

DISCUSSION PAPER SERIES

IZA DP No. 17519

**Coordination Games Played by Children
and Teenagers:
On the Influence of Age, Group Size and
Incentives**

Daniela Glätzle-Rützler
Matthias Sutter
Claudia Zoller

DECEMBER 2024

DISCUSSION PAPER SERIES

IZA DP No. 17519

Coordination Games Played by Children and Teenagers: On the Influence of Age, Group Size and Incentives

Daniela Glätzle-Rützler

University of Innsbruck

Claudia Zoller

Management Center Innsbruck

Matthias Sutter

*University of Innsbruck, Max Planck
Institute for Research on Collective Goods,
University of Cologne and IZA*

DECEMBER 2024

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Coordination Games Played by Children and Teenagers: On the Influence of Age, Group Size and Incentives*

Efficient coordination is a major source of efficiency gains. We study in an experimental coordination game with 718 children and teenagers, aged 9 to 18 years, the strategies played in pre-adulthood. We find no robust age effects in the aggregate, but see that smaller group sizes and larger incentives increase the likelihood of choosing the efficient strategy. Beliefs play an important role as well, as subjects are more likely to play the efficient strategy when they expect others to do so as well. Our results are robust to controlling for individual risk-, time-, and social preferences.

JEL Classification: C91

Keywords: coordination game, age, group size, incentives, children, experiment

Corresponding author:

Matthias Sutter
Max Planck Institute for Research on Collective Goods
Kurt-Schumacher-Str. 10
53113 Bonn
Germany
E-mail: matthias.sutter@coll.mpg.de

* We thank three referees, Ragan Petrie as editor, and Stefania Bortolotti, Alexander Cappelen, Werner Güth, Bertil Tungodden, and workshop participants in Munich and Bergen for helpful comments and suggestions. We are particularly grateful to Thomas Plankensteiner from the State Board of Education in Tyrol (Landesschulrat für Tirol) and the headmasters of the involved schools (Peter Arnold, Max Gnigler, Sigmund Heel, Gottfried Heiss, Ulrike Künstele, Hermann Lergetporer, Bernhard Schretter, and Peter Paul Steinringer) for making this study possible. Financial support was provided by the Austrian Central Bank (Jubilaumsfondsprojekte 12588 and 14680) and by the University of Cologne (through the Hans Kelsen Prize).

1. Introduction

An efficient organization of groups requires that group members are able to coordinate their actions successfully in order to achieve potentially large efficiency-gains from interaction. Examples abound across many different fields (Ochs, 1995; Camerer, 2003; Devetag and Ortmann, 2007; Cooper and Weber, 2020). Think of hunting together or defending one's territory against potential invaders, or companies coordinating on a new technological standard when competing platforms are available. In such situations, typically multiple equilibria exist, such as several platforms that can serve as technical standards or all group members hunting a large prey, like a stag, rather than each hunting individually a small animal like a rabbit.

The ability to coordinate actions has important welfare implications. In many cases, the multiple Nash equilibria in coordination games can be ranked according to their overall efficiency (Cooper et al., 1990; van Huyck et al., 1990; Duffy and Feltovich, 2006; Cason et al., 2012a, 2012b; Brandts et al., 2016). Coordination failure creates inefficiencies, which can take on two forms: either players do not coordinate on the same action, but rather pick different actions and therefore mis-coordinate, or they coordinate on an equilibrium that is not the most efficient one (Brandts and Cooper, 2006; Weber, 2006).

Given the importance of the ability to coordinate actions when subjects have to interact with each other and multiple equilibria exist, it is interesting to note that there is hardly any knowledge about when humans learn to coordinate efficiently and how the choice of efficient strategies develops in adolescence (Cooper and Weber, 2020). Except for very few papers discussed below it remains an open question whether children and teenagers are able to play efficient strategies and whether this ability develops with age and depends in predictable ways on economic incentives and the number of interaction partners in the group.

In this paper, we study how 718 children and teenagers, aged 9 to 18 years, played a simple one-shot coordination game – a stag-hunt game (Cooper et al., 1990) – in which there were only two strategies available that could lead to two different equilibria (in pure strategies); one equilibrium that is Pareto-efficient – which we will call the efficient equilibrium henceforth – and another equilibrium that is Pareto-dominated – we call this second equilibrium the inefficient one. Our primary interest lies in understanding the influence of age, incentives, and group size on the choice of strategies in this simple, one-shot coordination game. We can examine whether already children and teenagers react systematically to changes in the latter parameters of a coordination game and whether age moderates these potential effects. In addition to studying the influence of external parameters of the game, we are able to control for subjects' beliefs about other subjects' behavior.

According to Cooper and Weber (2020), there is hardly any work on how children play coordination games. Developmental psychologists have investigated how young children play such games. Grueneisen et al. (2015a) have studied whether theory of mind is important for efficient coordination in a type of stag-hunt coordination game, played by 6-year-old children. Theory of mind is developed around the age of 6, and it allows subjects to put themselves into the shoes of another person. This ability is supposedly influential in coordination games because the strategic uncertainty about the interaction partner's behavior is important for one's own choices. Grueneisen et al. (2015a) find that children with higher theory of mind are better able to coordinate on more efficient outcomes. Their study does not consider a broader age range, however. This is done in Grueneisen et al. (2015b) where the authors let children aged 3 to 8 years play coordination games. In these games, there is a focal point that can be used to coordinate the players' actions. Yet, all possible equilibria of the game are equally efficient, i.e. there is no ranking of equilibria possible. Grueneisen et al. (2015b) report that 5-year-old children and 8-year-old children are able to coordinate on the focal point with a likelihood that is significantly greater than chance, but 3-year-old children are not yet able to do so. Our paper differs from Grueneisen et al. (2015b) in several aspects. First, our coordination games have two Pareto-ranked equilibria (in pure strategies) which allow us to study whether children are able to choose the more efficient strategy. Second, we do not implement any focal point, thus making the choice of the efficient strategy considerably harder. Third, contrary to both Grueneisen et al. (2015a) and Grueneisen et al. (2015b) we do not only consider pairs of subjects playing a coordination game, but we also study behavior in larger groups, because many coordination games are played by multiple players. Our treatment variation with respect to group size (pairs of subjects vs. groups of five) allows investigating whether larger groups affect the likelihood of choosing the efficient strategy. Fourth, we consider a broader age range than Grueneisen et al. (2015a, 2015b) by considering children and teenagers aged 9 to 18 years. Brocas and Carrillo (2021a) study a similar age range in their coordination experiments. They run both a stag-hunt game (like we do) and a battle-of-the-sexes game. Children and adolescents of age 7 to 16 play these games for 96 periods (four supergames with 24 periods each). They find that older participants are more likely to reach the efficient and fair long-run outcome over the course of the many repetitions. In contrast to them, we study one-shot games that reveal subjects' reaction to strategic uncertainty (by varying group sizes) and different incentive schemes (by varying payoff matrices), absent of any convergence due to potential learning effects. We consider the paper by Brocas and Carrillo (2021a) and ours as complementary.

To start with, our focus on the influence of age on behavior in coordination games was exploratory, since at the time of running our experiment (in 2009) no evidence existed about how children and adolescents might play coordination games. Of course, from the (now existing) experimental economics literature on the behavior of children in economic games, one could form some expectations. Reasoning in economic games seems to develop below the age of 10 and is rather stable in the teenage years, as Brosig-Koch et al. (2015) have shown in an experiment on a backward-induction task (that has a unique equilibrium) and Brocas and Carrillo (2021b) in a dominance solvable game. Similarly, Czermak et al. (2016) have found in a series of normal form games (also with unique equilibria) that strategic thinking is fairly stable in the teenage years. A priori, it is unclear from these papers what the implications for coordination games would be. In coordination games, strategic uncertainty is prevalent because a player needs to form an expectation about the other players' actions in order to best respond. This does not necessarily imply that more sophisticated strategic thinking would lead to more frequent choices of the efficient strategy. The direction of influence will depend on expectations, and they might depend on age, so it's ultimately an empirical question whether age matters. In repeated coordination games, Brocas and Carrillo (2021a) have shown that older adolescents coordinate more efficiently, yet for one-shot games this is yet to be examined.

Besides investigating potential age effects, we also examine the influence of incentives and of group size. With respect to incentives, the prediction is straightforward (see Schmidt et al., 2003, or Brandts and Cooper, 2006, for coordination games with adults). The choice of the efficient strategy in our stag-hunt game should become more likely when it becomes more profitable – by increasing the payoffs in the efficient equilibrium – or when it is less costly – by increasing the payoffs in case of failure to coordinate on the efficient equilibrium. With respect to the effects of group size, experimental coordination games show typically more efficient play in smaller groups than in larger groups, and such patterns have been found consistently with adult participants (see Weber, 2006, and in particular his table 2 for a summary of group size effects, but also the surveys of Devetag and Ortmann, 2007, or Cooper and Weber, 2020). We vary group sizes by letting subjects play once in pairs and once in groups of five, expecting more frequent choices of the efficient strategy in smaller groups.

Our experiment has the following main results: (1) We don't find systematic and robust effects of age in our age range of 9-18 year-olds. This holds for the choice of the efficient strategy and the stated beliefs. Yet, with respect to the likelihood of successful coordination, we see some slight increase with age, mainly driven by an increase of inefficient coordination. (2) As expected, incentives play a systematic role. By and large, the efficient strategy is chosen

significantly more often as it gets more rewarding in 2-person groups. In 5-person groups we find no significant effect of the incentive scheme. (3) Efficient strategies are more often chosen in groups of two subjects than in groups with five subjects. (4) When subjects expect others in the group to play the efficient strategy, they are more likely to play it themselves, meaning that beliefs and actions are related to each other.

The rest of the paper is organized as follows. In Section 2 we present the experimental design and our experimental treatments. In Section 3 we present the results, and Section 4 concludes the paper.

2. Experimental design

2.1 Procedures

We conducted our study in seven different schools in Tyrol (Austria) in 2009, including 28 different classes that comprised fourth grade (with 9-10 years old children), sixth grade (11-12 years), eighth grade (13-14 years), tenth grade (15-16 years), and twelfth grade (17-18 years).¹ In total, we had 819 participants, with a slight surplus of girls (see Table 1), as one of the schools was a girls-only school.² Our experiment was part of a larger research project that investigated over the course of two academic years the development of economic decision-making of children. In the first year, we measured children's risk, time, and social preferences (see Fehr et al., 2013, Sutter et al., 2013) and we use these data as explanatory variables for the coordination decision that we elicited in the second year of the research project. In the notes to Table 2 we describe in detail how we measured those preferences. Risk preferences might matter in coordination games, because such games entail strategic risk. Social preferences might matter for the choice of strategy because miscoordination yields different degrees of unequal payoffs across our treatments. Time preferences have been shown to be related to cooperative behavior (Kölle and Wenner, 2023), and similarly they might be related to coordination behavior.

Our study was approved by the state board of education in Tyrol. Further consent was given by the headmasters of each school. Prior to the experiment, parents received an information leaflet explaining the nature of the study, without revealing details of the

¹ In the Austrian school system, children attend 4 years of elementary school (grade 1-4) and thereafter choose to either attend a Gymnasium (which qualifies for an academic track; grades 5-12) or a Middle school (grade 5-8). Roughly the top 25% of children pass the entry qualification to attend a Gymnasium. Thus, one needs to be cautious when evaluating the representativeness of the results of our study. We expect the socio-economic status (SES) background of the subjects in our sample to be somewhat biased to higher levels of SES compared to the general population.

² The fraction of female students varies between 41% and 64% among the schools except for two schools. One of the elementary schools has a fraction of females of only 21% and one of the Gymnasiums is a female-only school.

experiment, though. Participation was, of course, voluntary for children, but all of them agreed to participate. The experiment was run during regular school hours. Within a given school we randomly selected classes of each age cohort to participate in our research project. Given that all children of the randomly selected classes in each school participated, there was no self-selection into experimental participation.

Table 1: Distribution of observations by treatment, gender, and understanding

School grade (age of students)	Number of subjects who passed the control questions per treatment [#]			Total (female)	Failed control questions (in % of total)
	“9/1”	“9/5”	“13/1”		
4 (9/10 years)	37	32	32	101 (52)	21 (17%)
6 (11/12 years)	76	44	51	171 (95)	35 (17%)
8 (13/14 years)	74	51	74	199 (112)	24 (11%)
10 (15/16 years)	54	34	40	128 (80)	14 (10%)
12 (17/18 years)	47	28	44	119 (64)	7 (6%)
Total (female)	288 (173)	189 (86)	241 (144)	718 (403)	101 (12%)

Notes: Experimental participants earned points in the experiment that were exchanged into Euro at the following age-dependent rates: €0.3 per point for 9-10-year-olds, €0.5 for 11-12-year-olds, €0.7 for 13-14-year-olds, €1.4 for 15-16-year-olds and €2.2 for 17-18-year-olds.

