003*	-0.005**			0.001	0.003
			(0.002)	(0.002)			(0.002)	(0.002)
Girl (= 1)			0.048	0.002			0.013	0.019
			(0.030)	(0.035)			(0.030)	(0.033)
Siblings			0.015	0.022			0.017	-0.016
			(0.015)	(0.021)			(0.023)	(0.020)
Std. cognitive abilities			-0.020	0.003			0.009	0.018
			(0.014)	(0.019)			(0.018)	(0.018)
Patience			0.005	0.006			-0.016	-0.016
			(0.019)	(0.025)			(0.014)	(0.016)
Theory of mind (= 1)			0.042	0.056			0.029	0.046
			(0.031)	(0.040)			(0.041)	(0.039)
SES				-0.009				-0.006
				(0.013)				(0.011)
Parental warmth				0.020				-0.020
				(0.031)				(0.019)
Kindergarten FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	864	864	860	680	888	888	876	672

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable, and the partner's image score is used as the main independent variable. The sample consists of subjects in IR (columns 1 to 4) and DR (columns 5 to 8). Reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). Additional variables include the round, a dummy variable indicating whether a partner from the previous round cooperated (= 1), a subject's image score, a dummy variable indicating whether a subject cooperated in the previous round (= 1), age as the number of months, a gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), a theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

§ The effect of the partner's image score in column 4 is not significant (p = 0.135), but it becomes significant when imputing missing SES and parental warmth values from age, gender, standardized cognitive abilities, theory of mind, number of siblings and patience, and repeating the regression (p = 0.025, AME = 0.021).

Table A10. Probit regression estimates of the effect of experiencing punishment on cooperation using inverse probability weighting

	Dependent variable: Cooperation (= 1)			
	(1)	(2)	(3)	(4)
Experienced punishment in previous round (= 1)	0.252***	0.292***	0.322***	0.346***
	(0.059)	(0.062)	(0.065)	(0.068)
Round			-0.004	0.010
			(0.021)	(0.023)
Age			0.004	0.005
			(0.004)	(0.005)
Girl (= 1)			-0.042	-0.025
			(0.086)	(0.088)
Siblings			-0.075	-0.064
			(0.072)	(0.079)
Cooperation of partner in previous round (= 1)			-0.158**	-0.150**
			(0.061)	(0.075)
Std. cognitive abilities			0.063	0.052
			(0.039)	(0.042)
Patience			-0.035	0.001
			(0.052)	(0.050)
Theory of mind (= 1)			0.055	-0.053
			(0.110)	(0.123)
SES				0.021
				(0.033)
Parental warmth				0.002
				(0.072)
Kindergarten FE	No	Yes	Yes	Yes
Observations	221	214	206	177

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable, and a dummy variable indicating whether a subject was punished in the round before (= 1) as the main independent variable. The sample consists only of subjects from TPP who kept their token (were selfish) in the previous round. The reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). Additional variables include the round, age as the number of months, gender dummy variable (girl = 1), number of siblings, dummy variable indicating whether partner from previous round cooperated, standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher number indicates higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** p<0.01. ** p<0.05. * p<0.1.

Table A11. OLS regression estimates of the effect of cooperation on the subject's round payoff using inverse probability weighting

	Dependent variable: subject's round payoff				
	CTR	IR	DR	TPP	All treatments
	(1)	(2)	(3)	(4)	(5)
Cooperation (= 1)	-0.882***	-0.819***	-0.938***	0.452***	-0.882***
	(0.068)	(0.064)	(0.109)	(0.098)	(0.068)
IR					-0.096*
					(0.050)
DR					-0.019
					(0.060)
TPP					-0.623***
					(0.087)
IR × cooperation					0.063
					(0.093)
DR × cooperation					-0.056
					(0.128)
TPP × cooperation					1.334***
					(0.118)
Constant	1.582***	1.486***	1.563***	0.960***	1.582***
	(0.041)	(0.029)	(0.044)	(0.078)	(0.041)
Observations	1,002	1,080	1,110	870	4,062
R-squared	0.160	0.139	0.179	0.041	0.132

The table reports regression results from OLS models using inverse probability weighting. A subject's round payoff is used as the dependent variable. Columns 1 to 4 use a dummy variable indicating whether the subject cooperated (= 1) as the main independent variable, while column 5 uses the dummy variable indicating whether the subject cooperated, the treatment dummy variables, and their interaction terms. The reported coefficients represent weighted least squares estimates. Weights are predicted inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

Table A12. Probit regression estimates of the effect of age on cooperation using inverse probability weighting

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age (in months)	-0.007***	-0.003	-0.003*	-0.005**	-0.001	-0.000	0.006*	0.007**	-0.007***	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.277	-0.174
									(0.171)	(0.196)
DR									-0.383**	-0.342*
									(0.167)	(0.189)
TPP									-0.428**	-0.323
									(0.218)	(0.212)
Age × IR									0.004	0.002
									(0.003)	(0.003)
Age × DR									0.006[§]	0.005[§]
									(0.002)	(0.003)
Age × TPP									0.012°	0.011°
									(0.003)	(0.003)
Observations	997	787	1,080	850	1,110	840	870	740	4,057	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable. Columns 1 to 8 use age (in months) as the main independent variable, while columns 9 and 10 use age (in months), treatment dummy variables, and their interaction terms. The reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** $p < 0.01$. ** $p<0.05$. * $p<0.1$.

