Third party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation

Philipp Lergetporer[a,b], Silvia Angerer[a], Daniela Glätzle-Rützler[a] and Matthias Sutter[c,d]

03 February 2014

[a] University of Innsbruck, Department of Public Finance, A-6020 Innsbruck, Austria.

[b] Ifo Institute at the University of Munich, D-81679 Munich, Germany.

[c] European University Institute, Department of Economics, I-50133 Firenze, Italy.

[d] University of Cologne, Department of Economics, D-50923 Köln, Germany.

The human ability to establish cooperation even in large groups of genetically unrelated strangers depends upon the enforcement of cooperation norms. Third party punishment is one important factor to explain high levels of cooperation among humans whereas there seems to be no evidence that other animal species use this mechanism for promoting cooperation. We study the effectiveness of third party punishment to increase children's cooperative behavior in a large scale cooperation game. Based on an experiment with 1,120 children, aged seven to eleven years, we find that the threat of third party punishment more than doubles cooperation rates, despite the fact that children are rarely willing to execute costly punishment. We can show that the higher cooperation levels with third party punishment are driven by two components. First, cooperation is a rational (expected payoff-maximizing) response to incorrect beliefs about the punishment behavior of third parties. Second, cooperation is a conditionally cooperative reaction to correct beliefs that third party punishment will increase a partner's level of cooperation.

Significance statement

Cooperation among humans depends strongly upon the willingness of others to take costly action in order to enforce the social norm to cooperate. Such enforcement behavior is often coined third-party punishment. Here we show that third-party punishment is already effective as means to increase cooperation in children. Most importantly, we can identify why this is the case. First, children expect (mistakenly) third parties to punish quite often and therefore become more cooperative. Second, however, the presence of third parties lets children become (rightfully) more optimistic about the cooperation levels of the interaction partner in a simple prisoner's dilemma game. As a reaction to more optimistic expectations, children cooperate more themselves. The experiment has been run with about 1,100 children aged seven to eleven years.

Human cooperation rates among genetically unrelated strangers in large groups are unusually high and exceed cooperation among all other animal species by far (1,2). Already in childhood, humans begin to conform to cooperative social norms (3). This raises the question how such social norms are enforced, because norm enforcement may explain why humans cooperate more than other animals. While in repeated interactions reciprocity (4,5) may account for the higher cooperation rates – given that, unlike chimpanzees, humans become reciprocal already in early childhood (6) – in non-repeated (i.e., one-shot) settings a different mechanism must be at work (7). In fact, a growing body of literature suggests that the punishment of defectors is key to trigger and sustain cooperation in such contexts (7-11).

Punishment can take on the form of second party punishment, where those who are the victims of defection can punish norm-violators (7,12-18), or third party punishment, where unaffected bystanders can execute sanctions against norm-violators, even though the bystanders are not materially affected by a norm violation (8,19-29). While both humans and other animals use second party sanctioning to promote cooperation (30,31), it is interesting to note that humans also engage in third party punishment in order to increase cooperation rates. In contrast, it is unclear whether aggression against conspecifics among non-human animals ever aims to punish non-cooperators or whether such behavior can even be called third-party punishment at all.(*) Some recent evidence even shows that chimpanzees do not engage in third party punishment (36). Thus, the deliberate punishment of defectors by third parties is a likely candidate to explain higher cooperation rates in humans. In this paper, we study the ontogeny of cooperation and third party punishment during childhood, focusing on the questions whether third parties increase cooperation rates already in children and, if they do, for which reasons. So far, the experimental literature has documented positive effects of third party punishment on human cooperation exclusively for adult (typically student) populations (26-29). Since cooperation norms become internalized much earlier, already in childhood (3,37), we consider it as important to study how norm enforcement works in childhood. In their recent paper on the ontogeny of cooperative behavior (unrelated to norm enforcement through punishment), House et al. (3) conclude that future research should examine "institutions that influence cooperative behavior and how their acquisition and application shapes children's behavior across development" (p. 14590). We work in this direction by studying the effects of punishment institutions on the cooperative behavior of children.

We are particularly interested in whether children become more cooperative when a third party (of the same age) may punish them. If so, we try to disentangle the reasons for such a behavioral response by examining whether children become more cooperative because they are afraid of getting punished, or because they expect the partner in the cooperation game to cooperate in the presence of a third party, in which case third party punishment works through the channel of conditional cooperation (38-40). Studying the behavior of children, as players in a cooperation game or as third parties, allows determining whether norm enforcement through third party punishment works already at a young age where many norms are only beginning to become internalized (3,41). This is of particular interest as potential third party intervention is important among peers in school (for instance, punishment threats of peers towards free-riders in cooperative learning environments in order to foster student's commitment; 42), but it is unclear to date whether it shifts expectations about others' behavior or whether the third party punishment itself promotes norm enforcement.

The experiment was run in the city of Meran/Italy, with more than 1,100 primary school children, aged seven to eleven years. We chose these age cohorts because (i) important behavioral and economically relevant traits evolve during this period of life (43), (ii) peer interactions in primary school classes prepare children for their adult roles, teaching each other values and attitudes such as cooperation (44,45), and (iii) middle childhood (starting from around age 6) may be when children begin to conform to cooperative social norms (3). A necessary prerequisite for strategic interaction experiments to provide reliable results is that participants can understand them and, in our case, have developed the ability to take another person's perspective. In contrast to non-human primates like chimpanzees who "do not have a full-blown human-like theory of mind" (46, p. 156) both conditions are entirely met in humans in the age cohorts considered in this paper (3,46-49). The participating children represent 86% of all primary school children in grades two to five in this city with its 38,000 inhabitants.

Following previous literature on adults (26), we let our subjects play a one-shot, simultaneous prisoner's dilemma game as a baseline (see Fig. 1 for an illustration of the game and Methods as well as Supplementary Information, SI, for details). Mutual cooperation yields the Pareto-efficient outcome of [4,4]. However, both players have a dominant strategy to defect, leading to the inefficient Nash-equilibrium of [2,2].

554 children played this two-player game in a control treatment (CTR) without any third party. Matching was random and anonymous and pairs were always formed from the same age cohort. After having played the game once (but before being informed about the choice of their partner in the PD), these children were asked to act as third parties for another set of children. Children did not know about this additional task before completing play in the PD.

A different set of 566 children was assigned to a third-party punishment treatment (TPP). Children in TPP were randomly and anonymously paired (within their age cohort) and then played the PD once. Each child in a pair of TPP was assigned one child from CTR as the third party, and children in TPP were aware of this before making decisions. Of course, they were not informed about the third party's decision before choosing to cooperate or defect. The third party (the child in CTR) had to decide whether to invest a token to punish the assigned child (in TPP) in case this child would defect in the PD (†). As a consequence of punishment, the child in TPP lost all gains from the PD-experiment if it had defected. If the child in TPP had cooperated, or if the third party had not invested its token, then the third party kept the token which could be exchanged into a reward (‡).

After having made their own decisions, we asked children about their beliefs. Both in CTR and in TPP they were asked whether they expected the partner in their pair to cooperate. In TPP they were additionally asked whether they expected the third party to punish defection. Correct guesses were rewarded with one token. By asking children in TPP both about the expected punishment and the expected cooperation of their partner, we can check whether they cooperate in order to avoid punishment or because they expect the partner to cooperate as well.