[#]Treatments are explained in Section 2.2.

The experiment was computerized (using zTree, Fischbacher, 2007) and run with a portable lab.³ Instructions were explained verbally by taking great care that children and teenagers could understand the rules of the coordination game and in particular how to interpret the payoff matrix. Before proceeding to the experiment, all participants had to answer two control questions (see Appendix for control questions). Overall 718 subjects answered both control questions correctly with a slight increase in comprehension by age (from 83% for 9-10 year olds to 94% for 17-18 year olds, $p=0.009$ χ^2 -test). Note that we base our analysis on those 718 children and teenagers who passed both control questions. We used partition walls to prevent participants from looking at other participants’ screens. Moreover, experimenters enforced rules that questions could only be answered privately and that no communication between participants was allowed.

³ In each of the schools we turned one classroom into a lab using our own partition walls and laptops.

Table 2: Descriptive statistics

Variables	Mean	SD	N
Choosing strategy HIGH (=1)	.212	.409	718
Expecting strategy HIGH (=1)	.329	.470	718
Treatment 9/1 (=1)	.401	.490	718
Treatment 9/5 (=1)	.263	.441	718
Treatment 13/1 (=1)	.336	.4724	718
Group size 5 (=1)	.500	.500	718
Order (=1) [§]	.503	.500	718
School 1 (=1)	.021	.143	718
School 2 (=1)	.306	.461	718
School 3 (=1)	.220	.414	718
School 4 (=1)	.272	.445	718
School 5 (=1)	.068	.252	718
School 6 (=1)	.061	.240	718
School 7 (=1)	.052	.221	718
Grade 4 (=1)	.141	.348	718
Grade 6 (=1)	.238	.426	718
Grade 8 (=1)	.277	.448	718
Grade 10 (=1)	.178	.383	718
Grade 12 (=1)	.166	.372	718
Female (=1)	.561	.496	718
Risk aversion (r) [#]	.556	.245	670
Impatience [#]	80.561	45.082	629
Spiteful (=1) [#]	.214	.410	533
Egalitarian (=1) [#]	.351	.477	533
Generous (=1) [#]	.396	.489	533
Math grade [§]	2.511	1.111	718

Notes:

[§] Order is defined as 1 when subjects first played in the 2-person group, and afterwards in the 5-person group. The reverse order is coded as 0.

[#]Risk aversion, impatience, and other regarding preferences (i.e. spiteful, egalitarian, generous) of subjects were measured in different experiments (one academic year earlier). Risk aversion was elicited utilizing the framework of Ellsberg's two-color choice task (Ellsberg 1961) (see Figure A1). Subjects could choose either a lottery that paid €10 or zero with equal probability, or a safe amount of money that increased in increments of €0.50 from €0.50 to €10. As a measure of a subject's risk aversion r we use $r = 1 - CE/10$, where CE denotes the certainty equivalent of the risky prospect. The variable risk aversion can range from 0 (very risk-loving) to 1 (very risk-averse), with 0.5 as risk-neutral. For roughly one quarter of the sample we examined the effects of stake sizes by increasing the lottery prize from €4 for 10-year-olds in steps of €2 up to €12 for 18-year-olds. We did not find any stake size effect and hence control for this stake size variation by normalizing the CE to €10. A subject's impatience was measured using time preferences decision sheets, giving students the option of choosing a fixed amount at an earlier point in time or waiting for an increased payoff at a later point in time (see Figure A2). Each subject faced 8 different decision sheets with 20 decisions each. Only one of the choices was paid out in the end. The decision sheets varied in a 2x2x2 factorial design the size of the fixed payoff at the earlier point in time (either €10.10 or €4.05), the waiting time between the earlier payoff and the later payoff date (3 weeks or 1 year) and the up-front delay (no up-front delay or 3 weeks up-front delay). The variable "Impatience" measures the frequency with which a subject chose the earlier (and smaller) payoff. Larger values indicate higher impatience. See Sutter et al. (2013) for more details on the risk and the time preference experiment. Other-regarding preferences were measured with three simple allocation choices which allow classifying the subjects' preference type to either be egalitarian, altruistic, or spiteful. See Fehr et al. (2013) for more details on the social preference elicitation experiment.

[§]Math grade are self-reported grades in mathematics ranging from 1 (best grade) to 5 (worst grade). Columns 1 to 3: means, standard deviations (SD), and number of subjects who passed our comprehension check (i.e.

answered the two control questions correctly). Note, that we have two observations on choosing strategy HIGH as well as expecting strategy HIGH per child, one for group size 2 and one for group size 5. Note that treatments are explained in Section 2.2.

The experiment was incentivized, and each point earned in the experiment was exchanged into Euros at a pre-specified exchange rate which was announced prior to the experiment. This exchange rate differed by school grades, and was proportional to the average weekly allowance across the whole age range (see Fehr et al., 2013 for average weekly allowance for this age range). In the legend to Table 1 we present the exchange rate for each age group. Overall, the experiment took about 25 minutes.

2.2 The effects of incentives – Between-subjects treatments

The experimental design had both a between-subjects and a within-subjects variation. The between-subjects variation concerned the payoff matrix of the coordination game. Here we had three different one-shot treatments which are shown in Table 3.⁴ In each treatment, subjects could choose between action LOW (presented as action “One” in the instructions, see experimental instructions in the Appendix) and action HIGH (presented as action “Two” in the instructions). The combination of one’s own action and the minimum action chosen by the other players in one’s group determined the payoffs. This general structure yielded two pure strategy Nash equilibria of the game, one where all players in a group play HIGH – which is the Pareto-efficient equilibrium – and one in which all players play LOW – which yields the inefficient equilibrium.⁵ The choice of LOW always yielded 8 points for the active player.

The three treatments shown in Table 3 differed with respect to the possible payoffs from choosing strategy HIGH. The treatment labels indicate the two possibilities in case of picking HIGH. In treatment “9/1”, presented on the left-hand side of Table 3, a player who plays HIGH earned 9 points if the minimum choice in the group was HIGH, but earned only 1 point if someone in the group chose LOW. Of course, the efficient equilibrium (of all players earning 9 points) requires all group members to choose HIGH, but this is risky, because if only one group member chooses LOW, then the payoff drops to 1 point for all players who chose HIGH.

⁴ The treatments were assigned to classes (note that we did not randomize treatments within a given class) in order to guarantee a more or less balanced sample. Due to two last-minute cancellations of two classes (one in grade 6 and one in grade 12), we have a lower number of subjects in some treatments (i.e. especially in grade 12 in treatment 9/5, see Table 1). In cases where we had to allocate treatments unevenly (i.e. because we had one or two additional class(es)), we assigned first treatment 9/1 followed by treatment 13/1.

⁵ Of course, there also exists a Nash equilibrium in mixed strategies in the game.

Table 3: Payoff matrices in the three one-shot treatments

Treatments	“9/1”		“9/5”		“13/1”	
	<i>Minimum action of other group members</i>					
<i>Own action</i>	HIGH	LOW	HIGH	LOW	HIGH	LOW
HIGH	9, 9	1, 8	9, 9	5, 8	13, 13	1, 8
LOW	8, 1	8, 8	8, 5	8, 8	8, 1	8, 8

Notes: The Pareto-efficient equilibrium (HIGH, HIGH) is depicted in the top left corner of each treatment, the inefficient equilibrium (LOW, LOW) in the bottom right corner. The first numbers in a cell present the row player’s payoff. All players saw themselves in the role of the row player. The options LOW and HIGH in the columns of each treatment indicate the minimum choice of all other players in the group (excluding the row player), and second numbers in each cell show the payoffs of the second player only in the games with a group size of two subjects.

The other two treatments, shown in the middle and on the right-hand side of Table 3, vary either the payoff in the efficient equilibrium or the payoff in case of mis-coordination (i.e., the payoff for the row player in the upper right cell of the matrix). In treatment “9/5” it becomes less costly to play HIGH because of a payoff of 5 – rather than 1 as in “9/1” – if another group member chooses LOW. Finally, in treatment “13/1” the payoff in the efficient equilibrium increases from 9 – as in “9/1” – to 13, keeping everything else constant.

Given that playing HIGH becomes less costly in “9/5” in case of miscoordination and more profitable in “13/1” in case of efficient coordination, we expect to observe more choices of HIGH in these two treatments than in “9/1”. It is unclear what to expect from a comparison of “9/5” and “13/1”, however. Both treatments yield an expected payoff of 7 from playing HIGH under the assumption that it is equally likely that the minimum chosen by the other players is LOW or HIGH. If increased payoffs in the efficient equilibrium (13 points instead of 9) weigh in larger than the increased payoffs in case of mis-coordination (5 points instead of 1), then we should observe more choices of HIGH in “13/1”, which would then lead to the largest frequency of HIGH in “13/1”, an intermediate one in “9/5”, and the lowest in “9/1”.⁶

2.3 The effects of group size – Within-subjects variation

While the effects of incentives were examined in a between-subjects variation, we examined the possible effects of different group sizes by letting each subject play two one-shot games (with the same incentives each) with different group sizes. The “2-person groups” had

⁶ We note here that the often invoked concepts of payoff-dominant, respectively risk-dominant equilibrium do not help in making predictions in our case, as the efficient equilibrium of all players choosing HIGH is both payoff- and risk-dominant.

two members, meaning that the row player's payoff in any of the treatments shown in Table 3 depended on the own choice and the choice of the only other player in the group. The "5-person groups" had five group members. Given the row player's choice, the payoff for the row player from choosing HIGH was then determined by the minimum choice of the other four players in the group. In other words, choosing HIGH paid off for the row player only if all four other group members chose HIGH as well. We expect to find more choices of HIGH in 2-person groups compared to 5-person groups as the strategic uncertainty about the other group members' choices is much lower in smaller groups.

Both one-shot games were explained at the beginning of the experiment (see experimental instructions in the Appendix), yet we balanced the order in which the two games were introduced in order to control for potential order effects. After we had worked through the instructions and participants had answered two control questions, we let them make their decisions in the order as explained in the instructions. At the end of the experiment, it was randomly determined which game (first or second) was payoff relevant.

2.4 The effects of beliefs

We also elicited subjects' beliefs on the behavior of the other group members. Therefore, following the actual choice in the 2-person, respectively 5-person, groups we asked all participants to guess the minimum choice of their other group members. The belief elicitation was incentivized with one token for a correct guess. We expect that coordination choices are highly correlated with subjects' beliefs. Note that in order to make the question about beliefs easy also for our youngest participants, we only asked for a point belief (whether someone expected HIGH or LOW as the minimum choice of the others in the group). This means that we did not elicit any kind of probability for HIGH or LOW as the minimum.

Subjects had no information about the other group members other than that they were attending the same school grade, and would therefore be of the same age. It was additionally explained that the matched subjects did not attend the same school to exclude any potential influence of personal relationships in the decision-making process.⁷ Given the latter matching procedure, all subjects were only paid after all data had been collected.⁸ This was done within

⁷ This is true for all age groups and treatments except for grade 4 treatment 9/1 and grade 4 treatment 9/5. In these two subgroups only one school participated and thus we matched subjects across different classes within their school. The subjects were informed in these two cases about the matching with someone from another class in the same school.

⁸ This implies that the matching of players into groups was not done on the spot. Later it was then done manually, but due to miscommunication with research assistants it was not recorded. For this reason, we cannot provide

1-2 weeks after executing the experiments. The money was handed over in sealed envelopes with an anonymous ID-code on them.⁹

3. Experimental results

The results section will first present the results for the 2-person groups and then the 5-person groups. Recall that the effects of group size were examined in a within-subject design in which the group size was randomly ordered (first 2-person group and then 5-person group, or vice versa). We did not find any order effects (see Table A2 in the Appendix) for which reason we pool the two different orders. The figures below show the coordination behavior of our participants separately for each grade and for each of the three different incentives. Actions (i.e. choosing the efficient strategy HIGH) are discussed first and then followed-up with the beliefs (i.e. expecting the partner(s) to choose the efficient strategy HIGH). Likewise, in the regression analysis we control for our main explanatory variables grade (i.e. age), incentives (i.e. stake sizes), and group size as well as interactions thereof.