[#] Control variables include the round, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

[§] In column 9, Age × DR interaction effect is positive and significant for all observations at the 5% level. In column 10, it is positive for all and significant at the 10% level for the majority of observations.

° Age × TPP interaction effect is positive and significant at the 1% level across all observations, both in columns 9 and 10.

Table A13. Probit regression estimates of the effect of cognitive abilities, patience, and theory of mind (TOM) on cooperation using inverse probability weighting

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Std. cognitive	-0.061***	-0.031	-0.019	0.019	-0.004	0.018	0.047**	-0.005	-0.060***	-0.044**
abilities	(0.020)	(0.021)	(0.014)	(0.020)	(0.017)	(0.018)	(0.023)	(0.029)	(0.020)	(0.021)
Patience	0.020	0.011	0.002	-0.003	-0.005	-0.002	0.049*	0.069***	0.020	0.020
	(0.017)	(0.021)	(0.016)	(0.022)	(0.015)	(0.015)	(0.029)	(0.023)	(0.017)	(0.021)
Theory of mind	-0.033	-0.020	0.003	0.012	0.025	0.007	0.100*	-0.002	-0.032	-0.008
(= 1)	(0.040)	(0.047)	(0.037)	(0.044)	(0.039)	(0.045)	(0.058)	(0.045)	(0.039)	(0.046)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.064	-0.053
									(0.054)	(0.066)
DR									-0.032	-0.043
									(0.055)	(0.064)
TPP									0.266***	0.297***
									(0.079)	(0.084)
Std. cog. abilities									0.047[§]	0.047[§]
× IR									(0.025)	(0.028)
Std. cog. abilities									0.061°	0.058°
× DR									(0.027)	(0.027)
Std. cog. abilities									0.122˜	0.092˜
× TPP									(0.034)	(0.035)
Patience × IR									-0.018	-0.006
									(0.024)	(0.030)
Patience × DR									-0.029	-0.027
									(0.023)	(0.026)
Patience × TPP									0.033	0.043
									(0.036)	(0.038)
TOM × IR									0.048	0.022
									(0.056)	(0.067)
TOM × DR									0.063	0.048
									(0.056)	(0.065)
TOM × TPP									0.157≈	0.129≈
									(0.077)	(0.077)
Observations	982	787	1,075	850	1,100	840	860	740	4,017	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable. Columns 1 to 8 use standardized cognitive abilities, patience (from 0 to 2; number represents the number of tokens saved for next day), and theory of mind dummy variable (= 1) as the main independent variables, while columns 9 and 10 use the same three variables, but also treatment dummy variables and their interaction terms. Reported coefficients represent (weighted) average marginal effects. Weights are predicted inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their reported error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

[#] Control variables include age as the number of months, round, gender dummy variable (girls = 1), number of siblings, SES as the highest education level of parents, parents' self-reported parental warmth, and kindergarten fixed effects.

[§] Std. cog. abilities × IR is positive for all observations and significant at the 10% level for the majority of observations, in both columns 9 and 10.

° Std. cog. abilities × DR is positive and significant for all observations (mostly at the 5% level), both in columns 9 and 10.

˜Std. cog. abilities × TPP is positive and significant (at least) at the 5% level for all observations, both in columns 9 and 10.

≈TOM × TPP is positive for all and significant for most observations, predominantly at the 5% level in column 9, and at the 10% level in column 10.

Table A14. Probit regression estimates of the effect of SES and parental warmth on cooperation using inverse probability weighting (correcting for selection in passing control questions)

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SES	-0.025**	-0.019	-0.012	-0.015	-0.005	-0.012	0.052***	0.044**	-0.024**	-0.020*
	(0.011)	(0.013)	(0.012)	(0.014)	(0.011)	(0.012)	(0.019)	(0.018)	(0.010)	(0.012)
Parental warmth	0.011	-0.002	0.033	0.026	-0.018	-0.043**	0.061	0.047	0.011	0.019
	(0.027)	(0.029)	(0.029)	(0.033)	(0.015)	(0.020)	(0.045)	(0.041)	(0.026)	(0.028)
Control variables[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.229	-0.156
									(0.244)	(0.277)
DR									0.106	0.297
									(0.244)	(0.225)
TPP									-0.323*	-0.213
									(0.177)	(0.229)
IR × SES									0.017	0.010
									(0.017)	(0.020)
DR × SES									0.016	0.009
									(0.016)	(0.015)
TPP × SES									0.062[§]	0.067[§]
									(0.047)	(0.037)
IR × parental warmth									0.020	0.012
									(0.038)	(0.041)
DR × parental warmth									-0.028	-0.052
									(0.031)	(0.036)
TPP × parental warmth									0.041	0.032
									(0.047)	(0.050)
Observations	807	787	850	850	850	840	745	740	3,252	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable. Columns 1 to 8 use SES as the highest education level of parents (from 1 to 8; higher number indicates higher education) and parents' self-reported parental warmth as the main independent variables, while columns 9 and 10 use SES, parental warmth, treatment dummy variables, and their interaction terms. Reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of passing the control questions, taking into account age and standardized cognitive abilities of the subject (see Appendix A.1.1 for more details on the construction of weights). The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their reported error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** p < 0.01. ** p<0.05. * p<0.1.