Results

Fig. 2 shows cooperation rates across the four age cohorts for players in CTR and in TPP. Overall, we find that 58% of players cooperate in the one-shot PD in TPP, while only 25% do so in CTR (P = 0.000 overall and in each age group separately, χ^2-tests). This means that the presence of a third party with an opportunity to punish defectors more than doubles cooperation rates. Looking at cooperation rates across age cohorts, we find no significant age effects within any treatment (P = 0.339 in CTR and P = 0.552 in TPP, Cuzick's Wilcoxon-type test for trend).

Fig. 3 illustrates players' beliefs about their partner's likelihood of cooperation. We find that beliefs significantly differ across treatments: 64% of subjects in TPP, but only 51% in CTR belief that their partner will cooperate (P=0.000 across all age groups, χ^2-test). This means that subjects anticipate that the presence of third parties will have an impact on the partner's willingness to cooperate. In TPP, in all four age cohorts the expected likelihood of cooperation matches the actual rate of cooperation fairly closely. Comparing the dark grey bars across Fig. 2 and Fig. 3 does not yield significant differences in any age cohort (P = 0.16 for 7/8 years; P = 0.16 for 8/9 years; P = 0.09 for 9/10 years; P = 0.42 for 10/11 years;

McNemar's tests). However, in CTR all four age cohorts are too optimistic, because expected cooperation rates are always significantly higher than actual cooperation rates (see the light grey bars in Fig. 2 and Fig. 3; $P < 0.01$ in each cohort; McNemar's tests). This latter result suggests an intention to free-ride on the (expected) contributions of partners, which implies a willingness to accept advantageous inequality (50,51). As soon as a third party is present (in TPP), however, expectations and actual behavior with respect to cooperation get well calibrated. This effect can explain why (potential) third party punishment increases cooperation rates if subjects are conditional cooperators (38-40). For someone who conditions the level of cooperation on the interaction partner's willingness to cooperate, third party punishment shifts the expectations upwards, and hence triggers more cooperation, even in the absence of actual punishment. In fact, Fig. S3 in SI shows that the average expectations of conditional cooperators (that is, subjects whose belief about the cooperative behavior of the partner is aligned with their own decision) are significantly higher in TPP than in CTR ($P = 0.000$ in each cohort; χ^2-tests).

Fig. 4 juxtaposes actual and expected punishment rates. It turns out that third parties use the punishment option very rarely, overall in less than 10% of cases (§). Players in TPP expect third parties to punish in 51% of cases on average, however. The difference is highly significant throughout ($P = 0.000$ in each age cohort; χ^2-tests), indicating a strong mismatch between beliefs and actual punishment behavior. A similar mismatch, albeit of smaller size, has been found in previous studies of third party punishment when subjects share a pie very unevenly in a simple allocation task (26). Hence, the mismatch is not an artefact of our design or subject pool (¶). Moreover, the fact that only a small fraction of primary school children incurs costs in order to punish defectors is consistent with the finding that children's behavior is typically closer to payoff-maximization than adult behavior (53; see also the evidence of payoff-maximizing behavior of chimpanzees; 55).

It is interesting to note that, given *actual* punishment behavior, players in TPP have higher expected payoffs from defection than from cooperation in all age cohorts, meaning that it would be a payoff-maximizing strategy to defect. Hence, if only actual punishment was important, it should not have any effect on cooperation rates (contrary to what we see in Fig. 2). However, given *expected* punishment rates, cooperation yields higher expected payoffs than defection for all cohorts, except for the oldest (where cooperation and defection have practically the same expected payoff; see Tab. S2 in SI).

Hence, cooperation in TPP is driven by two components. First, it is a rational (expected payoff-maximizing) response to incorrect beliefs about the punishment behavior of third parties. Second, it becomes more likely as a conditionally cooperative reaction to an increase in the expected cooperation rate of a subject's partner. The latter increase, in turn, is due to the presence of third parties (||).

From a societal point of view, TPP is more efficient than CTR. Given the actual cooperation rate of 24.6%, the expected payoff of a player is 2.49 tokens in CTR. Taking into account the 8% chance of losing all earnings through punishment, and considering the cooperation rate of 58% in TPP, a player in TPP earns on average 3.01 tokens. Subtracting from this the average costs of 0.03 tokens for the third party through punishing (8% of the 42% of defectors get punished by third parties, which costs them 1 token), yields a net surplus of 2.98 tokens, which is 20% higher than in CTR.

Discussion

Third party punishment increases cooperation already among children, aged seven to eleven years. Across these age cohorts, we have found no significant differences in reactions to potential punishment. Most noteworthy, third party punishment works through two channels, one of which relies on a misalignment of actual and expected punishment behavior. Subjects expect to get punished for defection much more often than third parties are actually willing to incur the costs of punishment. Yet, the expectation of punishment suffices to increase cooperation rates. This misalignment between actual and expected punishment may, in fact, explain why field data suggest that third party sanctioning is hardly observed whereas cooperation rates are found to be substantial at the same time (56). The second channel through which third parties increase cooperation rates is their effect on expected cooperation rates of other players. The presence of third parties with a punishment option is expected to make *others* more cooperative, which in turn triggers *own* cooperation as a consequence of conditional cooperation. In fact, while the prevalence of conditional cooperation has been shown for adults (38,39), our study can be interpreted as showing that already children are conditional cooperators. Moreover, our study establishes a link between conditional cooperation and the cooperation-enhancing effect of third party punishment.

Among the avenues for future research we see three straightforward extensions of our work. First, it would be interesting to see whether third party reward is equally efficient in increasing cooperation as is third party punishment or whether positive and negative incentives work differently (57). Second, it would be a worthwhile project to study even younger children than we did in this paper. We consider it an intriguing question whether at a very early age (potential) third party punishment would be executed on the one hand, and would be effective to increase cooperation on the other hand. Third, studying how the presence of third party observers who cannot punish affects cooperation in children would be interesting. As we measure the joint effect of observation and punishment on cooperation rates, disentangling both channels contributes to the understanding whether the presence of third parties, the possibility of punishment, or the interaction of both promotes cooperation among humans.

Methods

We conducted our experiment in all fourteen elementary schools in Meran (South Tyrol, Italy) in November 2012. Meran is the second largest city in the province of South Tyrol, with about 38,000 inhabitants, of which roughly 50% are Italian speaking and 50% German speaking. Our experiment was part of a larger research project which investigated economic decision making of elementary school children. In Italy, elementary school comprises grades 1 to 5.

Before starting the project we obtained approval from the Internal Review Board of the University of Innsbruck, the South Tyrolean State Board of Education, and from the headmasters as well as informed consent from the parents of the involved children to run a series of six experimental sessions in the academic years 2011/12 and 2012/13. We got permission from 86% of parents of all elementary school children in Meran. The permissions were either granted implicitly or explicitly. Each school district separately decided whether the parents had to sign a consent form in order to give their child the permission to participate (opt-in) or whether participation was implicit and the parents had to sign in order to prohibit participation (opt-out). One out of five school districts decided to implement the explicit participation consent form, the others implemented the implicit participation. In all cases, parents received a letter explaining the general purpose of the two-year research project prior to the start of the experiments. Participation in each experimental session was, of course, voluntary for children, but all except a single child consented to participate.

The experiment on cooperation and punishment was the second experiment conducted with the children in the second year of the study. In that year we worked with children in grades 2 to 5 (while we had grades 1 to 5 in the first year). In total, we had 1,141 children participating in this experiment.