When presenting the results in the following, it helps to explain the general structure of Tables 4 and 5 already here. Table 4 shows the results of the OLS regression for the likelihood of choosing strategy HIGH, split by actions (models 1 to 3) and subsequently for beliefs (models 4 and 5). Model 1 shows the basic regression, model 2 includes measures for risk aversion, impatience, social preferences, and math grade, and model 3 includes the interaction between grade and beliefs. Similarly, model 4 shows the basic regression with the expected strategy as dependent variable, while model 5 includes additional controls. Table 5 shows the Wald-tests where we test for age effects, differences between the treatments (i.e. stake sizes), and differences between the group sizes. Models 1-5 follow the same structure as in Table 4. Additionally, columns 6 and 7 present the non-parametric test results for comparison. To correct for multiple hypothesis testing, we have adjusted our analysis by performing a Bonferroni correction in which we take into account that we have tested 6 main hypotheses at the same time (i.e. 3 incentive structures x 2 group sizes). Finally, in Table A3 and A4 in the Appendix we show probit regressions which confirm that these estimation results are largely consistent with our OLS regressions analyses.¹⁰

statistics on how often a particular equilibrium was played, yet in Section 3.4 we present simulation results on this question.

⁹ This experiment was part of a larger project (see Sutter et al., 2019), for which reason children had already experience with economic experiments and receiving delayed payoffs. Moreover, we control for impatience in our regressions (see Table 4 and Table A3 in the Appendix) which does not have an effect on participants' behavior in the coordination game.

¹⁰ We follow Haushofer and Shapiro (2016) in effectively calculating minimum detectable effect sizes (MDEs) by multiplying the standard error of our main explanatory variables in our regressions (Table 5) by 2.8. For this

3.1 2-person groups

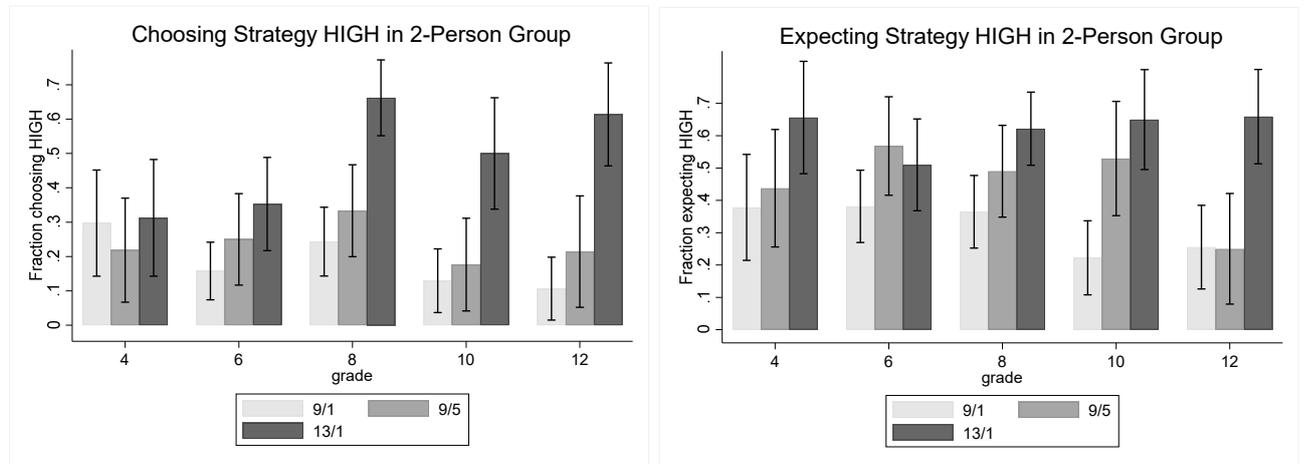
Age. We begin by looking at subjects' choices with respect to age in the 2-person group. Figure 1a shows the likelihood of choosing strategy HIGH (on the vertical axis) by grade (on the horizontal axis) for each payoff treatment. The overall frequency of choosing HIGH across all age groups and treatments is 31.2%. When we investigate decisions by grade and treatment, we do not observe a clear age trend. While the non-parametric tests show a significant increase in choosing HIGH with age in the "13/1" treatment ($p=0.019$), this is not confirmed by the Wald-tests (see Table 5). We conclude that there is no robust relationship between age and choosing HIGH across treatments in 2-person groups. Subsequently, we proceed in Figure 1b by examining the subjects' beliefs regarding their expectation of the partner's action. Similar to the findings pertaining to the subjects' actions, there is no discernible age-related trend observed in terms of anticipating the partner's inclination towards selecting HIGH in any of the treatments (see Table 5).

Incentives. When investigating how the different incentives affect choices, we observe that our subjects do differentiate between the payoff treatments. Pooling across all age groups, the frequency of choosing HIGH is 18.4% in the "9/1" treatment, 24.9% in the "9/5" treatment, and 51.5% in the "13/1" treatment (see Table A1). While coordination is less sensitive to incentives among the youngest cohort, we find stronger reactions towards incentives among our older participants (see Figure 1a). Specifically, we see significantly higher rates of choosing HIGH in treatment "13/1" compared to treatments "9/1" and "9/5". This is true across all age groups, with the exception of the youngest age group (i.e. grade 4, see Table 5).

The beliefs by and large match this pattern as a higher fraction of subjects expects their partner to choose HIGH in treatment "13/1" than in "9/1" and "9/5". The Wald-tests confirm a significant difference across all age groups between treatments "9/1" and "13/1", as well as for older age cohorts (starting at grade 8) between treatments "9/5" and "13/1" (see Table 5). The latter effect, however, is not robust to controlling for additional factors (see Table 5, model 5).

calculation, we focus on our main regression in column (1). With respect to the age effects the MDEs lie in the range of 4 to 5 percentage points (80% power and a significance level of $\alpha = 0.05$). With respect to treatment differences according to the incentive structure as well as the group size the MDEs lie in the range of 7 to 18 percentage points. We conclude that while MDEs below 10 percentage points indicate that we have enough statistical power to detect economically meaningful effects, in cases where MDEs lie above 10 percentage points the statistical power of our analysis is relatively low. When looking at the results of our paper one needs to keep in mind that null effects might be due to lack of statistical power. Yet, despite correcting for multiple hypotheses testing we still find many significant effects that are meaningful in size and robust across various different models. In that sense, we take a relatively conservative approach and only report results that are robust across different specifications.

When comparing coordination behavior between treatments “9/1” and “9/5”, no clear distinction in strategies can be found (see Table 5). Contrarily, the beliefs show that subjects expect their partners to choose HIGH more frequently in “9/5”. This is the case for all grades



with the exception of grade 12.

Figure 1: a. Relative frequency of choosing HIGH in 2-person groups; b. relative frequency of expecting the other player to choose HIGH (error bars show 95% CI)

3.2 5-person groups

Next, we show and compare choices and beliefs for the 5-person groups. Recall that each subject played both, the 2- and 5-person game, in random order. We first investigate the differences within 5-person groups by age and incentives and then contrast coordination behavior in large versus small groups.

a. **Age.** Similar to our results for 2-person groups we do not find any robust differences for action HIGH when comparing the different grade levels. While there is a significant negative effect of age on the choice of HIGH in treatment “9/5”, it is not robust to the inclusion of risk-, time-, and social preferences as well as the math grade (see Table 5). We conclude that there are no robust age effects in large groups. In line with the actions, the beliefs do not show any significant differences or developments with age either (see Table 5 and Figure 2b).

b. **Incentives.** When comparing the choice of strategy HIGH between the three treatments in 5-person groups, no robust significant differences are found (see Figure 2a and Table 5). Only when comparing treatments “9/1” and “13/1”, we see a significantly higher frequency of HIGH in the two oldest grade levels in “13/1”. However, this effect is not robust to the inclusion of economic preferences and the math grade (Table 5, model 2). The beliefs are once again aligned

with the choices as there are no significant differences in expectations (see Figure 2b and Table 5).

Group size. Finally, we investigate differences in coordination behavior across group sizes (i.e. comparing frequencies of choosing/expecting strategy HIGH in Figures 1 and 2). From the comparison of Figures 1 and 2 it stands out that the frequency of choosing strategy HIGH aggregated over all grade levels and treatments is much lower in 5-person groups (11.3%) than in 2-person groups (31.2%). A similar pattern is found in beliefs (19.6% vs. 46.1%). Both the lower fraction of actions and of beliefs (of HIGH) reflect a response to the increased strategic uncertainty of coordinating with four partners rather than just one.

Investigating group size effects by treatments we find the biggest effect on actions in the 13/1 incentive, where the average choice of HIGH is 13.7% in large groups compared to 51.5% in 2-person groups. This effect is significant across all grade levels. Similarly, the frequency of playing HIGH in treatment “9/1” is lower in the 5-person groups (8.7% versus 18.4%) and this effect is driven by the older age cohorts (namely, grades 10 and 12) as confirmed by the Wald tests in Table 5. No robust effect is found for treatment “9/5”. While the non-parametric tests show a significant effect (see Table 5, column 6), this is not confirmed by the Wald tests. Analyzing group size differences in beliefs largely confirms our results for actions. Subjects across all age groups expect in all three treatments the minimum choice of HIGH significantly less often in 5-person groups than in 2-person groups (compare Figure 2b to Figure 1b; Table 5, models 4-5 and column 7).

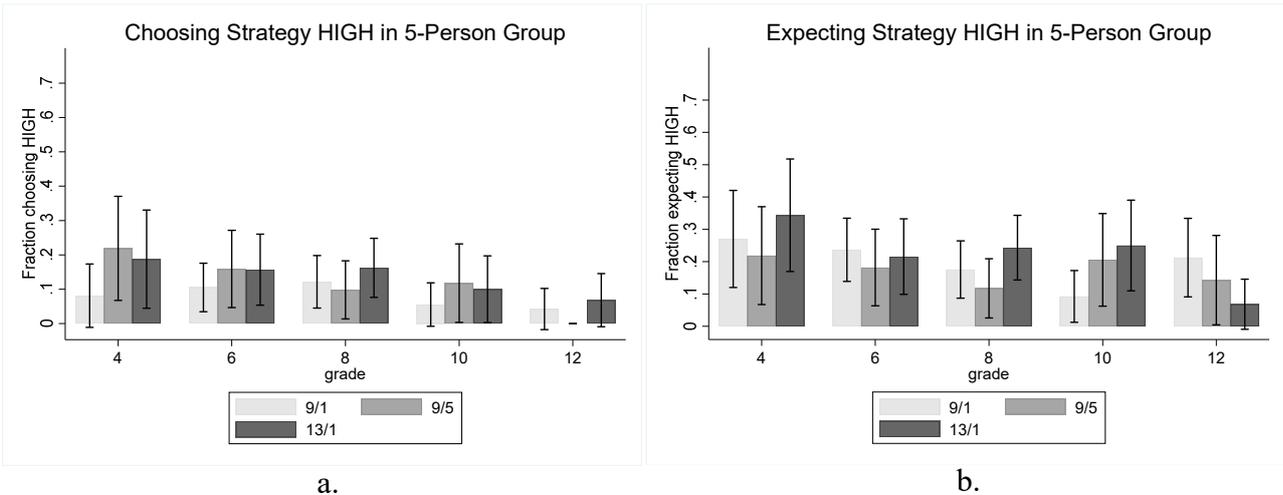


Figure 2: a. Relative frequency of choosing HIGH in 5-person groups; b. relative frequency of expecting HIGH as the minimum choice of the other four players in the 5-person group (error bars show 95% CI)

Table 4: OLS estimation of determinants of choosing strategy HIGH and expecting strategy HIGH.

Dependent variable: (1) – (3) likelihood to choose strategy HIGH, (4) – (5) likelihood to expect strategy HIGH as the minimum of the other players in one's group.

