If You Could Read My Mind – An Experimental Beauty-Contest Game With Children

BY Henning Hermes & Daniel Schunk

DISCUSSION PAPER

NHH

Institutt for samfunnsøkonomi
Department of Economics

SAM 23/2019

ISSN: 0804-6824

March 2021

This series consists of papers with limited circulation, intended to stimulate discussion.

If You Could Read My Mind—An Experimental Beauty-Contest Game with Children

Henning Hermes*[1] and Daniel Schunk[2]

[1]NHH Bergen, FAIR / Department of Economics
[2]University of Mainz, Department of Economics

March 2021

Abstract

We develop a new design for the experimental beauty-contest game (BCG) that is suitable for children in school age and test it with 114 schoolchildren aged 9–11 years as well as with adults. In addition, we collect a measure for cognitive skills to link these abilities with successful performance in the game. Results demonstrate that children can successfully understand and play a BCG. Choices start at a slightly higher level than those of adults but learning over time and depth of reasoning are largely comparable with the results of studies run with adults. Cognitive skills, measured as fluid IQ, are predictive only of whether children choose weakly dominated strategies but are neither associated with lower choices in the first round nor with successful performance in the BCG. In the implementation of our new design of the BCG with adults we find results largely in line with behavior in the classical BCG. Our new design for the experimental BCG allows to study the development of strategic interaction skills starting already in school age.

JEL classifications: C72, C92

Keywords: children, experimental beauty-contest game, guessing game, strategic interaction, decision-making, cognitive skills, noncognitive skills

Acknowledgments

We are grateful to the editor and two anonymous referees for very helpful feedback. In addition, we would like to thank our colleague Aleksa Kaurin for teaming up in the data collection as well as Helen Chairopoulou, Kristina Jakob, Helena Rohlik, Marion Schmale, and Vivien Voigt for excellent research assistance. We also thank Eirik Berger, Matthias Heinz, Fanny Landaud, Matthias Sutter, Bertil Tungodden, Justin Valasek, and Constantin Weiser for helpful comments as well as Rosemarie Nagel for providing raw data from her experiment. In addition, the paper benefited from feedback at the ESA Conference 2019 in Abu Dhabi, the Workshop on Microeconomics in Lueneburg, the EEG seminar at MPI Bonn, and NCBEE 2019 in Kiel. This project was generously supported in part by the Jacobs Foundation and the research unit Interdisciplinary Public Policy at the University of Mainz (IPP Mainz). Henning Hermes gratefully acknowledges financial support from the German National Academic Foundation (Studienstiftung des deutschen Volkes) and the Research Council of Norway through its Centers of Excellence Scheme, FAIR (projects No 262675, 262636, and 250170F10).

*Corresponding author: henning.hermes@nhh.no; NHH Bergen, Helleveien 30, 5045 Bergen, Norway

1 Introduction

An important skill for economic actors is the ability to anticipate the actions of others in strategic settings and to choose one's own actions accordingly. There is a very wide range of situations in which these strategic interaction skills are crucial.[1] For example, in their well-known study on signaling games, Cooper and Kagel (2005) argue, based on recording and coding of the dialogues of experimental participants, that "a critical step in monopolists' learning to play strategically is putting themselves in the entrant's shoes, reasoning from the entrant's point of view to infer likely responses to their choice as a monopolist." Similarly, empirical evidence demonstrates that individual investors base their investment decisions on their beliefs about the return expectations of *other* investors (Rangvid, Schmeling, and Schrimpf 2013; Egan, Merkle, and Weber 2014). Another example are most forms of matching markets; specifically, Braun et al. (2014) show this in the context of a laboratory study that investigates students' behavior in university admissions procedures. Finally, the dynamics and outcomes of weakest-link games, representing, for example, coordination problems in a firm, depend on the players' ability to anticipate the actions of their peers (e.g., Brandts and Cooper 2006).