Each child was fetched individually from the classroom and brought to a separate room where an experimenter explained the experiment one-to-one to the child. In this room, there were four to eight experimenters running the experiment with four to eight children at the same time, visually separated from one another. Treatments were randomized within each experimenter and experimenters had to memorize the instructions of the game and explain the game orally (in the mother-tongue of the child), with detailed visual support (see SI for experimental instructions). The duration of the experiment was approximately 20 minutes and it was conducted with pen and paper. Following the general procedure when conducting experiments with children (43,58), children had to repeat the rules of the game in their own words after the explanation by the experimenter. In case of mistakes, the experimenter repeated the respective passages, and asked the child to repeat the rules once more. 21 children did not manage to correctly repeat the rules, in particular the consequences of each combination of actions. Given our one-on-one explanation technique, this is a reasonable rate (43), leaving us with 1,120 children with full understanding for the analysis (see Tab. S1). Of course, the 21 children without correct understanding were allowed to participate in the experiment until the end. Including their choices would not change any of our results.

A subject either participated in treatment CTR or TPP. Both experimental treatments had two stages. This was not known to children at the beginning of the experiment. Only at the end of the first stage (after having made all decisions and after having answered our questions on expectations) they were informed about the second stage and its rules. Children were informed about the outcome of the two stages of the experiment only three months later when they received the presents (**).

In each stage, a child was anonymously matched with another child from the same age cohort (i.e., grade) and language group, but from a different school. This was common knowledge. The baseline game in the experiment was a one-shot prisoner's dilemma (shown in Fig. 1). In this game, one child and its partner were endowed with two tokens each, which could be either kept or passed on to the other player. In the latter case, the tokens were doubled. This game was played in the first stage of treatment CTR where no external observer was present, hence the game was played without a third party who could punish defection. The prisoner's dilemma game without third party was also played in the *second* stage of treatment TPP. This second stage in TPP served as a within-subjects control whether subjects who had experienced third party punishment in stage one would change their behavior when third party punishment was removed. The change was as expected, with cooperation rates dropping significantly in the absence of third party punishment, and the data are shown in Fig. S1 and Fig. S2 in SI. In the paper we do not report the second stage of treatment TPP.

In the first stage of treatment TPP the prisoner's dilemma game was extended to a third-party punishment experiment: each player in the prisoner's dilemma of this stage was paired with an exclusive third-party observer. This was a subject in stage 2 of CTR. The observer was *not* affected by the play of the children in treatment TPP, but was endowed with one token. This token could either be kept by the child in stage 2 of CTR or could be spent to destroy the whole payoff of the paired player in TPP *if* this player chose defection. Since children in the role of an observer had played the game themselves before (in stage 1 of CTR), they were familiar with the rules of the game and could easily condition their decision on the paired player's choice to cooperate or defect. Of course, the observers (in CTR) were not informed about the actual choice of the observed player (in TPP) before making their decision on how to spend the token, meaning that we implemented a so-called strategy method (52). The decision of the observer was only implemented in case of defection. Thus, we did not allow for spiteful punishment (61), because we are primarily interested in the enforcement of a cooperation norm and not whether children are willing to punish cooperative acts. All involved participants were exactly informed about the punishment mechanism. It is also noteworthy that the observed players knew that their partner in the prisoner's dilemma also faced a punishment threat in case of defection, because the partner had also one (different) child assigned as an observer with an opportunity to punish defection. At the very end of the session children completed a post-experimental questionnaire on demographic data (on siblings, gender and age). Total earnings in the experiment were determined by actual decisions and also by the stated expectations. The latter were also incentivized. Subjects earned an extra token per correct guess.

As incentives, we used sweets (lollipops, small chocolates, candies), fruits (bananas, apples, oranges) and other small presents (stickers, balloons, pencils, wristbands). Children could exchange the tokens earned in the experiment into items of their choice in a so-called "experiment-store". The cost of each item ranged from one to three tokens.

As control variables, we measured the IQ and the extent of altruism and intertemporal preferences of our participants one to six months before the experiment on third party punishment. IQ was elicited with a shortened version of Raven's test. Altruism was elicited in a dictator game. Subjects were endowed with 6 tokens and we let them decide anonymously how many tokens to keep for themselves (and exchange them into presents in the "experiment-store") or to donate to one of the province's largest charities, "Menschen in Not: Kinderarmut durch Kinderreichtum", respectively "Umanità che ha bisogno: famiglia numerosa = famiglia povera?", an initiative to support underprivileged children in South Tyrol. This charity is run by the well-known Caritas diocese Bolzano-Bressanone. For each token allocated to the charity we donated 50 Euro Cents to the charity. Intertemporal preferences were measured with the use of a choice list. Each child had to make three decisions in which to choose either 2 tokens at the end of the experiment or a larger number of tokens with a delay of 4 weeks. The delayed payoff was either 3 tokens, 4 tokens, or 5 tokens. From these choices, we can identify very impatient children (who always choose the 2 tokens immediately) and check whether they behave differently from more patient children. We have data for 977 children who participated in the third party punishment experiment and in both the altruism experiment and the intertemporal choice task, and these children are the basis for the regressions shown in Tab. S3 in SI.

Acknowledgements

The authors thank Rudolf Meraner from the South Tyrolean State Board of Education (*Pädagogisches Institut für die deutsche Sprachgruppe in Südtirol*), the headmasters of the participating schools (Gabriella Kustatscher, Maria Angela Madera, Eva Dora Oberleiter, Brigitte Öttl, Ursula Pulyer, Vally Valbonesi), the parents of the involved children for making this study possible, and the children for participation. We received helpful comments from two referees, Loukas Balafoutas, Nikos Nikiforakis, Karl Sigmund, the audiences at the Maastricht Behavioral and Experimental Economics Symposium 2013 and the University of Innsbruck. Financial support from the Government of South Tyrol and the "Aktion D. Swarovski" at the University of Innsbruck is gratefully acknowledged.

Footnotes

* Several studies report behavior which resembles third-party punishment in non-human animals (32-35). It is likely, however, that such behavior is motivated by selfishness and yields cooperation only as a by-product (36).

† While both altruistic and spiteful punishment, i.e. the punishment of defectors and cooperators, is usually permitted in experiments with adults (26) we restricted our participants' action space to altruistic punishment, because we are primarily interested in the enforcement of a social norm to cooperate and not whether children are willing to punish cooperative acts.

‡ We chose this binary punishment technology in order to assure comprehension. While adults who act as third parties in such games are usually asked to choose the number of tokens to be invested into punishment (26,29), we kept the design as simple as possible and thus only allowed for a binary decision.

§ One possible reason for the low incidence of punishment rates could be the use of the strategy method to decide upon punishment. Third parties had to make decisions conditional on the player in the PD game choosing defection. This may leave the third party in an emotionally "cold" state, while choosing punishment after having seen the player defect in the PD game may create an emotionally "hot" state and then trigger more punishment. A recent survey comparing the strategy method with the latter type of direct-response method has failed to find a systematic behavioral impact of the strategy method, however (52). In their study on third-party-punishment with adults, Fehr and Fischbacher (26) also use the strategy method for eliciting the decisions of third-party observers and find between 21% and 50% punishment rates (depending on the decision of the interaction partner in the PD). Our somewhat smaller punishment rates are compatible with the finding that children's behavior is typically closer to payoff-maximization (which predicts no punishment) than adult behavior (53).