	(1) actions	(2) actions	(3) actions	(4) beliefs	(5) beliefs
Treatment 9/5	0.078 (0.087)	0.032 (0.111)	0.081 (0.087)	0.153 (0.119)	0.366* (0.149)
Treatment 13/1	0.113 (0.086)	0.138 (0.106)	0.125 (0.087)	0.212 (0.115)	0.241 (0.145)
Group size 5	0.050 (0.050)	0.039 (0.062)	0.016 (0.052)	-0.103 (0.062)	-0.171 (0.082)
Grade	-0.002 (0.016)	-0.016 (0.020)	-0.018 (0.015)	-0.029 (0.023)	-0.015 (0.029)
Grade*Treatment 9/5	-0.013 (0.023)	0.002 (0.030)	-0.014 (0.024)	0.001 (0.033)	-0.045 (0.042)
Grade*Treatment 13/1	0.041 (0.023)	0.037 (0.028)	0.034 (0.023)	0.024 (0.031)	0.016 (0.040)
Grade*group size 5	-0.034* (0.013)	-0.030 (0.017)	-0.022 (0.013)	-0.010 (0.017)	0.012 (0.022)
Treatment 9/5*group size 5	0.028 (0.042)	0.024 (0.051)	0.027 (0.042)	-0.171*** (0.053)	-0.201** (0.066)
Treatment 13/1*group size 5	-0.185*** (0.041)	-0.206*** (0.051)	-0.177*** (0.041)	-0.262*** (0.051)	-0.244*** (0.063)
Expected coordination	0.356*** (0.027)	0.381*** (0.033)	0.223*** (0.065)		
Grade*expected coordination			0.046 (0.020)		
Female	-0.027 (0.020)	-0.018 (0.026)	-0.027 (0.020)	0.043 (0.027)	0.015 (0.034)
Risk aversion		-0.151* (0.058)			0.184** (0.068)
Impatience		-0.000 (0.000)			0.000 (0.000)
Math grade		-0.008 (0.012)			0.006 (0.017)
Spiteful		0.018 (0.095)			-0.093 (0.095)
Egalitarian		-0.014 (0.092)			-0.110 (0.092)
Generous		0.004 (0.094)			-0.121 (0.091)
School fixed effects	YES	YES	YES	YES	YES
Constant	0.071 (0.114)	0.175 (0.160)	0.138 (0.116)	0.536*** (0.138)	0.531** (0.188)
R-squared	0.286	0.323	0.290	0.125	0.139
N	1,436	962	1,436	1,436	962
Number of independent observations	718	481	718	718	481

***/**/* Denotes significance at the 1%/5%/10% level. P-values corrected for multiple hypotheses testing by performing a Bonferroni correction. Robust standard errors in parentheses. Clustered on the level of individual subjects. School fixed effects are included in each regression (i.e. we include for each of the seven schools, except for school 1 as the baseline school, a dummy. In no single regression, these dummies are significant.

Table 5: Post-estimation Wald tests for regressions presented in Table 4 (columns 1 to 5) and results of corresponding non-parametric tests (columns 6 for actions and 7 for beliefs).

	(1)	(2)	(3)		(4)	(5)	(6)		(7)
	actions	actions	Wald tests		beliefs	beliefs	actions	beliefs	
H0: No age effect for ...									
... treatment 9/1 and group size 2 ^A	-0.002	-0.016	-0.018	-0.029	-0.015				
... treatment 9/5 and group size 2 ^B	-0.015	-0.013	-0.033	-0.028	-0.060				
... treatment 13/1 and group size 2 ^C	0.039	0.022	0.016	-0.006	0.001	**			
... treatment 9/1 and group size 5 ^D	-0.035	-0.046	-0.041**	-0.039	-0.003				
... treatment 9/5 and group size 5 ^E	-0.048**	-0.043	-0.055***	-0.038	-0.047	*			
... treatment 13/1 and group size 5 ^F	0.005	-0.008	-0.007	-0.015	0.013				*
H0: No treatment effect between treatment 9/1 and 9/5 and group size 2 for ...									
... grade 4 ^G	0.066	0.034	0.067	0.154	0.322**				
... grade 6 ^G	0.053	0.036	0.052	0.155*	0.277***				
... grade 8 ^G	0.040	0.038	0.038	0.156**	0.232***				
... grade 10 ^G	0.027	0.040	0.023	0.157**	0.187**				**
... grade 12 ^G	0.014	0.042	0.009	0.159	0.143				
H0: No treatment effect between treatment 9/1 and 13/1 and group size 2 for ...									
... grade 4 ^H	0.153	0.176	0.159	0.236**	0.257				
... grade 6 ^H	0.194***	0.213***	0.193***	0.260***	0.273***	*			
... grade 8 ^H	0.234***	0.250***	0.227***	0.284***	0.289***	***		**	
... grade 10 ^H	0.275***	0.288***	0.261***	0.307***	0.305***	***		***	
... grade 12 ^H	0.316***	0.325***	0.295***	0.331***	0.321***	***		***	
H0: No treatment effect between treatment 9/5 and 13/1 and group size 2 for ...									
... grade 4 ^I	0.088	0.142	0.092	0.082	-0.064				
... grade 6 ^I	0.141**	0.177*	0.141**	0.104	-0.004				
... grade 8 ^I	0.195***	0.212***	0.189***	0.127*	0.057	***			
... grade 10 ^I	0.248***	0.247***	0.237***	0.150**	0.117	**			
... grade 12 ^I	0.301***	0.283***	0.286***	0.173*	0.178	***		***	
H0: No treatment effect between treatment 9/1 and 9/5 and group size 5 for ...									
... grade 4 ^J	0.093	0.058	0.094	-0.016	0.121				
... grade 6 ^J	0.080	0.060	0.080	-0.015	0.076				
... grade 8 ^J	0.068	0.062	0.065	-0.014	0.031				
... grade 10 ^J	0.055	0.065	0.051	-0.013	-0.014				
... grade 12 ^J	0.042	0.067	0.036	-0.012	-0.058				
H0: No treatment effect between treatment 9/1 and 13/1 and group size 5 for ...									

... grade 4 ^K	-0.032	-0.030	-0.017	-0.026	0.013		
... grade 6 ^K	0.009	0.007	0.016	-0.002	0.029		
... grade 8 ^K	0.050	0.044	0.050	0.022	0.045		
... grade 10 ^K	0.090**	0.082	0.084**	0.046	0.061		
... grade 12 ^K	0.131**	0.119	0.118**	0.070	0.076		
H0: No treatment effect between treatment 9/5 and 13/1 and group size 5 for ...							
... grade 4 ^L	-0.125	-0.088	-0.112	-0.010	-0.108		
... grade 6 ^L	-0.071	-0.053	-0.063	0.013	-0.047		
... grade 8 ^L	-0.018	-0.018	-0.015	0.036	0.014		
... grade 10 ^L	0.035	0.017	0.033	0.059	0.074		
... grade 12 ^L	0.089	0.052	0.081	0.082	0.135		
H0: No large group effect in treatment 9/1 for ...							
... grade 4 ^M	0.017	0.009	-0.006	-0.112	-0.159*		
... grade 6 ^M	-0.017	-0.021	-0.029	-0.122***	-0.146**		
... grade 8 ^M	-0.050	-0.051	-0.051	-0.132***	-0.134***		*
... grade 10 ^M	-0.084***	-0.081**	-0.073**	-0.142***	-0.122**		
... grade 12 ^M	-0.117***	-0.111**	-0.096**	-0.152***	-0.109		
H0: No large group effect in treatment 9/5 for ...							
... grade 4 ^N	0.044	0.034	0.021	-0.283***	-0.360***		
... grade 6 ^N	0.011	0.004	-0.001	-0.293***	-0.347***		***
... grade 8 ^N	-0.023	-0.026	-0.024	-0.303***	-0.335***	**	***
... grade 10 ^N	-0.056	-0.056	-0.046	-0.312***	-0.323***		**
... grade 12 ^N	-0.090	-0.086	-0.068	-0.322***	-0.310***	*	
H0: No large group effect in treatment 13/1 for ...							
... grade 4 ^O	-0.168***	-0.197***	-0.183***	-0.374***	-0.403***		*
... grade 6 ^O	-0.202***	-0.227***	-0.205***	-0.384***	-0.391***		**
... grade 8 ^O	-0.235***	-0.257***	-0.228***	-0.394***	-0.378***	***	***
... grade 10 ^O	-0.269***	-0.287***	-0.250***	-0.403***	-0.366***	***	***
... grade 12 ^O	-0.302***	-0.317***	-0.273***	-0.413***	-0.354***	***	***
H0: No expected coordination effect for ...							
... grade 4 ^P			0.269***				
... grade 6 ^P			0.315***				
... grade 8 ^P			0.361***				
... grade 10 ^P			0.407***				
... grade 12 ^P			0.453***				

***/**/* Denotes significance at the 1%/5%/10% level. P-values corrected for multiple hypotheses testing by performing a Bonferroni correction.

^A ($\beta_{\text{grade}=0}$)

^B ($\beta_{\text{grade}+ \beta_{\text{grade}*\text{treatment } 9/5}=0}$)

^C ($\beta_{\text{grade}+ \beta_{\text{grade}*\text{treatment } 13/1}=0}$)

^D ($\beta_{\text{grade}+ \beta_{\text{grade}*\text{group size } 5}=0}$)

- ^E ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 9/5 + \beta_{\text{grade}} * \text{group size } 5 = 0$)
- ^F ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 13/1 + \beta_{\text{grade}} * \text{group size } 5 = 0$)
- ^G ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)
- ^H ($\beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)
- ^I ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] = \beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}]$)
- ^J ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0$)
- ^K ($\beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5 = 0$)
- ^L ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0 = \beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5$)
- ^M ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)
- ^N ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0$)
- ^O ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5 = 0$)
- ^P ($\beta_{\text{expected coordination}} + \beta_{\text{grade}} * \text{expected coordination} * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)

3.3 Best response play

Figure 3 addresses the question of how many of our subjects are playing best response to their expectations about others' behavior. As evident from Figure 3 subjects play best response with a very high probability when they believe the minimum choice of the other players in their group is LOW. When subjects believe the minimum choice is HIGH, they are less likely to play best response. This pattern is equally found in small and large groups. In 2-person groups approximately 90% of our subjects (across all age groups) play best response if they believe their partner chooses LOW, while only 56% play best response if they believe their partner chooses HIGH. In 5-person groups we find an even larger gap in best response rates. If subjects believe that LOW is the minimum choice of the other players in their group, they best respond on average in 94% of times. However, if subjects believe that HIGH is the minimum choice of the other players in their group, the best response rate drops to only 32%. Our design choice of only asking for binary beliefs (either HIGH or LOW) may have contributed to this difference in best response rates, contingent on subjects' beliefs. It may be the case that subjects expecting LOW are almost certain about the other's behavior, and thus have no reason at all to choose HIGH themselves. On the contrary, subjects expecting HIGH may be less sure about it, assigning also some probability (but less than 50%) to LOW, in which case choosing LOW themselves may be a reasonable choice.

Concerning the effects of age, we see that best response rates seem to increase somewhat in older subjects. While 85% of 4th-graders in 2-person groups play best response if they expect their partner to choose LOW, this fraction increases to 94% in 12th-graders, yet this is not significant ($p=0.7$ Cuzick's Wilcoxon-type test for trend). Given the expectation that the partner chooses HIGH, best response rates increase significantly from 41% (4th-graders) to 71% (12th-graders) in 2-person groups ($p=0.007$ Cuzick's Wilcoxon-type test for trend). A similar pattern can be found in 5-person groups for subjects who expect the minimum choice of the other players in their group to be LOW. While 88% of 4th-graders in this group play best response, this fraction increases significantly to 99% for 12th-graders ($p=0.029$ Cuzick's Wilcoxon-type test for trend). No increase with age can be found for subjects who expect the minimum choice of the other players in their group to be HIGH. Therefore, the ability to incorporate the beliefs in own decisions and best respond to these beliefs is by and large more pronounced in older children and teenagers.