But why studying these skills in children? In the last two decades, a large number of studies (summarized in Kautz et al. 2014) has documented that preferences and skills are shaped in the early years, they form the basis for future investments, and fundamentally determine adult life outcomes. For this reason, the early development of preferences and skills has been studied in great detail in recent years in the economic literature (see Sutter, Zoller, and Glätzle-Rützler 2019, for a review). By contrast, children's strategic interaction skills have been studied much less, despite being of similar importance in their own right as well as for the life cycle of skills. The study of strategic interaction skills among children is inherently difficult, reflecting Jean Piaget's view that up to a certain age children systematically lack the ability to adopt other's perspectives (Piaget 1962). Hence, most existing work on this topic has so far considered only settings, in which children interact either with a single peer child or with a computer (see the literature review below). However, this disregards the fact that most strategic interactions outside the laboratory involve several (and often large numbers of) peers, and thus require a more comprehensive notion of strategic sophistication.

[1] Brocas and Carrillo (2018b) define strategic interaction to be the intrinsic ability to anticipate the actions of others and to act accordingly.

Here, we develop a novel design to study strategic interaction skills in children, which is based on the experimental beauty-contest game (henceforth, BCG) introduced by Nagel (1995, at that point called "guessing game"): N decision-makers simultaneously choose a number x between 0 and 100 and the person with the number closest to $p * \bar{x}$ wins a fixed prize (mostly, $p = 2/3$). This economic game has been used to study strategic interaction in groups for more than two decades. The experimental BCG has a strong advantage compared with most other interactive games used in economics to study decision-making (e.g., ultimatum game, public good game, etc.): Neither social, nor time or risk preferences should affect decision-making (cf. Kocher and Sutter 2005).[2] A substantial body of knowledge has been developed concerning how various game parameters such as repetition, feedback, or time pressure affect performance in a BCG (Duffy and Nagel 1997; Ho, Camerer, and Weigelt 1998; Weber 2003; Kocher and Sutter 2006) and how individuals vs. teams of different sizes behave (Kocher and Sutter 2005; Sutter 2005). Yet, there is no study using a BCG to measure strategic interaction skills in groups with children. Given the relevance of strategic interaction skills for economic decision-making (see above), having such a measure of strategic interaction skills for children would not only be interesting in itself but would also enable us to shed light on which other skills (cognitive or noncognitive) are the building blocks to developing strong strategic interaction skills over the life cycle.

To achieve this goal, we simplify the experimental BCG into a board game—we make it less abstract, provide concrete and visually illustrated operations with a spatial interpretation corresponding to each step in the game, and use the median, integers (0–100), $p = 1/2$, and only five players per group. Applying this new design for the experimental BCG, we study the behavior of $n = 114$ children aged 9–11 years in the game and demonstrate that they are capable of successfully playing an experimental BCG. We also validate the new design using an adult student sample with $n = 120$. In a second step, aimed at understanding the building blocks of children's strategic interaction skills, we analyze the link between children's performance in this game and their fluid IQ, an important part of cognitive skills. Our results demonstrate that fluid IQ is predictive only of whether children choose weakly dominated strategies, but is not associated with lower choices in the first round or with successful performance in the subsequent BCG. Hence, we contribute to the literature on strategic interaction in two ways, namely by (i) developing a tool to study strategic interaction in groups with young children and adolescents, and (ii) analyzing the role of cognitive skills for strategic interaction in a less abstract game form.

[2] To keep instructions simple, in our design children receive a payoff in *every* round of the BCG (see Section 2.3 for details). In this setting, social preferences actually could affect decision making, for example, because children might be inequality averse.

Our study is related to a literature on strategic interaction in children. Some early studies (starting with Murnighan and Saxon 1998; Harbaugh and Krause 2000; Harbaugh et al. 2003; Harbaugh, Krause, and Liday 2003) have examined children's behavior in simple interactive games, such as dictator, ultimatum, trust, and public good games. However, there is not much research about more complex strategic interactions in children, i.e., settings in which the cognitive demand of the task is higher and in which it matters to what extent children are able to take the perspective of others and translate this into their own strategic decision-making.[3]