¶ One might also argue that the mismatch between beliefs about the behavior of third parties and actual punishment rates is due to the difficulty of children playing the PD game to put themselves into the role of the third party and take her perspective. However, psychological studies on the development of Theory of Mind (47,54) show that normally developing children are able to differentiate the other's view from their own one by the age of four to six years, and then take the perspectives of other persons into account.

|| In SI we show support for this with a regression (in Tab. S3) in which the likelihood to cooperate is the dependent variable. Expecting the partner to cooperate increases a subject's likelihood of cooperation by 42 percentage points ($P = 0.000$), and expecting the third party to punish defection raises the likelihood of cooperation by 38 percentage points ($P = 0.000$). A Wald-test shows that both factors are equally strong and not significantly different from each other. In this regression, we also control for age, gender, IQ and other covariates. Of the latter, only altruistic giving in an experiment on voluntary donations to a charity turns out to be significant (and positive, as expected).

** As the total earnings of each child were dependent not only on own choices, but also on the decision of the partner in the experiment (who was from another school), it was not possible to calculate the final earnings of the children immediately at the end of a session. Thus the tokens earned in the experiment were handed over at our next visit, three months after this experiment took place. Given our delayed payment procedure and the high discount rates among children (59,60) we checked whether impatient children behaved differently from more patient ones, because the former might have perceived the incentives in the experiment as less valuable than the latter. To tackle this important issue (suggested by a referee) we compared the choices (in TPP and CTR) of very impatient children with their more patient peers (we measured patience in an independent experiment on intertemporal preferences conducted six month before the experiment for the present paper was conducted). Table S3 in SI shows that children who are categorized as more impatient do not behave differently in the PD game.

References

1. Richersen PJ, Boyd R (2005) *Not by Genes Alone: How Culture Transformed Human Evolution* (Univ of Chicago Press, Chicago).
2. Fehr E, Fischbacher U (2004) Social norms and human cooperation. *Trends Cogn Sci* 8: 185-190.
3. House B et al. (2013) Ontogeny of prosocial behavior across diverse societies. *Proc Natl Acad Sci USA* 110: 14586-14591.
4. Hamilton WD (1964) The genetical evolution of social behavior. I, II. *J Theor Biol* 7: 1-52.
5. Trivers RL (1971) The evolution of reciprocal altruism. *Q Rev Biol* 46: 35-57.
6. House B, Henrich J, Sarnecka B, Silk JB (2013) The development of contingent reciprocity in children. *Evol Hum Behav* 34: 86-93.
7. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415: 137-140.
8. Boyd R, Gintis H, Bowles S, Richerson P J (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100: 3531-3535.
9. Hauert C, Traulsen A, Brandt H, Nowak M A, Sigmund K (2007) Via freedom to coercion: The emergence of costly punishment. *Science* 316: 1905-1907.
10. Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328: 617-620.
11. Mussweiler T, Ockenfels A (2013) Similarity increases altruistic punishment in humans. *Proc Natl Acad Sci USA* Early Edition: 1-6.
12. Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Am Econ Rev* 90: 980-994.
13. Nikiforakis N, Normann H T (2008) A comparative statics analysis of punishment in public-good experiments. *Exp Econ* 11: 358-369.
14. Nikiforakis N (2010) Feedback, punishment and cooperation in public good experiments. *Games Econ Behav* 68: 689-702.
15. Dreber A, Rand D G, Fudenberg D, Nowak M A (2008) Winners don't punish. *Nature* 452: 348-351.
16. Yamagishi T (1986) The provision of a sanctioning system as a public good. *J Pers Soc Psychol* 51: 110-116.
17. Gächter S, Herrmann B (2009) Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Phil Trans R Soc B* 364: 791-806.
18. Gürerk Ö, Irlenbusch B, Rockenbach B (2006) The competitive advantage of sanctioning institutions. *Science* 312: 108-111.
19. Gintis H (2000) Strong reciprocity and human sociality. *J Theor Biol* 206: 169-179.
20. Fehr E, Fischbacher U, Gächter S (2002) Strong reciprocity, human cooperation and the enforcement of social norms. *Hum Nature* 13: 1-25.
21. Gintis H, Bowles S, Boyd R, Fehr E (2003) Explaining altruistic behavior in humans. *Evol Hum Behav* 24: 153-172.
22. Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425: 785-791.
23. Fowler J H (2005) Altruistic punishment and the origin of cooperation. *Proc Natl Acad Sci USA* 102: 7047-7049.
24. Mathew S, Boyd R (2011) Punishment sustains large-scale cooperation in prestate warfare. *Proc Natl Acad Sci USA* 108: 11375-11380.

25. Shinada M, Yamagishi T (2007) Punishing free riders: direct and indirect promotion of cooperation. *Evol Hum Behav* 28: 330-339.
26. Fehr E, Fischbacher U (2004) Third-party punishment and social norms. *Evol Hum Behav* 25: 63-87.
27. Carpenter J P, Matthews P H, Ong'ong'a O (2004) Why punish? Social reciprocity and the enforcement of prosocial norms. *J Evol Econ* 14: 407-429.
28. Charness G, Cobo-Reyes R, Jiménez N (2008) An investment game with third-party intervention. *J Econ Behav Organ* 68: 18-28.
29. Carpenter J P, Matthews P H (2012) Norm enforcement: Anger, indignation or reciprocity? *J Eur Econ Assoc* 10: 555-572.
30. Hauser MD (1992) Costs of deception: Cheaters are punished in rhesus monkeys (*Macaca mulatta*). *Proc Natl Acad Sci USA* 89: 12137-12139.
31. Clutton-Brock TH, Parker GA (2005) Punishment in animal societies. *Nature* 373: 209-216.
32. Ratnieks FLW, Wenseleers T (2005) Evolution. Policing insect societies. *Science* 307: 54-56.
33. Monnin T, Ratnieks FLW, Jones GR, Beard R (2002) Pretender punishment induced by chemincal signaling in a queenless ant. *Nature* 419: 61-65.
34. Ratnieks FLW, Wenseleers T (2008) Altruism in insect societies and beyond: Voluntary or enforced? *Trends Ecol Evol* 23: 45-52.
35. Raihani NJ, Grutter AS, Bshary R (2010) Punishers benefit from third-party punishment in fish. *Science* 327: 171.
36. Riedl K, Jensen K, Call J, Tomasello M (2012) No third-party punishment in chimpanzees. *Proc Natl Acad Sci USA* 109: 14824-14829.
37. Fan C-P (2000) Theaching children cooperation – An application of experimental game theory. *J Econ Behav Organ* 41: 191-209.
38. Keser C, van Winden F (2000) Conditional cooperation and voluntary contributions to public goods. *Scand J Econ* 102: 23-39.
39. Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Econ Lett* 71: 397-404.
40. Kurzban R, Houser D (2005) Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proc Natl Acad Sci USA* 102: 1803-1807.
41. Schmidt M F H, Tomasello M (2012) Young children enforce social norms. *Current Directions in Psychological Science* 21: 232-236.
42. Joyce WB (1999) On the free-riding problem in cooperative learning. *Journal of Education for Business* 74: 271-274.
43. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454: 1079-1084.
44. Parsons T (1959) The school class as a social system: some of its functions in American Society. *Harvard Educational Review* 29: 297-318.
45. Johnson DW (1981) Student-student interaction: The neglected variable in education. *Educational Researcher* 10: 5-10.
46. Tomasello M, Call J, Hare B (2003) Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Science* 7: 153-156.
47. Astington JW (1993) *The child's discovery of the mind* (Harvard University Press, Cambridge).
48. Harbaugh W, Krause K, Berry TR (2001) GARP for kids: On the development of rational choice behavior. *American Economic Review* 91: 1539-1947.
49. Call J, Tomasello M (2008) Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science* 12: 187-192.

50. Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114: 817-868.
51. Bolton G E, Ockenfels A (2000) ERC: A theory of equity, reciprocity, and competition. *Am Econ Rev* 90: 166-193.
52. Brandts J, Charness G (2011) The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics* 14: 375-398.
53. Camerer FC (2003) *Behavioral game theory: Experiments in strategic interaction* (Princeton University Press, Princeton).
54. Selman R L (1971) Taking another's perspective: Role-taking development in early childhood. *Child Dev* 42: 1721-1734.
55. Jensen K, Call J, Tomasello M (2007) Chimpanzees are rational maximizers in an ultimatum game. *Science* 318: 107-109.
56. Balafoutas L, Nikiforakis N (2012) Norm enforcement in the city: A natural field experiment. *Eur Econ Rev* 56: 1773-1785.
57. Gneezy, U, Meier, S, Rey-Biel, P (2011) When and why incentives (don't) work to modify behavior. *J Econ Perspectives* 25(4): 191-210.
58. Harbaugh W, Krause K (1999) Economic experiments that you can perform at home on your children. Working Paper, University of Oregon. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/56/UO-1999-1_Harbaugh_Economic_Experiments%5B2%5D.pdf?sequence=1
59. Castillo, M, Ferraro, P J, Jordan, J L, Petrie, R (2011) The today and tomorrow of kids: Time preferences and educational outcomes of children. *J Public Econ* 95: 1377-1385.
60. Bettinger E, Slonim R (2007) Patience among children. *J Public Econ* 91: 343-363.
61. Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. *Science* 319: 1362-1367.

Author Contributions P.L., S.A., D.G.-R. and M.S. contributed equally to all parts of the research.

Additional Information

The corresponding author hereby confirms that all experiments were performed in accordance with relevant guidelines and regulations.

Competing financial interests: The authors declare that they have no competing financial interests.

Figure 1: In the prisoner's dilemma, players can either cooperate (C) or defect (D). While mutual cooperation yields the socially optimal outcome of 4 tokens per player, each subject has an incentive to defect as a dominant strategy. Defection of both players is the Nash equilibrium of the game, yielding a payoff of 2 for each player. The first (second) number in each cell indicates player 1's (player 2's) payoff.

		Player 2	
		C	D
Player 1	C	4, 4	0, 6
	D	6, 0	2, 2

Figure 2: Average cooperation rates by age and treatment (N=554 in CTR and N=566 in TPP). Error bars, mean ± SEM.

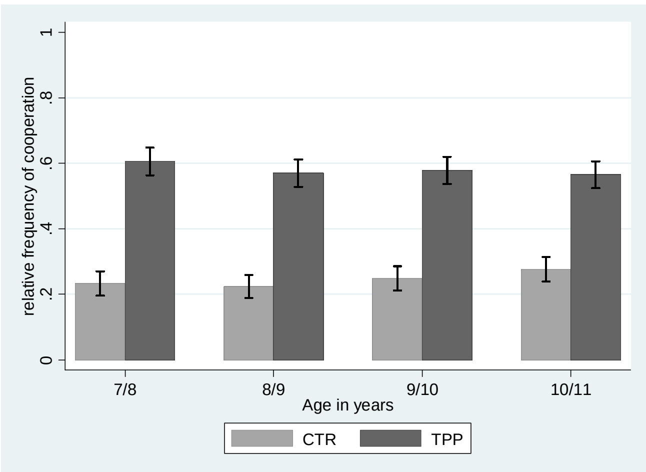


Figure 3: Average expectation of cooperative behavior of the partner by age and treatment (N=554 in CTR and N=566 in TPP). Error bars, mean ± SEM.

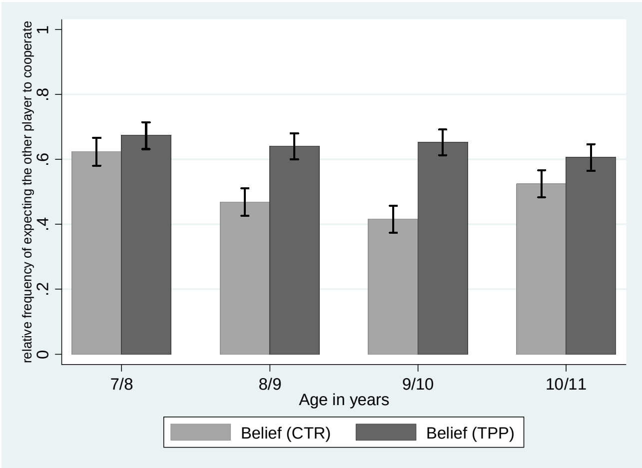
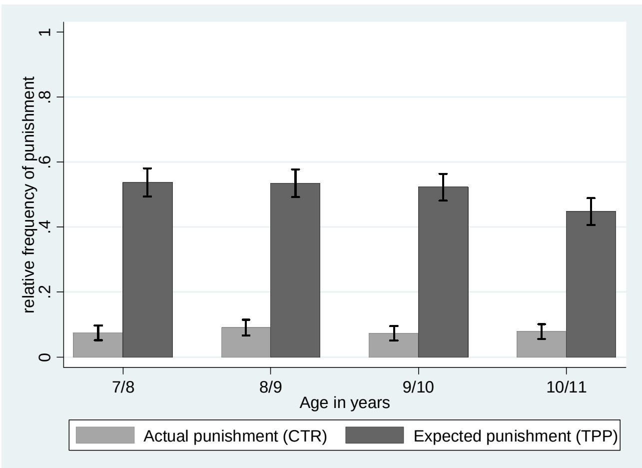


Figure 4: Average actual punishment rates of third parties (CTR; N=554) and expected rates of punishment by players in TPP (N=566) by age. Error bars, mean ± SEM.



Supplementary Information

Third party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation

by

Philipp Lergetporer, Silvia Angerer, Daniela Glätzle-Rützler and Matthias Sutter

Content

Supplementary tables and figures

Table S1: Number of participants by age and gender with the number of children not understanding the game in brackets.

	Gender	
Age (in years)	Female	Male
7/8 years	126 (5)	139 (9)
8/9 years	121 (1)	164 (2)
9/10 years	131 (2)	153 (1)
10/11 years	133 (0)	153 (1)
ALL (N=1,141)	**511 (8)**	**609 (13)**

In total, 21 children (10 in CTR and 11 in TPP) were excluded from the analysis because they did not understand the experiment.

Table S2: A comparison of expected payoffs in TPP, given beliefs about expected behavior and given actual behavior of the opponents.

Age (in years)	Observer's probability to punish (P)		Partner's probability to cooperate (C)		Expected payoff (given behavior) $		Expected payoff (given beliefs) $	
	Behavior	Beliefs	Behavior	Beliefs	Coop.	Defect	Coop.	Defect
7/8 yrs.	7.5%	53.8%	60.6%	67.4%	2.42	**4.09**	**2.70**	2.17
8/9 yrs.	9.1%	53.5%	57.0%	64.1%	2.28	**3.89**	**2.56**	2.12
9/10 yrs.	7.3%	52.4%	57.8%	65.3%	2.31	**4.00**	**2.61**	2.20
10/11 yrs.	7.8%	44.8%	56.6%	60.7%	2.26	**3.93**	2.43	**2.44**

$ Calculation of expected payoff

… for cooperation: C*4

… for defection: (1-P)*C*6 + (1-P)*(1-C)*2

Table S3: Clustered probit regression on cooperation

	Cooperation in CTR	Cooperation in TPP
Age in years	0.011	0.001
	(0.015)	(0.020)
Female (=1)	-0.036	0.077
	(0.037)	(0.052)
German school (=1)	0.035	0.055
	(0.035)	(0.050)
Belief partner (1=cooperation)[#]	0.320***	0.416***
	(0.035)	(0.065)
Relative IQ[$]	-0.061	0.015
	(0.103)	(0.127)
Number of siblings	0.001	0.007
	(0.020)	(0.023)
Altruism[&]	0.083***	0.065***
	(0.015)	(0.020)
Impatient (=1)[+]	-0.051	-0.027
	(0.043)	(0.058)
Belief observer (1=punishment)		0.379***
		(0.076)
Belief partner*Belief observer		0.114
		(0.106)
# Observations	489	488

Notes. ***, **, * denote significance at the 1%, 5%, 10% level, robust standard errors in parentheses. Clustered on class level.