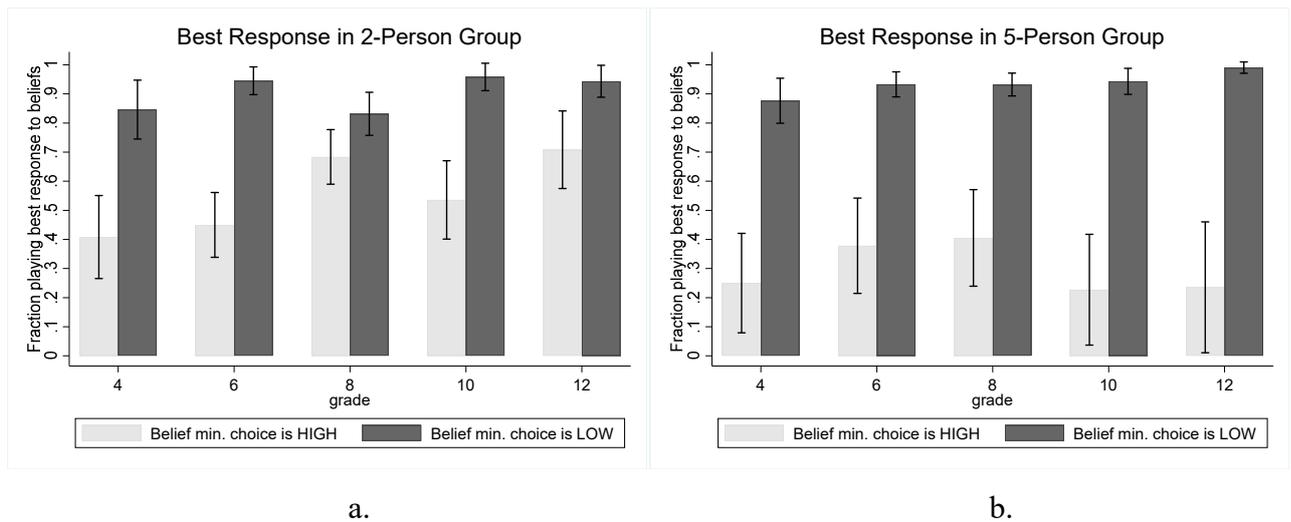


Figure 3: Relative frequency of playing best response to one's own beliefs about the minimum choice of the others in one's group; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

Figure 3 has shown best response rates, conditional on own beliefs. It has not yet shown how the beliefs themselves relate to one's own choices. This is shown in Figure 4, in panel a for 2-person groups and in panel b for 5-person groups. The bars indicate how likely subjects expect HIGH as the minimum choice of the other players in their group, conditional on own choices. The left bar in each grade refers to subjects choosing HIGH themselves, the right bar to the others choosing LOW. In both panels, we see a very strong difference between left and right bars, as one would expect. Subjects choosing HIGH themselves are much more likely to expect HIGH as the minimum of the others in their group. In 2-person groups, this relative frequency is 83.5% over all age groups, while for subjects playing LOW themselves we see that only 29.1% expect HIGH as the minimum. For 5-person groups, the corresponding numbers are 55.6% (for those playing HIGH) and 15.1% (for those playing LOW). These clear differences indicate that beliefs and actions are closely related. In the final subsection, we deal with this relationship from yet another angle.

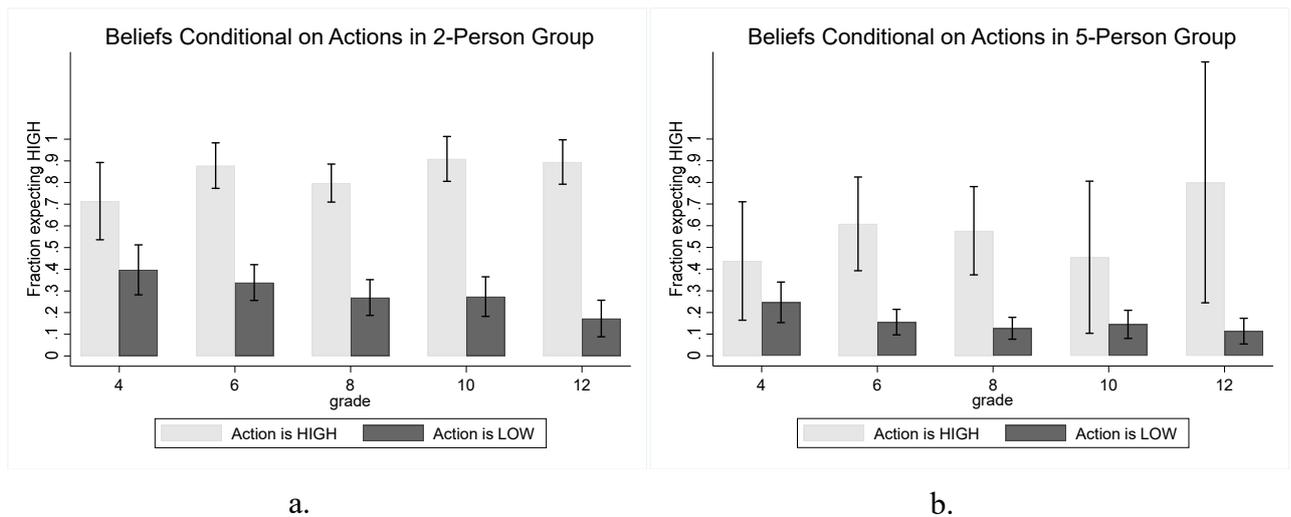


Figure 4: Relative frequency of expecting HIGH to be the minimum choice of the others in one's group contingent on action HIGH and LOW; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

3.4 Simulations on the level of (mis)coordination and the accuracy of beliefs

Recall that due to some miscommunication with helpers we did not recall who was actually matched with whom (see also footnote 8 above). Nevertheless, we can run simulations to examine the relative frequency with which coordination – on the efficient or the inefficient equilibrium – or miscoordination (when actions are not identical within a group) can be expected. To do so, participants were randomly matched 1,000 times in either pairs (2-person game) or in groups of 5 (see Section 2.4 for the matching procedure). Figure 5 illustrates the likelihood of successful coordination – meaning that the subject's action and the minimum choice of the other members of the group is the same (independent of which action) – by age and incentives for both group sizes. Figures A3 and A4 in the Appendix disentangle this further into cases of successful coordination on the efficient or the inefficient equilibrium (absolute frequency of respective equilibria among 1,000 simulations on the vertical axis). The accuracy of beliefs is demonstrated in Figure 6 by showing how many times (out of 1,000 simulations) one's beliefs match the others' actions. Figures A5 and A6 in the Appendix disentangle this for the different beliefs (of expecting others to play HIGH or LOW). Table A5 in the Appendix presents parametric and non-parametric test results for age, incentive, and group size effects. The overall frequency of successful coordination across all age groups and treatments is 75.1%. This is a relatively large fraction, yet it is driven by a very large degree of coordination on the

inefficient equilibrium, as 69 percentage points stem from inefficient coordination, and only 6.1 percentage points from efficient coordination.

Age. Looking at subjects' coordination success with respect to age, Figure 5 and Table A5 show that successful coordination increases significantly with age in treatment "9/1" and "9/5" in 2-person groups (see Figure 5a) and in treatment "9/5" in 5-person groups (see Figure 5b). A look at Figures A3 and A4 reveals that this is driven by a significant increase of inefficient coordination by age. Conversely, Figures A3 and A4 illustrate that for 2-person groups in treatment "13/1" we find a significantly declining age trend in inefficient coordination and a significantly increasing one in efficient coordination. The net effect on overall successful coordination is insignificant for this sub-group. Subsequently, we proceed by examining the correctness of subjects' beliefs. Similar to the results pertaining to successful coordination, we find a significant increase in correct beliefs in treatment "9/1" for 2- and 5-person groups (see Figure 6) which is driven by an increase of correct beliefs conditional on own LOW beliefs (see Figure A5). Moreover, there is a significant increase of correct beliefs given a LOW belief in treatment "9/5" and a significant decrease in treatment "13/1" in 2-person groups (see Figure A5).

Incentives. When investigating how the different incentives affect the occurrence of different equilibria, we observe that our subjects do differentiate between the payoff treatments. Pooling across all age groups, the occurrence of successful coordination is 80.3% in the "9/1" treatment, 74.7% in the "9/5" treatment, and 69.3% in the "13/1" treatment. Again, this is driven by inefficient coordination (78.5 percentage points in treatment "9/1", 71.6 percentage points in treatment "9/5", and 55.6 percentage points in treatment "13/1"). While successful coordination is less sensitive to incentives among the youngest cohort, we find stronger reactions towards incentives among our older participants in 2-person groups (see Figure 5a, Figure A3a, Figure A4a). Largely consistent with our findings for actions in Section 3.1 we see by and large more successful coordination in treatments "9/1" and "9/5" compared to treatment "13/1", driven by more inefficient and less efficient coordination. No significant incentive effects are found for 5-person groups. The accuracy of beliefs generally aligns with this pattern, as a greater proportion of subjects hold correct beliefs when their belief is LOW, and a smaller proportion hold correct beliefs when their belief is HIGH. These incentive effects are observed in the accuracy of beliefs for all treatment comparisons in 2-person groups starting from grade 6 onwards. No clear pattern in the accuracy of beliefs is found for 5-person groups.

Group size. Finally, we investigate differences across group sizes. Overall, the pattern is very clear - both regarding the occurrence of different equilibria as well as the accuracy of beliefs. We observe significantly fewer instances of successful coordination in 2-person groups compared to 5-person groups (see Figure 5a and Figure 5b) and this effect is driven by a lower likelihood of reaching the inefficient equilibrium (see Figure A3a and A3b). A similar pattern is found in the accuracy of beliefs. The correctness of beliefs is significantly higher in 5-person groups compared to 2-person groups (see Figure 6a and Figure 6b) and this is driven by a higher accuracy of LOW-beliefs in large groups (see Figure A5a and Figure A5b).

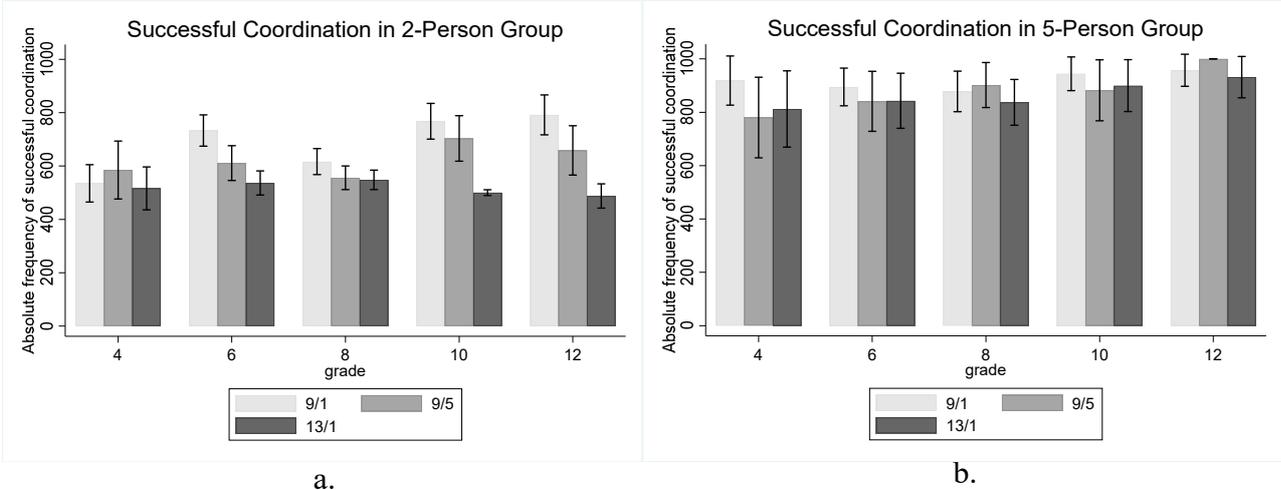


Figure 5: Absolute frequency of successful coordination for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

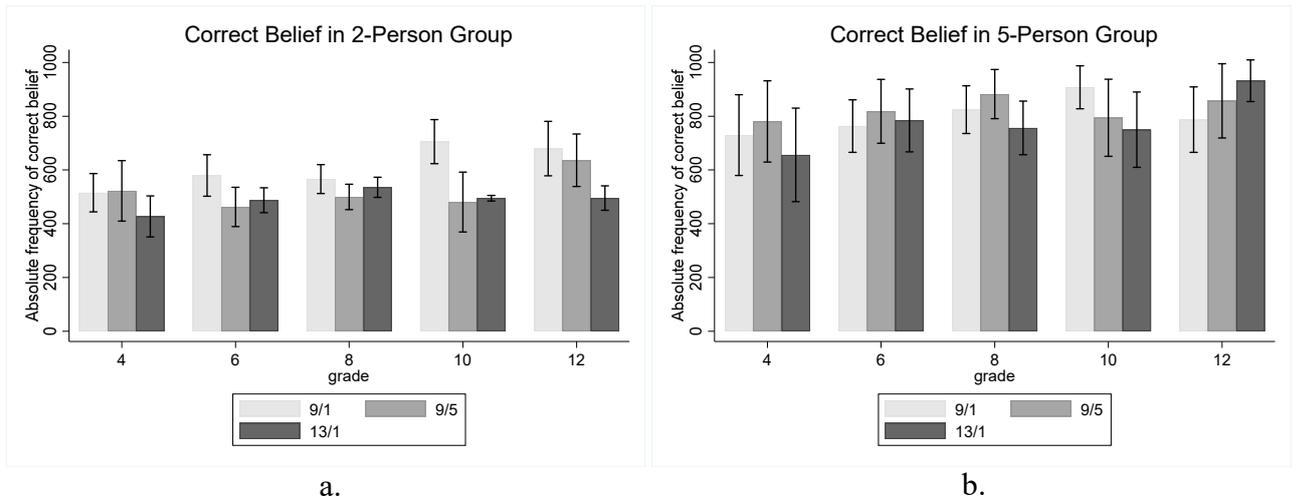


Figure 6: Absolute frequency of correct beliefs for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

4. Conclusion

We have studied the evolvement of coordination behavior of children and teenagers in a large experiment with 718 participants, aged nine to 18 years. We have focused on the influence of age, group size, and incentives on the likelihood to play the efficient strategy in an experimental stag-hunt game. While the effects of group size and monetary incentives have been investigated previously in coordination games with adults (van Huyck et al., 1990, Brandts and Cooper, 2006; Weber, 2006), the relationship of age and the ability to choose the more efficient strategy has not received sufficient attention so far. Given that coordination problems abound also in the pre-adulthood age (e.g., selecting on which social media to be active), we consider it as important to understand the development of choosing the efficient strategy in the (pre)teenage years.