[#] Control variables include age as the number of months, round, a gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), and kindergarten fixed effects.

[§] Both in columns 9 and 10, TPP × round interaction effect is positive for all observations, and it is significant (at least) at the 5% level for all observations whose predicted probability of cooperating is more than 0.5. For observations whose predicted probability of cooperating is less than 0.5, the interaction effect is not significant for the large majority of observations in column 9, and significant for approximately half of observations (with a significance level ranging from 1% to 10%) in column 10.

Table A15: Probit regression estimates of the effects of the current partner's image score (cooperation percentage) on a child's cooperation

	Dependent variable: Cooperation (= 1)							
	IR				DR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Partner's image score	0.120***	0.108**	0.121***	0.106**	0.185***	0.164***	0.160**	0.126
	(0.042)	(0.043)	(0.043)	(0.050)	(0.054)	(0.054)	(0.064)	(0.077)
Round			-0.080***	-0.086***			-0.032***	-0.041***
			(0.012)	(0.014)			(0.012)	(0.015)
Cooperation of partner			-0.010	-0.015			0.017	0.020
in previous round (= 1)			(0.030)	(0.040)			(0.051)	(0.062)
Subject's image score			0.299***	0.290***			0.276***	0.280***
			(0.077)	(0.081)			(0.073)	(0.087)
Subject's cooperation in			-0.189***	-0.203***			-0.234***	-0.254***
previous round (=1)			(0.051)	(0.054)			(0.053)	(0.062)
Age			-0.003*	-0.005**			0.001	0.002
			(0.002)	(0.002)			(0.002)	(0.002)
Girl (= 1)			0.049	0.004			0.012	0.020
			(0.031)	(0.036)			(0.034)	(0.039)
Siblings			0.017	0.022			0.012	-0.022
			(0.016)	(0.022)			(0.025)	(0.023)
Std. cognitive abilities			-0.023	-0.001			0.009	0.021
			(0.016)	(0.021)			(0.020)	(0.021)
Patience			0.001	0.000			-0.015	-0.014
			(0.020)	(0.026)			(0.015)	(0.018)
Theory of mind (= 1)			0.049	0.069			0.036	0.051
			(0.034)	(0.043)			(0.045)	(0.043)
SES				-0.007				-0.006
				(0.014)				(0.012)
Parental warmth				0.019				-0.028
				(0.032)				(0.022)
Kindergarten FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	864	864	860	680	888	888	876	672

The table reports regression results from probit models using a dummy variable for cooperation (= 1) as the dependent variable and the partner's image score (where the value is calculated as the proportion of previous cooperation) as the main independent variable. The sample consists of subjects in IR (columns 1-4) and DR (columns 5-8). The coefficients represent average marginal effects. Additional variables include the round, a dummy variable indicating whether a partner from the previous round cooperated (= 1), a subject's image score (where the value is calculated as the proportion of previous cooperation), a dummy variable indicating whether a subject cooperated in the previous round (= 1), age as the number of months, a gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience (from 0 to 2; numbers represent the number of tokens saved for next day), a theory of mind dummy variable (= 1), SES as the highest education level of parents (from 1 to 8; higher numbers indicate higher education), parents' self-reported parental warmth, and kindergarten fixed effects. Clustered standard errors at the session level in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A16. Probit regression estimates of selection on observables concerning SES and parental warmth

	Dependent variable: parental SES and parental warmth variables obtained (= 1)
	(1)
Total amount of cooperation	0.009
	(0.009)
Siblings	-0.026
	(0.017)
Girl (= 1)	0.052*
	(0.028)
Age	-0.003*
	(0.002)
Std. cognitive abilities	0.019
	(0.016)
Kindergarten FE	Yes
p-value: Kindergarten FE coefficients = 0	0.004
Observations	805

The table reports regression results from a probit model using a dummy variable indicating whether the SES variable (highest educational level of the parents) and the parental warmth variable were obtained for the subject. Independent variables include a subject's total amount of cooperation over five rounds, number of siblings, gender dummy variable (girl = 1), age as the number of months, standardized cognitive abilities, and kindergarten fixed effects. The reported coefficients represent average marginal effects. Standard errors at the session level in parentheses. *** p<0.01. ** p<0.05. * p<0.1.