[$] The IQ was measured with Raven's Coloured Progressive Matrices. Consistent with the mean values of the norming sample of the Raven's Coloured Progressive Matrices, the share of correctly solved matrices in our subject pool increases with age. Therefore, we measured the IQ relative to the respective age group in order to avoid confounding age- and IQ-effects.

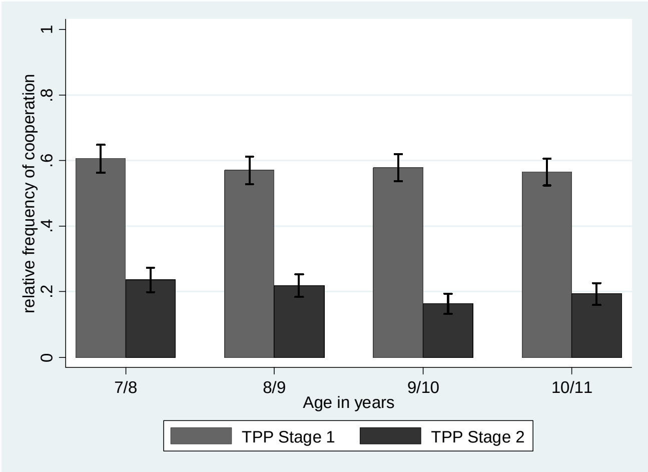
[&] Number of tokens donated to a charity (0-6) in an independent experiment. We include this variable as a proxy for social intelligence which is needed in social interactions (Kaukiainen et al. (1) find that social intelligence is significantly related to empathy).

[+] Patience was measured in an independent experiment with 3 binary choice problems. If a child decided in all three choice problems not to wait for the higher payoff, the child was classified as being impatient.

Reference:

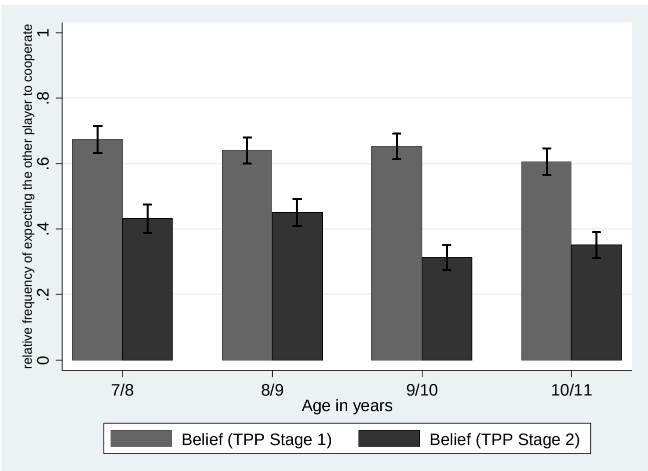
62. Kaukiainen A et al. (1999) The relationships between social intelligence, empathy, and three types of aggression. *Aggressive Behavior* 25: 81-89.

Figure S1: Average cooperation rates by age and stage in TPP. Stage 1 with third party punishment (reported in the paper); Stage 2 without third party punishment (not reported in the paper, N=566). Error bars, mean ± SEM.



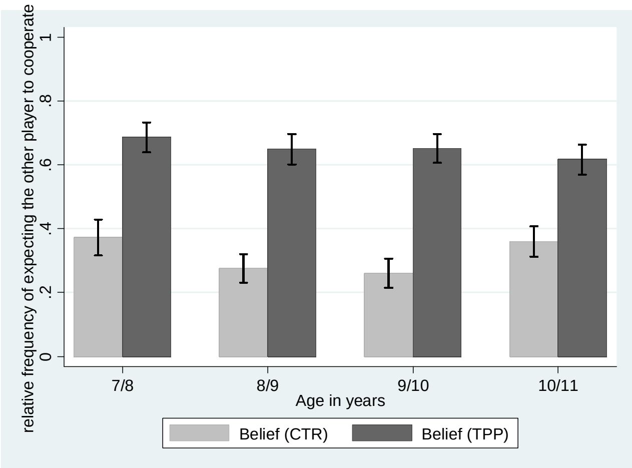
Overall, 20% cooperate in stage 2 of TPP without punishment, compared to 58% with punishment in stage 1 (P=0.000 overall and in each cohort; McNemar exact test). The cooperation rate in stage 2 is not significantly different from the rate in CTR (compare Fig. 2; P>0.05 overall and in each cohort, χ^2-tests).

Figure S2: Average expectation of cooperative behavior of the partner by age and stage in TPP. Stage 1 with third party punishment (reported in the paper); Stage 2 without third party punishment (not reported in paper, N=566). Error bars, mean ± SEM.



In the second stage of TPP (without punishment) children have significantly and much lower expectations about their partner's likelihood to cooperate (P < 0.01 in all age cohorts; Mc-Nemar test).

Figure S3: Average expectations of conditional cooperators towards the cooperative behavior of their partners by age in CTR (N=369) and TPP (N=418). Subjects are classified as conditional cooperators if their beliefs about the partner's choice are aligned with their own decision.



Experimental instructions (translated from German/Italian)

Note: Italic font is used for the instructions to the experimenter.

Assign child to treatment 1 or treatment 2 according to the randomization-list.

Today I prepared a game for you. In this game you can earn tokens. You can buy presents in our shop with these tokens. Bigger presents cost more than smaller ones. The presents are different from last time.

Treatment CTR

In this game you can earn these white tokens (*show tokens*). Could you please repeat the rules for the tokens? (*Child must repeat that it can select presents with the tokens and that bigger presents are more costly than smaller presents*).

There are two meeples: a yellow and a white meeple (*place meeples on table; the yellow meeple must be placed directly in front of the child*). You are the yellow meeple (*point at yellow meeple*). The white meeple is a child which we will select randomly. This child attends the x. grade of a German (Italian) school here in Meran just like you, but it attends a different school (*adapt explanation to grade and school-language; place white meeple on school-card*). This child can be a boy or a girl, but you don't know whom exactly you are playing with. This is a secret. Your partner doesn't know either who you are. Could you please repeat this part? (*Child must repeat all the information of this paragraph. If it misses some parts, ask explicitly*)

Stage 1: Prisoner's Dilemma without punishment

The game works like this:

(place decision-sheet "COOP" in front of the child). At the beginning you and your partner receive 2 tokens each (*place tokens in front of yellow and white meeple*). Both of you must decide whether to send ZERO or BOTH tokens to the partner. It is important to note that I have tokens as well (*show tokens*). If you send your tokens to your partner, I will add 2 more tokens for your partner (*show physically: shove the subject's tokens towards its partner and double the tokens upon arrival; restore original distribution after illustration*). If your partner decides to send you his/her tokens, I will double them as well (*show physically*). You must decide between sending ZERO and sending BOTH tokens on this sheet (*point at the respective box on the decision sheet*). Your partner has the same options as you: he/she can either send you ZERO or BOTH tokens. Let's go through some examples now (*The child must handle tokens in the examples*).