Overall we have found the following main results: age does not have a robust effect. While we find some interaction effects with group size and monetary incentives none of the age effects are robust to the inclusion of control variables. In sum, this means that in our one-shot coordination games, age plays no clear role. This is different from the finding of Brocas and Carrillo (2021a) that in repeated coordination games (across 96 periods in total), older adolescents become more efficient than younger ones. This might be due to learning, which was ruled out in our setting. In our one-shot games, age is no unambiguous predictor of behavior. Yet, for the other two main factors under investigation – group size and incentives – we find clear effects.

The second main finding refers to the effects of group size. We have seen a much higher likelihood of choosing the efficient strategy in the 2-person groups than in the 5-person groups. The strategic uncertainty in larger groups is obviously harmful for the choice of the efficient strategy. This result is completely consistent with earlier experimental work with adults, starting with van Huyck et al. (1990) and discussed extensively in Weber (2006) or Cooper and Weber (2020). Hence, children and teenagers react in a predictable way to strategic uncertainty in larger groups, as adults do. Our simulations show, however, that coordination on one equilibrium is much more likely for the inefficient equilibrium. Subjects who expect others to play the inefficient strategy choose this strategy themselves almost always, while those expecting others to play the efficient strategy deviate themselves more frequently.

Our third main finding concerns the relationship of monetary incentives and the choice of the efficient strategy. The larger these incentives, the more likely it is that subjects choose the efficient strategy. Here we have found that increasing the profits in the efficient equilibrium (in “13/1”) seems to work better than providing a cushion (in “9/5”) for subjects who choose the efficient strategy, but where groups fail to achieve the efficient equilibrium. Overall, our results about the effects of incentives seem to be in line with what has been found in experiments with adults (see Brandts and Cooper, 2006, Devetag and Ortmann, 2007, or Cooper and Weber, 2020), indicating that already children and teenagers react in a predictable way to monetary incentives in coordination games. Similarly, children and teenagers form beliefs that are related, in the aggregate, to their actions. Those who believe others in their group also play the efficient strategy are more likely to play the efficient strategy themselves.

For future research it would be interesting to investigate the influence of Theory of Mind (ToM) on children’s beliefs about the choice of their partner(s). This ability develops from early childhood into adolescence (Valle et al., 2015) and has been shown to affect strategic reasoning. Consequently, it may shape the proficiency of individuals in adopting equilibrium strategies within the context of game theoretical research designs. A study conducted by Brocas and Carillo (2020) has underscored the noteworthy development of ToM and logical reasoning during early childhood. Their findings suggest that children's strategic decision-making is primarily influenced by elements of complexity in the game. Following their example, we believe it would be interesting to investigate how ToM affects subjects’ choices with increasing age as well as increasing complexity of the game, i.e. group size.

References

- Ai, C., Norton, E. C., 2003. Interaction terms in logit and probit models. *Economics Letters* **80** (1), 123–129.
- Bettinger, E., Slonim, R., 2007. Patience among children. *Journal of Public Economics* **91**, 343–363.
- Brandts, J., Cooper, D. J., 2006. A change would do you good: An experimental study on how to overcome coordination failure in organizations. *American Economic Review* **96**, 669–693.
- Brandts, J., Cooper, D. J., Fatas, E., Qi, S., 2016. Stand by me – Experiments on help and commitment in coordination games. *Management Science* **62**, 2916–2936.
- Brocas, J., Carillo, J., 2020. Iterative dominance in young children: Experimental evidence in simple two-person games. *Journal of Economic Behavior and Organization* **179**, 623–637.
- Brocas, J., Carillo, J., 2021a. Dynamic coordination in efficient and fair outcomes: a developmental perspective. Working Paper, University of Southern California.
- Brocas, I., Castillo, J., 2021b. Steps of reasoning in children and adolescents. *Journal of Political Economy* **129** (7), 2067–2111.
- Brosig-Koch, J., Heinrich, T., Helbach, C., 2015. Exploring the capability to reason backwards: An experimental study with children, adolescents, and young adults. *European Economic Review* **74**, 286–302.
- Camerer, C. F., 2003. *Behavioral Game Theory*. Princeton: Princeton University Press.
- Cason, T. N., Savikhin, A. C., Sheremeta, R. M., 2012a. Behavioral spillovers in coordination games. *European Economic Review* **56**, 233–245.
- Cason, T. N., Sheremeta, R. M., Zhang, J., 2012b. Communication and efficiency in competitive coordination games. *Games and Economic Behavior* **76**, 26–43.
- Castillo, M., Ferraro, P. J., Jordan, J. L., Petrie, R., 2011. The today and tomorrow of kids: time preferences and educational outcomes of children. *Journal of Public Economics* **95**, 1377–1385.
- Cooper, D. J., Weber, R. A., 2020. Recent advances in experimental coordination games. In: Capra, C. M., Croson, R., Rigdon, M., Rosenblat, T. (eds). *Handbook of Experimental Game Theory*. Edward Elgar, Cheltenham, 149–183.
- Cooper, R., De Jong, D., Forsythe R., Ross, T., 1990. Selection criteria in coordination games: Some experimental results. *American Economic Review* **80**, 218–233.

- Czermak, S., Feri, F., Glätzle-Rützler, D., Sutter, M., 2016. How strategic are children and adolescents? Experimental evidence from normal-form games. *Journal of Economic Behavior and Organization* **128**, 265-285.
- Devetag, G., Ortmann, A., 2007. When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics* **10**, 331-344.
- Duffy, J., Feltovich, N., 2006. Words, deeds and lies: Strategic behavior in games with multiple signals. *Review of Economic Studies* **73**, 669-688.
- Ellsberg, D. 1961. Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics* **75**, 643-669.
- Fehr, E., Glätzle-Rützler, D., Sutter, M. 2013. The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review* **64**, 369-383.
- Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* **10**, 171-178.
- Grueneisen, S., Wyman, E., Tomasello, M., 2015a. “I know you don’t know I know ...” Children use second-order false-belief reasoning for peer coordination. *Child Development* **86**, 287-293.
- Grueneisen, S., Wyman, E., Tomasello, M., 2015b. Children use salience to solve coordination problems. *Developmental Science* **18**, 495-501.
- Kölle, F., Wenner, L., 2023. Time-inconsistent generosity: Present bias across individual and social contexts. *Review of Economics and Statistics* **105**, 683–699.
- Ochs, J. Coordination problems. 1995. In: Kagel, J. H., Roth, A. E. (eds.), *Handbook of Experimental Economics*, Princeton University Press, 195-252.
- Schmidt, D., 2003. Playing safe in coordination games: The roles of risk dominance, payoff dominance, and history of play. *Games and Economic Behavior* **42**, 281-299.
- Sutter, M., Kocher, M., Glätzle-Rützler, D., Trautmann, S., 2013. Impatience and uncertainty: Experimental decisions predict adolescents’ field behavior. *American Economic Review* **103**, 510-531.
- Sutter, M., Zoller, C., Glätzle-Rützler, D., 2019. Economic behavior of children and adolescents – A first survey of experimental economics results. *European Economic Review* **111**, 98-121.
- Valle, A., Massaro, D., Castelli, I., Marchetti, A., 2015. Theory of mind development in adolescence and early adulthood: the growing complexity of recursive thinking ability. *European Journal of Psychology* **11**, 112-124.

- Van Huyck, J. B., Battalio, R. C., Beil, R. O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review* **80**, 234-248.
- Weber, R. A., 2006. Managing growth to achieve efficient coordination in large groups. *American Economic Review* **96**, 114-126.

Appendix –intended for online-publication

Risk			
[1]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 0.50 euro for sure
[2]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 1 euro for sure
[3]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 1.50 euro for sure
[4]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 2 euro for sure
[5]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 2.50 euro for sure
[6]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 3 euro for sure
[7]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 3.50 euro for sure
[8]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 4 euro for sure
[9]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 4.50 euro for sure
[10]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 5 euro for sure
[11]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 5.50 euro for sure
[12]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 6 euro for sure
[13]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 6.50 euro for sure
[14]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 7 euro for sure
[15]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 7.50 euro for sure
[16]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 8 euro for sure
[17]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 8.50 euro for sure
[18]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 9 euro for sure
[19]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 9.50 euro for sure
[20]	draw from bag A	<input type="radio"/>	or <input type="radio"/> 10 euro for sure

Figure A1: Sample choice list for the measure of risk attitudes (subjects could either draw from bag A or choose fixed payment. Bag A contained 10 white and 10 orange balls – if right color was picked from bag payment of €10)

Time preferences – Decision sheet <i>(the order of decision sheets was random)</i>					
	amount today		or		amount in 3 weeks
[1]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	10.10 euro in 3 weeks
[2]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	10.30 euro in 3 weeks
[3]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	10.50 euro in 3 weeks
[4]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	10.70 euro in 3 weeks
[5]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	10.90 euro in 3 weeks
[6]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	11.10 euro in 3 weeks
[7]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	11.30 euro in 3 weeks
[8]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	11.50 euro in 3 weeks
[9]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	11.70 euro in 3 weeks
[10]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	11.90 euro in 3 weeks
[11]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	12.10 euro in 3 weeks
[12]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	12.30 euro in 3 weeks
[13]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	12.50 euro in 3 weeks
[14]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	12.70 euro in 3 weeks
[15]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	12.90 euro in 3 weeks
[16]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	13.10 euro in 3 weeks
[17]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	13.30 euro in 3 weeks
[18]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	13.50 euro in 3 weeks
[19]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	13.70 euro in 3 weeks
[20]	10.10 euro today	<input type="radio"/>	or	<input type="radio"/>	13.90 euro in 3 weeks

Figure A2: Sample choice list for the measure of impatience (Impatience calculated by summing up all the instances where a subject chose the earlier (and smaller) payoff across all 8 different choice lists High value – impatient, Low value – patient)

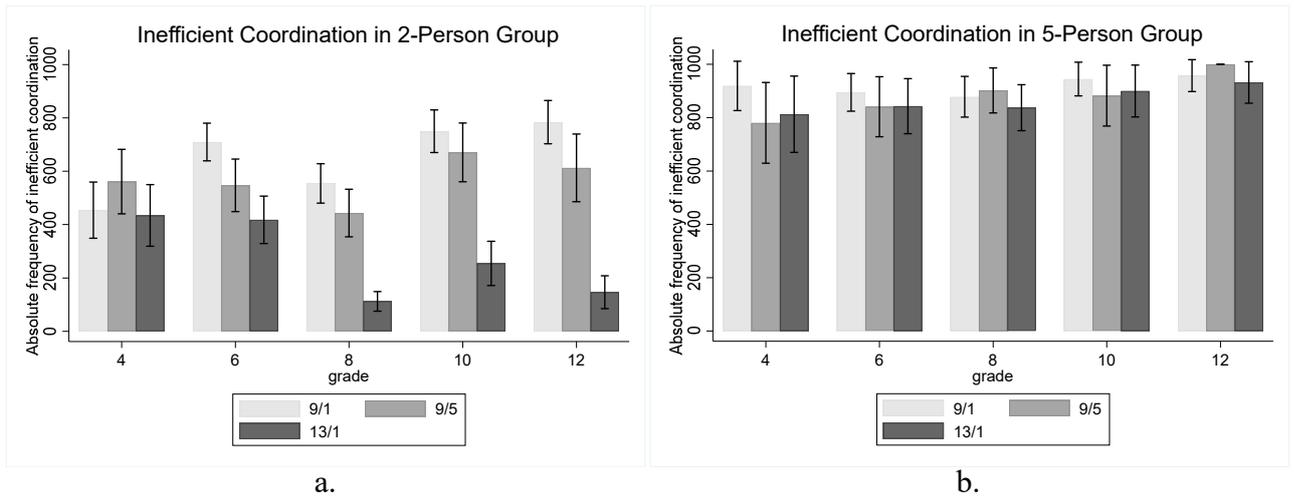


Figure A3: Absolute frequency of inefficient coordination for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