Table A17. Probit regression estimates of the effect of SES and parental warmth on cooperation using inverse probability weighting (correcting for selection in obtaining SES and parental warmth variables)

	Dependent variable: Cooperation (= 1)									
	CTR		IR		DR		TPP		All treatments	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SES	-0.024**	-0.019	-0.012	-0.016	-0.005	-0.011	0.053***	0.048***	-0.023**	-0.019*
	(0.010)	(0.012)	(0.012)	(0.014)	(0.011)	(0.012)	(0.019)	(0.017)	(0.010)	(0.011)
Parental warmth	0.009	-0.003	0.024	0.019	-0.015	-0.042**	0.052	0.046	0.009	0.015
	(0.027)	(0.028)	(0.030)	(0.033)	(0.014)	(0.021)	(0.043)	(0.040)	(0.026)	(0.027)
Controls[#]	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IR									-0.192	-0.135
									(0.280)	(0.286)
DR									0.084	0.292
									(0.257)	(0.237)
TPP									-0.298	-0.195
									(0.199)	(0.238)
IR × SES									0.015	0.008
									(0.018)	(0.019)
DR × SES									0.016	0.008
									(0.016)	(0.014)
TPP × SES									0.069[§]	0.069[§]
									(0.046)	(0.035)
IR × parental warmth									0.014	0.010
									(0.039)	(0.040)
DR × parental warmth									-0.023	-0.047
									(0.030)	(0.036)
TPP × parental warmth									0.036	0.030
									(0.046)	(0.050)
Observations	807	787	850	850	850	840	745	740	3,252	3,217

The table reports regression results from probit models using inverse probability weighting. A dummy variable for cooperation (= 1) is used as the dependent variable. Columns 1 to 8 use SES as the highest education level of parents (from 1 to 8; higher number indicates higher education) and parents' self-reported parental warmth as the main independent variables, while columns 9 and 10 use SES, parental warmth, treatment dummy variables, and their interaction terms. Reported coefficients represent (weighted) average marginal effects. Weights are predicted as inverse probabilities of obtaining the SES and parental warmth variables, taking into account age, gender, and kindergarten fixed effects (see Appendix A.1.2 for more details on the construction of weights). The omitted treatment category in columns 9 and 10 is the CTR treatment. Clustered standard errors at the session level in parentheses. The interaction coefficients in columns 9 and 10 represent average interaction effects and their reported error terms in parentheses represent average standard errors, calculated using the methodology of Norton et al. (2004). *** $p < 0.01$. ** $p<0.05$. * $p<0.1$.

[#] Control variables include age as the number of months, round, gender dummy variable (girl = 1), number of siblings, standardized cognitive abilities, patience, theory of mind dummy variable (= 1), and kindergarten fixed effects.

[§] Both in columns 9 and 10, TPP × round interaction effect is positive for all observations, and it is significant at the 5% level for all observations whose predicted probability of cooperating is more than 0.5. For observations whose predicted probability of cooperating is less than 0.5, the interaction effect is not significant for the majority of observations in column 9, and significant (predominantly at the 1% or the 5% level) for the majority in column 10.

References in online appendix

Bašić, Zvonimir, Armin Falk, and Fabian Kosse. 2020. "The Development of Egalitarian Norm Enforcement in Childhood and Adolescence." *Journal of Economic Behavior and Organization* 179: 667-80.

Falk, Armin, Fabin Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences." *Journal of Political Economy* 129 (9): 2504-2545.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2): 434-67.

Norton, Edward C., Hua Wang, and Chunrong Ai. 2004. "Computing Interaction Effects and Standard Errors in Logit and Probit Models." *The Stata Journal* 4 (2): 154-67.

References of papers used in the meta-analysis

Ahn, Toh Kyeong, Justin Esarey, and John T. Scholz. 2009. "Reputation and Cooperation in Voluntary Exchanges: Comparing Local and Central Institutions." *The Journal of Politics* 71 (2): 398-413.

Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica." *Journal of Public Economics* 92 (5-6): 1047-1060.

Akpalu, Wisdom, Babatunde Abidoye, Edwin Muchapondwa, and Witness Simbanegavi. 2017. "Public Disclosure for Carbon Abatement: African Decision-Makers in a PROPER Public Good Experiment." *Climate and Development* 9 (6): 548-558.

Barclay, Pat, and Nichola Raihani. 2016. "Partner Choice versus Punishment in Human Prisoner's Dilemmas." *Evolution and Human Behavior* 37 (4): 263-271.

Barrett, Scott, and Astrid Dannenberg. 2016. "An Experimental Investigation into 'Pledge and Review' in Climate Negotiations." *Climatic Change* 138: 339-351.

Boone, Christophe, Carolyn Declerck, and Toko Kiyonari. 2010. "Inducing Cooperative Behavior among Proselfs versus Prosocials: The Moderating Role of Incentives and Trust." *Journal of Conflict Resolution* 54 (5): 799-824.

Bornstein, Gary, and Ori Weisel. 2010. "Punishment, Cooperation, and Cheater Detection in "Noisy" Social Exchange." *Games* 1 (1): 18-33.

Brandts, Jordi, Arno Riedl, and Frans Van Winden. 2009. "Competitive Rivalry, Social Disposition, and Subjective Well-Being: An Experiment." *Journal of Public Economics* 93 (11-12): 1158-1167.