I) What happens if you tick this box here (*point at box with ZERO tokens*) and your partner sends you ZERO tokens? (*Answer: no tokens will be sent and the experimenter does not add any tokens*) And how many tokens will you earn in this case? (*Answer: 2*) And how many tokens will your partner earn? (*Answer: 2*)

II) What happens if you tick this box here (*point at box with BOTH tokens*) and your partner sends you BOTH of his tokens? (*Answer: the child and its partner send both tokens to one another and the experimenter doubles each transaction*) And how many tokens will you earn in this case? (*Answer: 4*) And how many tokens will your partner earn? (*Answer: 4*)

III) What happens if you tick this box here (*point at box with ZERO tokens*) and your partner sends you BOTH of his tokens? (*Answer: the child does not send tokens to partner, but the partner sends his/her tokens to the child. The experimenter doubles the transaction from the partner to the child*) And how many tokens will you earn in this case? (*Answer: 6*) And how many tokens will your partner earn? (*Answer: 0*)

IV) What happens if you tick this box here (*point at box with BOTH tokens*) and your partner sends you ZERO of his tokens? (*Answer: the child sends both tokens to partner, but the partner sends no tokens to the child. The experimenter doubles the transaction from the child to the partner*) And how many tokens will you earn in this case? (*Answer: 0*) And how many tokens will your partner earn? (*Answer: 6*)

Do you already know how many tokens your partner sends you? (*Answer: No.)* Exactly. Likewise, your partner does not know how many tokens you sent him when he decides on how many tokens to send you. Could you please repeat how the game works? (*Child must exhaustively repeat the rules of the game. If it misses some parts, ask explicitly*)

We don’t know yet the exact number of tokens you will earn in this game. You receive the tokens which you keep and those which your partner sends you. Since we don’t know yet how many tokens your partner will send you, you will receive the tokens from this part not today but only when we visit you next time. It is very important that your decision in this game is secret: the other children will never know how many tokens you sent.

Please take your decision now. Take as much time as you need. In the meantime I will turn around so that I don’t disturb you. Just call me when you are done. (*Hand over the pen to the child so that it can decide. Turn around and wait until the child signals that it has finished.*)

Thank you for your decision. You ticked this box (*point at the respective box*). Could you please explain what that means (*Child must explain what the [possible] consequences of its decision are*)?

I still have another question for you. What do you think decides your partner? If your guess is correct (meaning that your partner really does what you think), you will receive one extra token. If your guess is not correct, you don't earn an additional token. What do you think, will your partner send you ZERO or BOTH of his/her tokens (*Register answer of the child in computer*)? Thank you for your decision. This game is over now. (*Put away the material of this stage*)

Stage 2: Punishment Decision

(*Place two white meeples in front of the child and put orange tokens on the table*) Here I prepared a second game for you. In this game you can observe the play of two other children in a game which works just like the one we played before. These children attend the x. grade of a German (Italian) school here in Meran just like you, but they attend a different school (*adapt explanation to grade and school-language; place white meeples on school-cards*). We selected these children randomly and both are new children, i.e. none of them is your partner from before. You are the observer of one of these children (*point to the left meeple from the subject's point of view*). You do not know yet how this child decides in this game, but there are two possibilities: This sheet (*place explanation sheet 1 in front of the child*) shows the first option the child has. Can you tell me how many tokens this child would send to its partner (*Answer: 0*). On this sheet (*place explanation sheet 2 in front of the child and to the right of explanation sheet 1*) you see the second option the child has. Can you tell me how many tokens this child would send to its partner (*Answer: 2*). As an observer you have the option to deduct tokens from the child if it does not send the tokens to its partner (*point at explanation sheet 1*). Deducting tokens from the child works like this (*place orange-token-card (left) and deduction card (right) in front of the child*): You can either pick the card with the orange token (*point at orange-token-card)* or the deduction-card (*point at the deduction card*). If you pick the card with the orange token (*point at orange-token-card*), this token is yours and with this token you can select a present which costs 1 token when we visit you next time. In this case, the child you observe can keep all its tokens. If you pick the deduction-card instead (*point at deduction-card*), you don't receive the orange token but we will deduct all the tokens from the child which you observe. In this case, these tokens are lost and nobody gets them. This means that the child receives zero tokens in the game. If the child chooses to send the tokens to its partner (*point at explanation sheet 2*) you will receive the orange token and we will not deduct the tokens from the child no matter which card you picked. This means that you can only deduct points from a child who does not send its tokens.

It is important that the child knows at the time of deciding that you might deduct its tokens if it does not send them. The partner of the child you observe (*point at right meeple*) has an observer on its own and this observer can deduct the tokens of the partner if he/she does not send the tokens. In the game you played before nobody had the possibility to deduct your tokens. Could you please repeat how the game works? (*Child must exhaustively repeat the rules of the game. If it misses some parts, ask explicitly*). It is very important that your decision in this game is secret: the other kids will never know which card you picked. The other children really exist and you can really deduct the tokens of a child which sends nothing by picking the deduction-card.

Now you may choose between the two cards. Please leave the card which you want to pick on the table as it is so that I can see its picture and turn the other card upside down. If you pick the card with the orange token, just flip the deduction card (*demonstrate*). In this case you will receive the token in any case and we don't deduct the tokens from the child in any case. In contrast, if you pick the deduction card, flip the card with the orange token (*demonstrate*). In that case you will not receive the orange token and we will deduct the child's tokens if it doesn't send it. If it sends its tokens, you will receive the orange token and we don't deduct any token from that child. Please take your decision now. Take as much time as you need. In the meantime I will turn around so that I don't disturb you. Just call me when you are done. (*Hand over the pen to the kid so that it can decide. Turn around and wait until the child signals that it has finished. Then register answer of the child in the computer*.)

Thank you for your decision. You picked this card (*point at the respective card*). Could you please explain what that means (*Child must explain what the [possible] consequences of its decision are*).

Treatment TPP

In this game you can earn these white tokens (*show tokens*). Could you please repeat the rules for the tokens? (*Child must repeat that it can select presents with the tokens and that bigger presents are more costly than smaller presents*).

There are two meeples: a yellow and a white meeple (*place meeples on table; the yellow meeple must be placed directly in front of the child*). You are the yellow meeple (*point at yellow meeple*). The white meeple is a child which we will select randomly. This child attends the x. grade of a German (Italian) school here in Meran just like you, but it attends a different school (*adapt explanation to grade and school-language; place white meeple on school-card*). This child can be a boy or a girl, but you don't know whom exactly you are playing with. This is a secret. Your partner doesn't know either who you are. Could you please repeat this part? (*Child must repeat all the information of this paragraph. If it misses some parts, ask explicitly*)

Stage 1: Prisoner's Dilemma with punishment

The game works like this:

(place decision-sheet "COOP_PUN" in front of the child). At the beginning you and your partner receive 2 tokens each (*place tokens in front of yellow and white meeple*). Both of you must decide whether to send ZERO or BOTH tokens to the partner. It is important to note that I have tokens as well (*show tokens*). If you send your tokens to your partner, I will add 2 more tokens for your partner (*show physically: shove the subject's tokens towards its partner and double the tokens upon arrival; restore original distribution after illustration*). If your partner decides to send you his/her tokens, I will double them as well (*show physically*). You must decide between sending ZERO and sending BOTH tokens on this sheet (*point at the respective box on the decision sheet*). Your partner has the same options as you: he/she can either send you ZERO or BOTH tokens. Let's go through some examples now (*The child must handle tokens in the examples*).