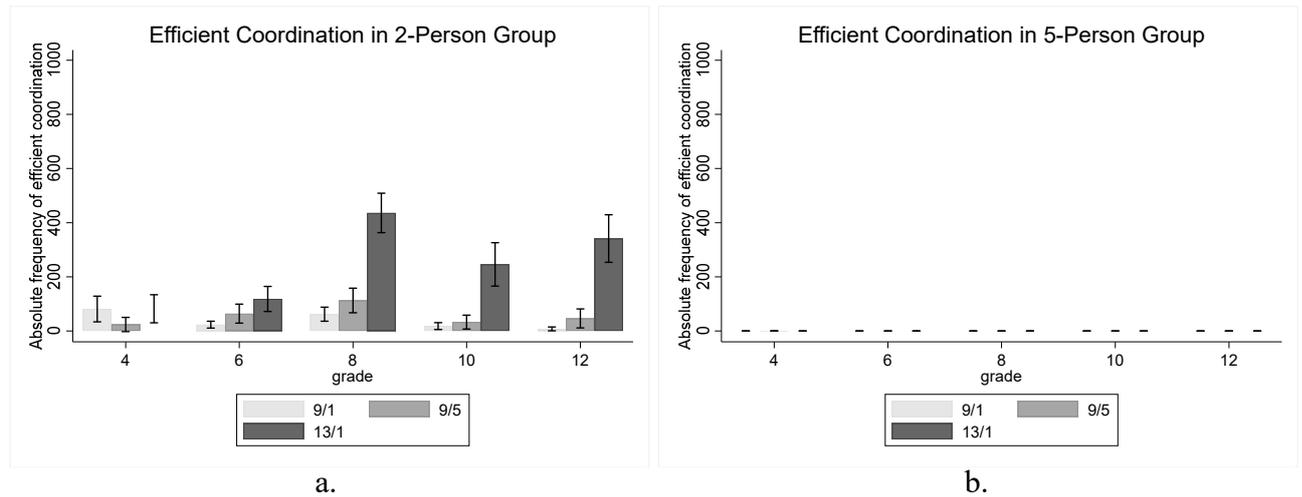
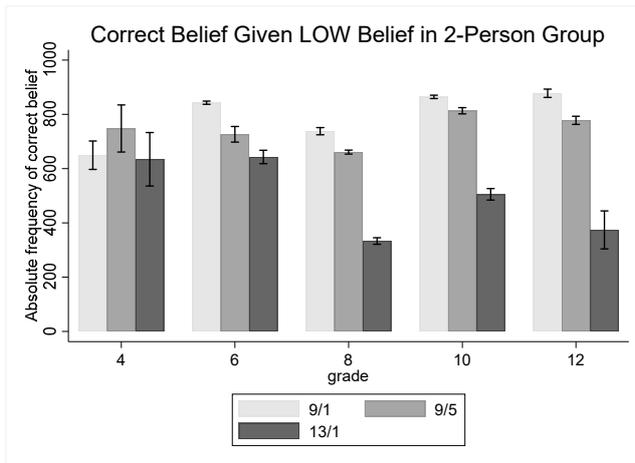
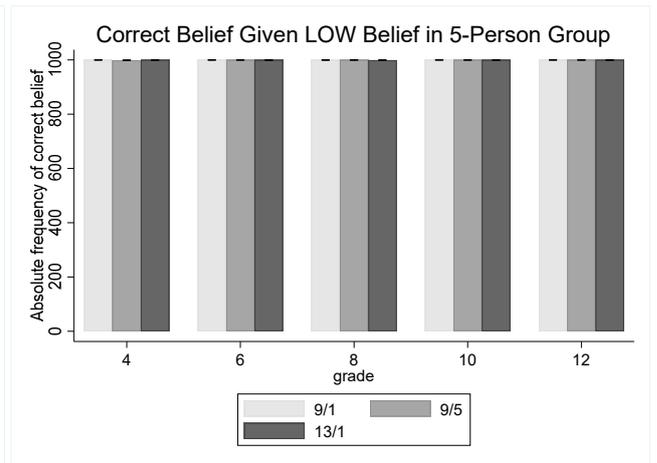


Figure A4: Absolute frequency of efficient coordination for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

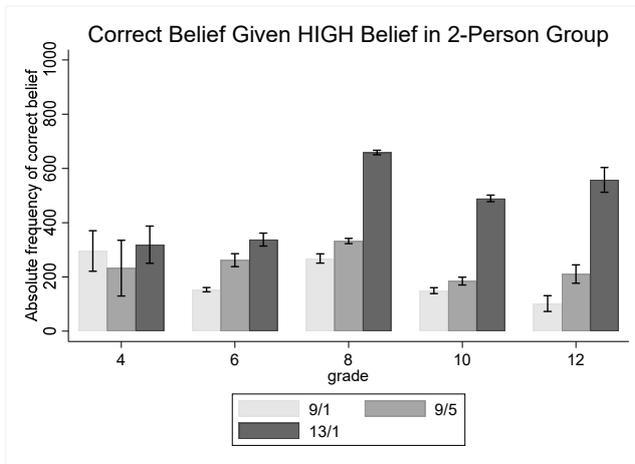


a.

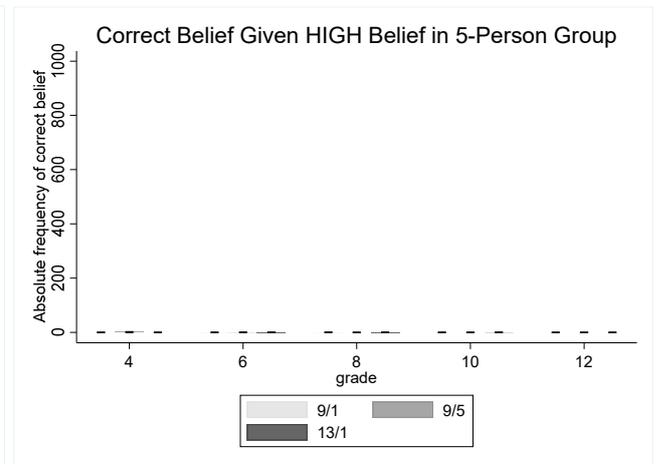


b.

Figure A5: Absolute frequency of correct beliefs given belief LOW for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)



a.



b.

Figure A6: Absolute frequency of correct beliefs given belief HIGH for 1,000 simulations; a. in 2-person groups; b. in 5-person groups (error bars show 95% CI)

Table A1: Summary results

Choosing strategy HIGH	2-person groups (actions)	5-person groups (actions)	2-person groups (beliefs)	5-person groups (beliefs)
Treatment 9/1				
Overall	.184	.087	.326	.194
Grade 4	.297	.081	.378	.270
Grade 6	.158	.105	.382	.237
Grade 8	.243	.122	.365	.176
Grade 10	.130	.056	.222	.093
Grade 12	.106	.043	.255	.213
Treatment 9/5				
Overall	.249	.122	.471	.169
Grade 4	.219	.219	.438	.219
Grade 6	.25	.159	.568	.182
Grade 8	.333	.098	.490	.118
Grade 10	.176	.118	.529	.206
Grade 12	.214	0	.250	.143
Treatment 13/1				
Overall	.515	.137	.614	.220
Grade 4	.313	.188	.656	.344
Grade 6	.353	.157	.510	.216
Grade 8	.662	.162	.622	.243
Grade 10	.5	.1	.650	.250
Grade 12	.614	.068	.659	.068

Table A2: OLS estimation of determinants of choosing strategy HIGH and expecting strategy HIGH.

Dependent variable: (1) likelihood to choose strategy HIGH, (2) likelihood to expect strategy HIGH as the minimum of the other players in one's group.

	(1)	(2)
Treatment 9/5	0.079 (0.087)	0.156 (0.118)
Treatment 13/1	0.111 (0.086)	0.207 (0.115)
Group size 5	0.050 (0.051)	-0.103 (0.062)
Grade	-0.002 (0.016)	-0.028 (0.023)
Grade*Treatment 9/5	-0.013 (0.023)	-0.001 (0.033)
Grade*Treatment 13/1	0.041 (0.023)	0.025 (0.031)
Grade*group size 5	-0.034* (0.013)	-0.010 (0.017)
Treatment 9/5*group size 5	0.028 (0.042)	-0.171*** (0.053)
Treatment 13/1*group size 5	-0.185*** (0.041)	-0.262*** (0.051)
Expected coordination	0.355*** (0.027)	
Female	-0.027 (0.020)	0.043 (0.027)
Order	-0.016 (0.020)	-0.054 (0.025)
School dummies	YES	YES
Constant	0.083 (0.115)	0.571*** (0.139)
R-squared	0.287	0.128
N	1,436	1,436
Number of independent observations	718	718

***/**/* Denotes significance at the 1%/5%/10% level.

Robust standard errors in parentheses. Clustered on the level of individual subjects. School fixed effects are included in each regression (i.e. we include for each of the seven schools, except for school 1 as the baseline school, a dummy. In no single regression, these dummies are significant.

*Table A3: Marginal probit estimation of determinants of choosing strategy HIGH and expecting strategy HIGH.
 Dependent variable: (1) – (3) likelihood to choose strategy HIGH, (4) – (5) likelihood to expect strategy HIGH as the minimum of the other players in one's group.*

	(1)	(2)	(3)	(4)	(5)
Treatment 9/5	YES	YES	YES	YES	YES
Treatment 13/1	YES	YES	YES	YES	YES
Group size 5	YES	YES	YES	YES	YES
Grade	YES	YES	YES	YES	YES
Grade*Treatment 9/5	YES	YES	YES	YES	YES
Grade*Treatment 13/1	YES	YES	YES	YES	YES
Grade*group size 5	YES	YES	YES	YES	YES
Treatment 9/5*group size 5	YES	YES	YES	YES	YES
Treatment 13/1*group size 5	YES	YES	YES	YES	YES
Expected coordination	0.358*** (0.028)	0.391*** (0.034)			
Grade*expected coordination			YES		
Female	-0.034 (0.023)	-0.033 (0.029)	-0.034 (0.023)	0.046 (0.029)	0.017 (0.037)
Risk aversion		-0.162* (0.061)			0.210** (0.077)
Impatience		0.000 (0.000)			0.000 (0.000)
Math grade		-0.008 (0.013)			0.007 (0.018)
Spiteful		0.012 (0.095)			-0.093 (0.089)
Egalitarian		-0.019 (0.087)			-0.112 (0.090)
Generous		-0.009 (0.091)			-0.128 (0.090)
School fixed effects	YES	YES	YES	YES	YES
Pseudo R-squared	0.273	0.311	0.279	0.099	0.110
N	1,436	962	1,436	1,436	962
Number of independent observations	718	481	718	718	481

***/**/* Denotes significance at the 1%/5%/10% level. P-values corrected for multiple hypotheses testing by performing a Bonferroni correction. The table shows marginal effects. Robust standard errors in parentheses. Clustered on the level of individual subjects. Note that standard software does not correctly calculate marginal effects and p-values of interaction terms in nonlinear models (see Ai and Norton, 2003). We corrected for this. School fixed effects are included in each regression (i.e. we include for each of the seven schools, except for school 1 as the baseline school, a dummy. In no single regression, these dummies are significant.

Table A4: Post-estimation Wald Tests for regressions presented in Table A3 (columns 1 to 5).