Bravo, Giangiacomo, and Flaminio Squazzoni. 2013. "Exit, Punishment and Rewards in Commons Dilemmas: An Experimental Study." *PloS one* 8 (8): e69871.

Brown-Kruse, Jamie, and David Hummels. 1993. "Gender Effects in Laboratory Public Goods Contribution: Do Individuals Put Their Money Where Their Mouth Is?." *Journal of Economic Behavior & Organization* 22 (3): 255-267.

Burton-Chellew, Maxwell N., Claire El Mouden, and Stuart A. West. 2017. "Evidence for strategic cooperation in humans." *Proceedings of the Royal Society B: Biological Sciences* 284 (1856): 20170689.

Caldwell, Michel D. 1976. "Communication and Sex Effects in a Five-Person Prisoner's Dilemma Game." *Journal of Personality and Social Psychology* 33 (3): 273-280.

Campos-Vazquez, Raymundo M., and Luis A. Mejia. 2016. "Does corruption affect cooperation? A laboratory experiment." *Latin American Economic Review* 25 (1): 1-19.

Cardenas, Juan Camilo. 2011. "Social Norms and Behavior in the Local Commons as Seen Through the Lens of Field Experiments." *Environmental and Resource Economics* 48 (3): 451-485.

Carpenter, Jeffrey, and Peter Hans Matthews. 2009. "What Norms Trigger Punishment?." *Experimental Economics* 12: 272-288.

Carpenter, Jeffrey P., and Peter Hans Matthews. 2012. "Norm enforcement: Anger, Indignation, or Reciprocity?." *Journal of the European Economic Association* 10 (3): 555-572.

Carpenter, Jeffrey, and Erika Seki. 2011. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." *Economic Inquiry* 49 (2): 612-630.

Casari, Marco, and Luigi Luini. 2009. "Cooperation Under Alternative Punishment Institutions: An Experiment." *Journal of Economic Behavior & Organization* 71 (2): 273-282.

Cason, Timothy N., and Lata Gangadharan. 2015. "Promoting Cooperation in Nonlinear Social Dilemmas Through Peer Punishment." *Experimental Economics* 18: 66-88.

Chen, Li. .2008. "An Experiment Study of the Effects of Sanction and Social Value Orientation on Trust and Cooperation." *Mimeo*.

Chen, Xiao-Ping, Madan M. Pillutla, and Xin Yao. 2009. "Unintended Consequences of Cooperation Inducing and Maintaining Mechanisms in Public Goods Dilemmas: Sanctions and Moral Appeals." *Group Processes & Intergroup Relations* 12 (2): 241-255.

Cherry, Todd L., and David M. McEvoy. 2013. "Enforcing Compliance with Environmental Agreements in the