I) What happens if you tick this box here (*point at box with ZERO tokens*) and your partner sends you ZERO tokens? (*Answer: no tokens will be sent and the experimenter does not add any tokens*) And how many tokens will you earn in this case? (*Answer: 2*) And how many tokens will your partner earn? (*Answer: 2*)

II) What happens if you tick this box here (*point at box with BOTH tokens*) and your partner sends you BOTH of his tokens? (*Answer: the child and its partner send both tokens to one another and the experimenter doubles each transaction*) And how many tokens will you earn in this case? (*Answer: 4*) And how many tokens will your partner earn? (*Answer: 4*)

III) What happens if you tick this box here (*point at box with ZERO tokens*) and your partner sends you BOTH of his tokens? (*Answer: the child does not send tokens to partner, but the partner sends his/her tokens to the child. The experimenter doubles the transaction from the partner to the child*) And how many tokens will you earn in this case? (*Answer: 6*) And how many tokens will your partner earn? (*Answer: 0*)

IV) What happens if you tick this box here (*point at box with BOTH tokens*) and your partner sends you ZERO of his tokens? (*Answer: the child sends both tokens to partner, but the partner sends no tokens to the child. The experimenter doubles the transaction from the child to the partner*) And how many tokens will you earn in this case? (*Answer: 0*) And how many tokens will your partner earn? (*Answer: 6*)

Do you already know how many tokens your partner sends you? (*Answer: No.)* Exactly. Likewise, your partner does not know how many tokens you sent him when he decides on how many tokens to send you. Could you please repeat how the game works? (*Child must exhaustively repeat the rules of the game. If it misses some parts, ask explicitly*)

I still have to explain you something very important: There is yet another meeple (*place black meeple next to the yellow meeple*) and this meeple is your observer in the game. This child attends the x. grade of a German (Italian) school here in Meran just like you, but it attends a different school (*adapt explanation to grade and school-language; place white meeples on school-cards*). We selected this child randomly and this is a new child, i.e. it is not your partner but another child. Your observer can see what you do in this game and he/she can deduct your tokens. This works as follows:

If you tick this box here (*point at box with ZERO tokens*) your observer can deduct all your tokens. The deduction works as follows: Your observer must choose between the card with the orange token (*show orange-token card*) and the deduction card (*show deduction-card*). If he/she goes for the orange token (*shove orange-token-card in front of the black meeple*) then he/she can use this token to buy a present in our shop for him/herself and you keep all your tokens. If however, he/she picks the deduction-card (*shove deduction-card in front of the black meeple*), he/she does not receive the orange token and we will deduct ALL your tokens. In this case, your tokens will be lost and nobody gets them. That is, you won't earn any tokens in this game. However, your observer can only deduct your tokens if you don't send any to your partner. If you tick this box here (which means that you send both tokens to your partner), your observer can't deduct tokens from you and he/she receives the orange token for him/herself. In this game your partner (*point at the white meeple*) has a separate observer which chooses between the orange-token- and the deduction-card. Could you please repeat how the game works? (*Child must exhaustively repeat the rules of the game. If it misses some parts, ask explicitly*). What happens to the deducted tokens? (*Answer: they are lost and nobody gets them*).

We don't know yet the exact number of tokens you will earn in this game. You receive the tokens which you keep and those which your partner sends you. If you don't send your tokens, your observer might deduct all your tokens. Since we don't know yet how the other children decide you will receive the tokens from this part not today but only when we visit you next time. It is very important that your decision in this game is secret: the other children will never know how many tokens you sent.

Please take your decision now. Take as much time as you need. In the meantime I will turn around so that I don't disturb you. Just call me when you are done. (*Hand over the pen to the child so that it can decide. Turn around and wait until the child signals that it has finished.*)

Thank you for your decision. You ticked this box (*point at the respective box*). Could you please explain what that means (*Child must explain what the [possible] consequences of its decision are*)?

I still have two more questions for you. What do you think will your partner do? If your guess is correct (meaning that the other child really does what you think), you will receive one extra token. If your guess is not correct, you don't earn an additional token. What do you think your partner will do? Remember that your partner has an observer who can deduct his/her tokens if he/she doesn't send you the tokens. Do you think that your partner will send you ZERO or BOTH of his/her tokens? (*Register answer of the child in computer*). Thank you for your decision.

Case 1: Subject did not send its tokens: And what do you think your observer will do? If your guess is correct (meaning that your observer really does what you think), you will get one extra token. If your guess is not correct, you don't earn an additional token. What do you think your observer will do? (*Register answer of the kid in computer*)

Case 2: Subject sent its tokens: And what do you think your observer would do if you didn't send your tokens? If your guess is correct (meaning that your observer really does what you think), you will get one extra token. If your guess is not correct, you don't earn an additional token. What do you think your observer would do? (*Register answer of the kid in computer*)

Thank you for your decision. This game is over now. (*Put away the decision sheet, the black meeple, the cards of the observer and the white tokens*)

Stage 1: Prisoner's Dilemma without punishment

Here I prepared a second game for you. You can earn these yellow tokens in this game (*show tokens*). It is important that you can't sum tokens of different colors. Apart from that, the same rules for the tokens apply.

There are two meeples: a yellow and a white meeple. You are the yellow meeple (*point at yellow meeple*). The white meeple is again a child which we select randomly. This child attends the x. grade of a German (Italian) school here in Meran just like you, but it attends a different school (*adapt explanation to grade and school-language; place white meeple on school-card*). This child can be a boy or a girl, but you don't know whom exactly you are playing with. This is a secret. Your partner doesn't know either who you are. It is important that this partner is a new child and not your partner or observer from the game we played before. Could you please repeat this part? (*Child must repeat all the information of this paragraph. If it misses some parts, ask explicitly*). This game works exactly the same as the one we played before. The only difference is that you don't have an observer now and thus, nobody can deduct your points if you decide not to send them. Your partner (*point at the white meeple)* has no observer either. Could you please repeat how the game works? (*Child must exhaustively repeat the rules of the game and mention the differences to the first stage. If it misses some parts, ask explicitly*)

We don't know yet the exact number of tokens you will earn in this game. You receive the tokens which you keep and those which your partner sends you. Since we don't know yet how many tokens your partner sends you, you will receive the tokens from this part not today but only when we visit you next time. It is very important that your decision in this game is secret: the other children will never know how many tokens you sent.

Please take your decision now. Take as much time as you need. In the meantime I will turn around so that I don't disturb you. Just call me when you are done. (*Hand over the pen to the child so that it can decide. Turn around and wait until the child signals that it has finished.*)

Thank you for your decision. You ticked this box (*point at the respective box*). Could you please explain what that means (*Child must explain what the [possible] consequences of its decision are*)?

I still have another question for you. What do you think will your partner do? If your guess is correct (meaning that your partner really does what you think), you will receive one extra token. If your guess is not correct, you don't earn an additional token. What do you think, will your partner send you ZERO or BOTH of his/her tokens (*Register answer of the child in computer*)? Thank you for your decision. This game is over now. (*Put away the material of this stage*)

Sample Decision Sheet