	(1)	(2)	(3)	(4)	(5)
	action	action	Wald tests action	beliefs	beliefs
H0: No age effect for ...					
... treatment 9/1 and group size 2 ^A	-0.010	-0.018	-0.038	-0.031	-0.013
... treatment 9/5 and group size 2 ^B	-0.022	-0.014	-0.055	-0.032	-0.066
... treatment 13/1 and group size 2 ^C	0.044	0.028	-0.007	-0.005	0.002
... treatment 9/1 and group size 5 ^D	-0.029	-0.036	-0.040**	-0.037	-0.002
... treatment 9/5 and group size 5 ^E	-0.062**	-0.056	-0.079***	-0.034	-0.039
... treatment 13/1 and group size 5 ^F	-0.014	-0.026	-0.034	-0.019	0.010
H0: No treatment effect between treatment 9/1 and 9/5 and group size 2 for ...					
... grade 4 ^G	0.067	0.024	0.072	0.161	0.344**
... grade 6 ^G	0.053	0.029	0.055	0.161	0.295***
... grade 8 ^G	0.040	0.033	0.039	0.160***	0.243***
... grade 10 ^G	0.029	0.037	0.026	0.159**	0.189*
... grade 12 ^G	0.019	0.039	0.016	0.156	0.136
H0: No treatment effect between treatment 9/1 and 13/1 and group size 2 for ...					
... grade 4 ^H	0.130	0.156	0.141	0.234	0.264
... grade 6 ^H	0.182**	0.204**	0.187**	0.261***	0.279***
... grade 8 ^H	0.235***	0.250***	0.231***	0.286***	0.294***
... grade 10 ^H	0.289***	0.295***	0.272***	0.311***	0.309***
... grade 12 ^H	0.344***	0.339***	0.311***	0.334***	0.324***
H0: No treatment effect between treatment 9/5 and 13/1 and group size 2 for ...					
... grade 4 ^I	0.063	0.133	0.069	0.073	-0.080
... grade 6 ^I	0.129	0.175	0.132	0.100	-0.016
... grade 8 ^I	0.195***	0.217***	0.191***	0.126*	0.052
... grade 10 ^I	0.260***	0.258***	0.246***	0.152*	0.120
... grade 12 ^I	0.325***	0.300**	0.295***	0.178	0.188
H0: No treatment effect between treatment 9/1 and 9/5 and group size 5 for ...					
... grade 4 ^J	0.156	0.103	0.149	-0.023	0.108
... grade 6 ^J	0.116	0.087	0.109	-0.019	0.065
... grade 8 ^J	0.080	0.067	0.074	-0.016	0.026
... grade 10 ^J	0.051	0.048	0.046	-0.013	-0.009
... grade 12 ^J	0.031	0.032	0.027	-0.010	-0.039
H0: No treatment effect between treatment 9/1 and 13/1 and group size 5 for ...					
... grade 4 ^K	0.019	0.004	0.021	-0.018	0.022
... grade 6 ^K	0.050	0.036	0.050	0.004	0.034
... grade 8 ^K	0.069	0.053	0.068	0.024	0.046
... grade 10 ^K	0.080**	0.058	0.077**	0.040	0.058
... grade 12 ^K	0.083*	0.055	0.078*	0.053	0.071
H0: No treatment effect between treatment 9/5 and 13/1 and group size 5 for ...					
... grade 4 ^L	-0.137	-0.099	-0.129	0.005	-0.086
... grade 6 ^L	-0.066	-0.051	-0.059	0.024	-0.031
... grade 8 ^L	-0.011	-0.015	-0.006	0.040	0.020
... grade 10 ^L	0.028	0.009	0.030	0.053	0.067
... grade 12 ^L	0.053	0.023	0.051	0.062	0.109
H0: No large group effect in treatment 9/1 for ...					
... grade 4 ^M	-0.007	0.006	-0.027	-0.113	-0.153*
... grade 6 ^M	-0.043	-0.037	-0.051	-0.125***	-0.143**
... grade 8 ^M	-0.067**	-0.063	-0.065**	-0.133***	-0.132***
... grade 10 ^M	-0.081***	-0.074**	-0.070**	-0.138***	-0.122**
... grade 12 ^M	-0.088**	-0.076*	-0.068*	-0.140***	-0.112
H0: No large group effect in treatment 9/5 for ...					
... grade 4 ^N	0.081	0.085	0.050	-0.297***	-0.389***
... grade 6 ^N	0.020	0.021	0.003	-0.305***	-0.373***
... grade 8 ^N	-0.027	-0.029	-0.030	-0.309***	-0.349***
... grade 10 ^N	-0.059	-0.063	-0.050	-0.309***	-0.320***
... grade 12 ^N	-0.076	-0.083	-0.057	-0.305***	-0.286***
H0: No large group effect in treatment 13/1 for ...					
... grade 4 ^O	-0.118	-0.147	-0.148*	-0.365***	-0.395***

... grade 6 ^O	-0.175***	-0.205***	-0.188***	-0.381***	-0.388***
... grade 8 ^O	-0.232***	-0.260***	-0.227***	-0.395***	-0.381***
... grade 10 ^O	-0.291***	-0.312***	-0.265***	-0.409***	-0.373***
... grade 12 ^O	-0.349***	-0.360***	-0.302***	-0.421***	-0.365***
H0: No expected coordination effect for ...					
... grade 4 ^P				0.280***	
... grade 6 ^P				0.369***	
... grade 8 ^P				0.457***	
... grade 10 ^P				0.540***	
... grade 12 ^P				0.615***	

***/**/* Denotes significance at the 1%/5%/10% level. P-values corrected for multiple hypotheses testing by performing a Bonferroni correction. Note that standard software does not correctly calculate marginal effects and p-values of interaction terms in nonlinear models (see Ai and Norton, 2003). We corrected for this.

^A ($\beta_{\text{grade}}=0$)

^B ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 9/5 = 0$)

^C ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 13/1 = 0$)

^D ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{group size } 5 = 0$)

^E ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 9/5 + \beta_{\text{grade}} * \text{group size } 5 = 0$)

^F ($\beta_{\text{grade}} + \beta_{\text{grade}} * \text{treatment } 13/1 + \beta_{\text{grade}} * \text{group size } 5 = 0$)

^G ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)

^H ($\beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)

^I ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] = \beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}]$)

^J ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0$)

^K ($\beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5 = 0$)

^L ($\beta_{\text{treatment } 9/5} + \beta_{\text{grade}} * \text{treatment } 9/5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0 = \beta_{\text{treatment } 13/1} + \beta_{\text{grade}} * \text{treatment } 13/1 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5$)

^M ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)

^N ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 9/5} * \text{group size } 5 = 0$)

^O ($\beta_{\text{group size } 5} + \beta_{\text{grade}} * \text{group size } 5 * 1[2/3/4/5 \dots \text{depending on the grade}] + \beta_{\text{treatment } 13/1} * \text{group size } 5 = 0$)

^P ($\beta_{\text{expected coordination}} + \beta_{\text{grade}} * \text{expected coordination} * 1[2/3/4/5 \dots \text{depending on the grade}] = 0$)

Table A5: Parametric and Non-parametric test results for Figure 5, Figure 6, and Figures A3 – A6

	(1) successful coordination	(2) inefficient coordination	(3) efficient coordination	(4) belief correct	(5) belief correct LOW	(6) belief correct HIGH
H0: No age effect for ...						
... treatment 9/1 and group size 2 ^A	***	***		***	***	***
... treatment 9/5 and group size 2 ^A	***	**			**	
... treatment 13/1 and group size 2 ^A		***	***		***	***
... treatment 9/1 and group size 5 ^A						
... treatment 9/5 and group size 5 ^A	***	***		**	***	
... treatment 13/1 and group size 5 ^A				**		
H0: No treatment effect between treatment 9/1 and 9/5 and group size 2 for ...						
... grade 4 ^B						
... grade 6 ^B	*	**	*		***	***
... grade 8 ^B					***	***
... grade 10 ^B				***	***	***
... grade 12 ^B			**		***	***
H0: No treatment effect between treatment 9/1 and 13/1 and group size 2 for ...						
... grade 4 ^B						
... grade 6 ^B	***	***	***		***	***
... grade 8 ^B		***	***		***	***
... grade 10 ^B	***	***	***	***	***	***
... grade 12 ^B	***	***	***	***	***	***
H0: No treatment effect between treatment 9/5 and 13/1 and group size 2 for ...						
... grade 4 ^B						
... grade 6 ^B					***	***
... grade 8 ^B		***	***		***	***
... grade 10 ^B	***	***	***		***	***
... grade 12 ^B	***	***	***	**	***	***
H0: No treatment effect between treatment 9/1 and 9/5 and group size 5 for ...						
... grade 4 ^B					***	
... grade 6 ^B						
... grade 8 ^B					**	
... grade 10 ^B						
... grade 12 ^B						
H0: No treatment effect between treatment 9/1 and 13/1 and group size 5 for ...						
... grade 4 ^B						

... grade 6 ^B						***
... grade 8 ^B						***
... grade 10 ^B						
... grade 12 ^B						
H0: No treatment effect between treatment 9/5 and 13/1 and group size 5 for ...						
... grade 4 ^B						***
... grade 6 ^B						**
... grade 8 ^B						***
... grade 10 ^B						
... grade 12 ^B						
H0: No large group effect in treatment 9/1 for ...						
... grade 4 ^B	***	***	***	*	***	***
... grade 6 ^B	***	***	***	**	***	***
... grade 8 ^B	***	***	***	***	***	***
... grade 10 ^B	***	***	**	***	***	***
... grade 12 ^B	***	***			***	***
H0: No large group effect in treatment 9/5 for ...						
... grade 4 ^B				**	***	**
... grade 6 ^B	***	***	***	***	***	***
... grade 8 ^B	***	***	***	***	***	***
... grade 10 ^B	*	*	*	***	***	***
... grade 12 ^B	***	***	*	*	***	***
H0: No large group effect in treatment 13/1 for ...						
... grade 4 ^B	***	***	**		***	***
... grade 6 ^B	***	***	***	***	***	***
... grade 8 ^B	***	***	***	***	***	***
... grade 10 ^B	***	***	***	***	***	***
... grade 12 ^B	***	***	***	***	***	***

^A Cuzick's Wilcoxon-type test for trend

^B T-test

Experimental instructions (for “9/1” treatment)¹¹

Welcome to our game. Before we start, we will explain the rules of our game. From now on, please don't talk to your neighbor and listen carefully. We will frequently stop during our explanation and allow you to ask questions. Therefore, please raise your hand and one of us will come to you to answer your question. You can earn money in this game. You will receive your payments in an envelope marked with your ID within 2 weeks.

Your payment is denoted in points. At the end of the session we will convert these points into Euros. We apply the following conversion rate:

1 point = 0.3 € (30 cents)¹²

You will receive a conversion table which will tell you how much your earned points are worth.

(Carefully explain the conversion table with a few examples.)

Everybody ok so far? *Leave time for questions and answer them privately.*

The game we will play consists of two parts but only one part will be paid out. At the end of the game we will randomly determine which of the two parts will be paid out. We do so by shuffling these two cards numbered with one and two and drawing one card blindly. The drawn card determines the part to be paid out.

For this game we will build random groups. You will play one part in a 2-person group and one part in a 5-person group. Your partner(s) are at a different school but attend the same grade level as you do. The screen will tell you whether you're playing in a 2- or 5-person group.¹³

Everybody ok so far? *Leave time for questions and answer them privately.*

In this game you have to pick a number (either 1 or 2). Each group member will decide independently of each other. Your payment will depend on your decision and on the decision(s) made by your partner(s). Your payment can be inferred from a table.

For example:

	Your partner's number	
	TWO	ONE
Your number	TWO 9	ONE 1
	ONE -	8

You have 1 partner!

¹¹ The other two treatments are analogous by changing the respective payoffs.

¹² Conversion rates were age dependent: €0.3 per point for 9-10-year-olds, €0.5 for 11-12-year-olds, €0.7 for 13-14-year-olds, €1.4 for 15-16-year-olds and €2.2 for 17-18-year-olds.

¹³ Here we present the order in which the 2-person game was explained first and the 5-person game second. This was balanced in other classes where we explained the 5-person game first and 2-person game second.

Playing with one partner:

If you choose 1, you will receive 8 points for sure. If you choose 2, your payment will depend on your partner's choice. If your partner chooses 2 as well, you will receive 9 points. If your partner chooses 1, then you will receive 1 point.

Everybody ok so far? *Leave time for questions and answer them privately.*

	Lowest number of all 4 partners	
	TWO	ONE
Your number	TWO 9	ONE 1
	ONE -	8

You have 4 partners!

Playing with four partners:

If you choose 1, you will receive 8 points for sure. If you choose 2, your payment will depend on your partners' choices. Specifically, the minimum number (lowest number) chosen by your four partners will count. If all four partners choose 2 (and you choose 2 as well), then you will receive 9 points. If one or more of your partners choose 1 (and you choose 2), then you will receive 1 point.

Everybody ok so far? *Leave time for questions and answer them privately.*

After completing these two parts, please answer an estimation-question which will appear on your computer screen. If you answer the estimation-question correctly, you will receive one additional point.

Before we start, you will have to answer two control questions, which we will check before we start playing the game.

Everybody ok so far? *Leave time for questions and answer them privately.*

If nobody has any further questions, please start the experiment.

Control questions to assess the participant's comprehension

Assume that you have chosen **ONE**.

How many points do you earn?

- This does not depend on my decision.
- 1 point
- 5 points
- 8 points
- 9 points
- 13 points

Assume that you have chosen **TWO**.

How many points do you earn if all other person also have chosen TWO?

- 1 point
- 5 points
- 8 points
- 9 points
- 13 points