Absence of Strong Institutions: An Experimental Analysis." *Environmental and Resource Economics* 54: 63-77.
Cox, Caleb A., and Brock Stoddard. 2015. "Framing and Feedback in Social Dilemmas With Partners and Strangers." *Games* 6 (4): 394-412.
Cubitt, Robin P., Michalis Drouvelis, and Simon Gächter. 2011. "Framing and Free Riding: Emotional Responses and Punishment in Social Dilemma Games." *Experimental Economics* 14: 254-272.
Danheiser, Priscilla R., and William G. Graziano. 1982. "Self-Monitoring and Cooperation as a Self-Presentational Strategy." *Journal of Personality and Social Psychology* 42 (3): 497-505.
De Geest, Lawrence R., John K. Stranlund, and John M. Spraggon. 2017. "Deterring Poaching of a Common Pool Resource." *Journal of Economic Behavior & Organization* 141: 254-276.
de Kwaadsteniet, Erik W., Erik van Dijk, Arjaan Wit, David De Cremer, and Mark de Rooij. 2007. "Justifying Decisions in Social Dilemmas: Justification Pressures and Tacit Coordination under Environmental Uncertainty." *Personality and Social Psychology Bulletin* 33 (12): 1648-1660.
Dekel, Sagi, and Sven Fischer. 2017. "Potential Pareto Public Goods." *Journal of Public Economics* 146: 87-96.
Dickinson, David L., David Masclet, and Marie Claire Villeval. 2015. "Norm Enforcement in Social Dilemmas: An Experiment With Police Commissioners." *Journal of Public Economics* 126: 74-85.
Drouvelis, Michalis, and Brit Grosskopf. 2016. "The Effects of Induced Emotions on Pro-social Behaviour." *Journal of Public Economics* 134: 1-8.
Duffy, John, and Jack Ochs. 2009." Cooperative Behavior and the Frequency of Social Interaction." *Games and Economic Behavior* 66 (2): 785-812.
Dugar, Subhasish. 2013. "Non-monetary Incentives and Opportunistic Behavior: Evidence From a Laboratory Public Good Game." *Economic Inquiry* 51 (2): 1374-1388.
Engelmann, Dirk, and Nikos Nikiforakis. 2015. "In the Long-Run We Are All Dead: On the Benefits of Peer Punishment in Rich Environments." *Social Choice and Welfare* 45: 561-577.
Eriksson, Kimmo, and Pontus Strimling. 2012. "The Hard Problem of Cooperation." *PloS one* 7 (7): e40325.
Eriksson, Kimmo, Pontus Strimling, and Micael Ehn. 2013. "Ubiquity and Efficiency of Restrictions on Informal Punishment Rights." *Journal of Evolutionary Psychology* 11 (1): 17-34.
Fatas, Enrique, and Guillermo Mateu. 2015. "Antisocial Punishment in Two Social Dilemmas." *Frontiers in Behavioral Neuroscience* 9: 107.
Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90 (4): 980-994.
Gächter, Simon, and Ernst Fehr. 1999. "Collective Action as a Social Exchange." *Journal of Economic Behavior & Organization* 39 (4): 341-369.
Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2010. "Culture and Cooperation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1553): 2651-2661.
Gaudeul, Alexia, and Caterina Giannetti. 2017. "The Effect of Privacy Concerns on Social Network Formation." *Journal of Economic Behavior & Organization* 141: 233-253.
Greiff, Matthias, and Fabian Paetzel. 2016. "Second-Order Beliefs in Reputation Systems with Endogenous Evaluations – An Experimental Study." *Games and Economic Behavior* 97: 32-43.
Grimalda, Gianluca, Andreas Pondorfer, and David P. Tracer. 2016. "Social Image Concerns Promote Cooperation More Than Altruistic Punishment." *Nature Communications* 7 (1): 12288.
Gürerk, Ozgur, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312 (5770): 108-111.
Hardy, Charlie L., and Mark Van Vugt. 2006. "Nice Guys Finish First: The Competitive Altruism Hypothesis." *Personality and Social Psychology Bulletin* 32 (10): 1402-1413.
Haruvy, Ernan, Sherry Xin Li, Kevin McCabe, and Peter Twieg. 2017. "Communication and Visibility in Public Goods Provision." *Games and Economic Behavior* 105: 276-296.
Hilbig, Benjamin E., Ingo Zettler, and Timo Heydasch. 2012. "Personality, Punishment and Public Goods: Strategic Shifts Towards Cooperation as a Matter of Dispositional Honesty–Humility." *European Journal of Personality* 26 (3): 245-254.
Howe, E. Lance, James J. Murphy, Drew Gerkey, and Colin T. West. 2016. "Indirect Reciprocity, Resource Sharing, and Environmental Risk: Evidence from Field Experiments in Siberia." *PloS one* 11 (7): e0158940.
Jacquet, Jennifer, Christoph Hauert, Arne Traulsen, and Manfred Milinski. 2011. "Shame and honour drive cooperation." *Biology Letters* 7 (6): 899-901.
Jerdee, Thomas H., and Benson Rosen. 1974. "Effects of opportunity to communicate and visibility of individual decisions on behavior in the common interest." *Journal of Applied Psychology* 59 (6): 712-716.
Joffily, Mateus, David Masclet, Charles N. Noussair, and Marie Claire Villeval. 2014. "Emotions, Sanctions, and Cooperation." *Southern Economic Journal* 80 (4): 1002-1027.
Kamijo Yoshio, Takeuchi Ai. 2009. "Voluntary Contribution Mechanism Hame and Endogenous Institution Selection." *The Waseda Journal of Political Science and Economics*. 368: 21-40.
Keser, Claudia, and Frans Van Winden. 2000. "Conditional Cooperation and Voluntary Contributions to Public Goods." Scandinavian Journal of Economics 102 (1): 23-39.

Kingsley, David C. 2016. “Endowment Heterogeneity and Peer Punishment in a Public Good Experiment: Cooperation and Normative Conflict.” *Journal of Behavioral and Experimental Economics* 60: 49-61.
Kingsley, David C., and Thomas C. Brown. 2016. “Endogenous and Costly Institutional Deterrence in a Public Good Experiment.” *Journal of Behavioral and Experimental Economics* 62: 33-41.
Kölle, Felix. 2015. “Heterogeneity and Cooperation: The Role of Capability and Valuation on Public Goods Provision.” *Journal of Economic Behavior & Organization* 109: 120-134.
Leibbrandt, Andreas, and Lauri Sääksvuori. 2012. “Communication in Intergroup Conflicts.” *European Economic Review* 56 (6): 1136-1147.
Liu, X., J. H. Ma, and Y. Zhu. 2010. “The Influence of Sanction System on Cooperation in Public Good Games from Attributional Perspective.” *Chinese Journal of Applied Psychology* 16: 332-340.
López-Pérez, Raúl, and Marc Vorsatz. 2010. “On Approval and Disapproval: Theory and Experiments.” *Journal of Economic Psychology* 31 (4): 527-541.
Lopez, Maria Claudia, James J. Murphy, John M. Spraggon, and John K. Stranlund. 2012. “Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia.” *Economic Inquiry* 50 (1): 131-142.
Maas, Alexander, Christopher Goemans, Dale Manning, Stephan Kroll, and Thomas Brown. 2017. “Dilemmas, Coordination and Defection: How Uncertain Tipping Points Induce Common Pool Resource Destruction.” *Games and Economic Behavior* 104: 760-774.
Marlowe, David, Kenneth J. Gergen, and Anthony N. Doob. 1966. “Opponent's Personality, Expectation of Social Interaction, and Interpersonal Bargaining.” *Journal of Personality and Social Psychology* 3 (2): 206-213.
Masclet, David, and Marie-Claire Villeval. 2008. “Punishment, Inequality, and Welfare: A Public Good Experiment.” *Social Choice and Welfare*: 475-502.
McCallum, Debra Moehle, Kathleen Harring, Robert Gilmore, Sarah Drenan, Jonathan P. Chase, Chester A. Insko, and John Thibaut. 1985. “Competition and Cooperation between Groups and between Individuals.” *Journal of Experimental Social Psychology* 21 (4): 301-320.
McCarter, Matthew W., Anya Samek, and Roman M. Sheremeta. 2014. “Divided Loyalists or Conditional Cooperators? Creating Consensus about Cooperation in Multiple Simultaneous Social Dilemmas.” *Group & Organization Management* 39 (6): 744-771.
Mengel, Friederike, and Ronald Peeters. 2011. “Strategic Behavior in Repeated Voluntary Contribution Experiments.” *Journal of Public Economics* 95 (1-2): 143-148.
Miettinen, Topi, and Sigrid Suetens. 2008. “Communication and Guilt in a Prisoner's Dilemma.” *Journal of Conflict Resolution* 52 (6): 945-960.
Milinski, Manfred, Dirk Semmann, Hans-Jürgen Krambeck, and Jochem Marotzke. 2006. “Stabilizing the Earth’s Climate is not a Losing Game: Supporting Evidence from Public Goods Experiments.” *Proceedings of the National Academy of Sciences* 103 (11): 3994-3998.
Morrison, Bruce John, Michael Engle, Toni Henry, Diana Dunaway, Michael Griffin, Kenneth Kneisel, and John Gimperling. 1971. “The Effect of Electrical Shock and Warning on Cooperation in a Non-zero-Sum Game.” *Journal of Conflict Resolution* 15 (1): 105-108.
Mulder, Laetitia B., Eric van Dijk, David De Cremer, and Henk A. M. Wilke. 2006. “Undermining Trust and Cooperation: The Paradox of Sanctioning Systems in Social Dilemmas.” *Journal of Experimental Social Psychology* 42 (2): 147-162.
Nelissen, Rob M.A., and Laetitia B. Mulder. 2013. “What Makes a Sanction “Stick”? The Effects of Financial and Social Sanctions on Norm Compliance.” *Social Influence* 8 (1): 70-80.
Nihonsugi, Tsuyoshi. 2016. “The Difference Between Partners and Strangers Designs in Public Goods Experiments.” *International Journal of Social Economics* 43 (6): 554-572.
Nihon Sugi, Tsuyoshi, and Koji Nakano. 2012. “An Analysis of the Difference in Contributions between Partners and Strangers Treatments in a Public Goods Supply Experiment: Measurement from Preferences and Beliefs.” *Osaka University Economics* 62 (2): 61-70.
Offerman, Theo, Joep Sonnemans, and Arthur Schram. 2001. “Expectation Formation in Step-Level Public Good Games.” *Economic Inquiry* 39 (2): 250-269.
Peeters, Ronald, and Marc Vorsatz. 2013. “Immaterial Rewards and Sanctions in a Voluntary Contribution Experiment.” *Economic Inquiry* 51 (2): 1442-1456.
Ramalingam, Abhijit, Sara Godoy, Antonio J. Morales, and James M. Walker. 2016. “An Individualistic Approach to Institution Formation in Public Good Games.” *Journal of Economic Behavior & Organization* 129: 18-36.
Rege, Mari, and Kjetil Telle. 2004. “The Impact of Social Approval and Framing on Cooperation in Public Good Situations.” *Journal of public Economics* 88 (7-8): 1625-1644.
Reuben, Ernesto, and Arno Riedl. 2013. “Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations.” *Games and Economic Behavior* 77 (1): 122-137.
Rodriguez-Sickert, Carlos, Ricardo Andrés Guzmán, and Juan Camilo Cárdenas. 2008. “Institutions Influence Preferences: Evidence from a Common Pool Resource Experiment.” *Journal of Economic Behavior & Organization* 67 (1): 215-227.
Sääksvuori, Lauri, Tapio Mappes, and Mikael Puurtinen. 2011. “Costly Punishment Prevails in Intergroup

Conflict." *Proceedings of the Royal Society B: Biological Sciences* 278 (1723): 3428-3436.
Savikhin Samek, Anya, and Roman M. Sheremeta. 2014. "Recognizing Contributors: An Experiment on Public Goods." *Experimental Economics* 17: 673-690.
Schroeder, David A., Thomas D. Jensen, Andrew J. Reed, Debra K. Sullivan, and Michael Schwab. 1983. "The Actions of Others as Determinants of Behavior in Social Trap Situations." *Journal of Experimental Social Psychology* 19 (6): 522-539.
Shinada, Mizuho, and Toshio Yamagishi. 2007. "Punishing Free Riders: Direct and Indirect Promotion of Cooperation." *Evolution and Human Behavior* 28 (5): 330-339.
Silverman, Dan, Joel Slemrod, and Neslihan Uler. 2014. "Distinguishing the Role of Authority "in" and Authority "to"." *Journal of Public Economics* 113: 32-42.
Skatova, Anya, and Eamonn Ferguson. 2013. "Individual Differences in Behavioural Inhibition Explain Free Riding in Public Good Games When Punishment Is Expected but Not Implemented." *Behavioral and Brain Functions* 9: 3.
Slusher, E. Allen, Kenneth J. Roering, and Gerald L. Rose. 1974. "The Effects of Commitment to Future Interaction in Single Plays of Three Games." *Behavioral Science* 19 (2) : 119-132.
Slusher, E. Allen, Gerald L. Rose, and Kenneth J. Roering. 1978. "Commitment to Future Interaction and Relative Power under Conditions of Interdependence." *Journal of Conflict Resolution* 22 (2): 282-298.
Tenbrunsel, Ann E., and David M. Messick. 1999. "Sanctioning Systems, Decision Frames, and Cooperation." *Administrative Science Quarterly* 44 (4): 684-707.
Tornatzky, Louis, and P. James Geiwitz. 1968. "The Effects of Threat and Attraction on Interpersonal Bargaining." *Psychonomic Science* 13 (2): 125-126.
Van Lange, Paul AM, Anthon Klapwijk, and Laura M. Van Munster. 2011. "How the Shadow of the Future Might Promote Cooperation." *Group Processes & Intergroup Relations* 14 (6): 857-870.
Van Vugt, Mark, and Wendy Iredale. 2013. "Men Behaving Nicely: Public Goods as Ceacock Tails." *British Journal of Psychology* 104 (1): 3-13.
Walker, James M., and Matthew A. Halloran. 2004. "Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings." *Experimental Economics* 7: 235-247.
Wang, Pei, and Li Chen. 2011. "The Effects of Sanction and Social Value Orientation on Trust and Cooperation in Public Goods Dilemmas." *Acta Psychologica Sinica* 43 (1): 52–64.
Weng, Qian, and Fredrik Carlsson. 2015. "Cooperation in Teams: The Role of Identity, Punishment, and Endowment Distribution." *Journal of Public Economics* 126: 25-38.
Wolf, Scott T., Taya R. Cohen, Jeffrey L. Kirchner, Andrea Rea, r. Matthew Montoya, and Chester A. Insko. 2009. "Reducing intergroup conflict through the consideration of future consequences." *European Journal of Social Psychology* 39 (5): 831-841.
Wu, Junhui, Daniel Balliet, and Paul AM Van Lange. 2016. "Gossip Versus Punishment: The Efficiency of Reputation to Promote and Maintain Cooperation." *Scientific Reports* 6 (1): 23919.
Wu, Jia-Jia, Chong Li, Bo-Yu Zhang, Ross Cressman, and Yi Tao. 2014. "The Role of Institutional Incentives and the Exemplar in Promoting Cooperation." *Scientific Reports* 4 (1): 6421.
Xiao, Erte, and Daniel Houser. 2011. "Punish in Public." *Journal of Public Economics* 95 (7-8): 1006-1017.
Xiao, Erte, and Howard Kunreuther. 2016. "Punishment and Cooperation in Stochastic Social Dilemmas." *Journal of Conflict Resolution* 60 (4): 670-693.
Xu. 2012. "Effects of Social Norm and Sense of Honor and Disgrace on Cooperation in Social Dilemmas." *Mimeo*
Xu, Qi, and Jiajun Wei. 2013. "The Effect of Sanctioning Systems on Trust and Cooperative Behaviors in Public Goods Dilemmas." *Conference on Psychology and Social Harmony*: 109-112.
Yamagishi, Toshio. 1988. "Seriousness of Social Dilemmas and the Provision of a Sanctioning System." *Social Psychology Quarterly*: 32-42.
Ye. 2007. "The Study on the Part Played by Reasoning Ability and Punishment in Public Goods Dilemma." Mimeo
Yoeli, Erez, Moshe Hoffman, David G. Rand, and Martin A. Nowak. 2013. "Powering up with indirect reciprocity in a large-scale field experiment." *Proceedings of the National Academy of Sciences* 110 (supplement 2): 10424-10429.
Zhang, Bo-Yu, Song-Jia Fan, Cong Li, Xiu-Deng Zheng, Jian-Zhang Bao, Ross Cressman, and Yi Tao. 2016. "Opting Out against Defection Leads to Stable Coexistence with Cooperation." *Scientific Reports* 6 (1):35902.
Zhang, Su, Wei Gao, and Binbin Fan. 2015. "Cognitive Ability and Cooperation: Evidence from the Public Goods Experiments." *Annals of Economics & Finance* 16 (1): 43-68.