Among the few notable exceptions are the following studies: First, there is a study on strategic interaction that involves observing child-experimenter interactions: Sher, Koenig, and Rustichini (2014) have children play two games, a sticker game, and a sender-receiver game. They argue that for successful strategic interaction, children require the ability to undertake recursive thinking, i.e., "the ability to use the output of one step of a reasoning process as input to a following step" as well as the ability to put themselves into another person's shoes, i.e., to understand another person's motives and emotions ("theory of mind") as well as their incentives (again, cf. Cooper and Kagel 2005). In their study, many children possess these skills from about the age of 7 years onward. Second, there is a number of two-person games, i.e., involving the interactions between two children: Brocas and Carrillo (2018a) examine iterative reasoning among children from pre-kindergarten to first grade in the context of four two-person games. They show that both logical thinking and theory of mind are key ingredients to successful strategic reasoning, and trace children's limitations in both these skills in their age groups. In another series of two-person games, Brocas and Carrillo (2018b) present further support for these findings and show that although preschoolers are able to think strategically in principle, this does not mean that they are capable of acting accordingly. Brocas and Carrillo (2020) play a two-person BCG with children and adolescents from 5 to 18 years. Their findings indicate that strategic sophistication is present early on and, while abstract reasoning is known to develop through adolescence, equilibrium play in their game improves through childhood and stabilizes already after age 10. The detected learning trajectory is mostly due to feedback-learning, i.e. observing the choice of the partner. Fe, Gill, and Prowse (2019) play a competitive game designed to trigger level-k thinking with pairs of children aged between 5 and 12 and find that theory-of-mind, cognitive ability, and age are related to the children's level-k-

[3] We should mention that there is a literature in developmental psychology that studies the extent to which children are able to take the perspectives of others (see Birch et al. 2017, for a summary). This literature focuses on how children reason about other children's motives and decisions, but not on the strategic interactions of the children.

behavior. Furthermore, Czermak et al. (2016) present 10–17-year-old children with simple experimental normal-form games for two children and elicit their beliefs, providing evidence that children of this age are clearly able to strategize in their choices. Third, there is one study that involves the interaction between three children: Brocas and Carrillo (2019) use a graphical paradigm of a dominance solvable game to study the evolution of the ability to think strategically between 8 years and adulthood. They report significant performance increases between 8 and 12 years and a stabilization afterwards.

While these above-mentioned studies are related to our approach, our experimental setting involves the interaction with a *group* of other children in the context of a normal multi-player BCG. That is, the game requires a substantive degree of logical thinking and perspective-taking. Research by Brosig-Koch, Heinrich, and Helbach (2015) has demonstrated that children at the age of 6 years already have the ability to reason backwards and that this ability increases with age. Thus, we expect that children in our sample (aged 9–11 years) will possess the necessary cognitive skills.[4] Moreover, empirical research using theory-of-mind tasks confirms that, by the end of preschool, most children should be well able to take the perspective of others (Wellman, Cross, and Watson 2001; Brocas and Carrillo 2018a); hence, we expect children to also possess sufficient perspective-taking abilities. Nevertheless, strategic interaction in the form of an experimental BCG is very challenging, even for adults. Grosskopf and Nagel (2008) conducted a two-player version of the BCG, which can be easily translated into a “whoever chooses the smaller number wins” contest. They found that only 10% of (adult) lab participants and 37% of professionals (i.e., participants at economic conferences) actually chose zero—in this form of the game, zero is the dominant strategy, irrespective of the other player’s choice (means were 35.6 and 21.7, respectively). Similarly, Bosch-Rosa and Meissner (2020) use one player guessing games with adults to show that even with this simple structure, many subjects fail to fully understand the structure of the game. The findings of our study suggest that—despite this challenging setting—children are able to understand the experimental BCG if it is presented in a visual manner and we use this new design to draw conclusions on the determinants of strategic interaction behavior in children.

The remainder of this paper is structured as follows: Section 2 describes the experimental procedures and the new design of the experimental BCG used in the present study, Section 3 presents the results and discussion, and Section 4 concludes.

[4] Castillo, Jordan, and Petrie (2018) show with a sample of 7th–9th graders that although many observed choices are not completely consistent with predictions from economic theory, they are not random either. They interpret this as an indication of rational behavior, which further supports that children of our age group should have the necessary level of rational thinking.

2 Experimental Design

The study was conducted in schools with children aged 9–11 years (third and fifth grade). The following subsections provide details on the procedures, participants, the new design of the BCG, and the outcomes measured.

2.1 Procedures

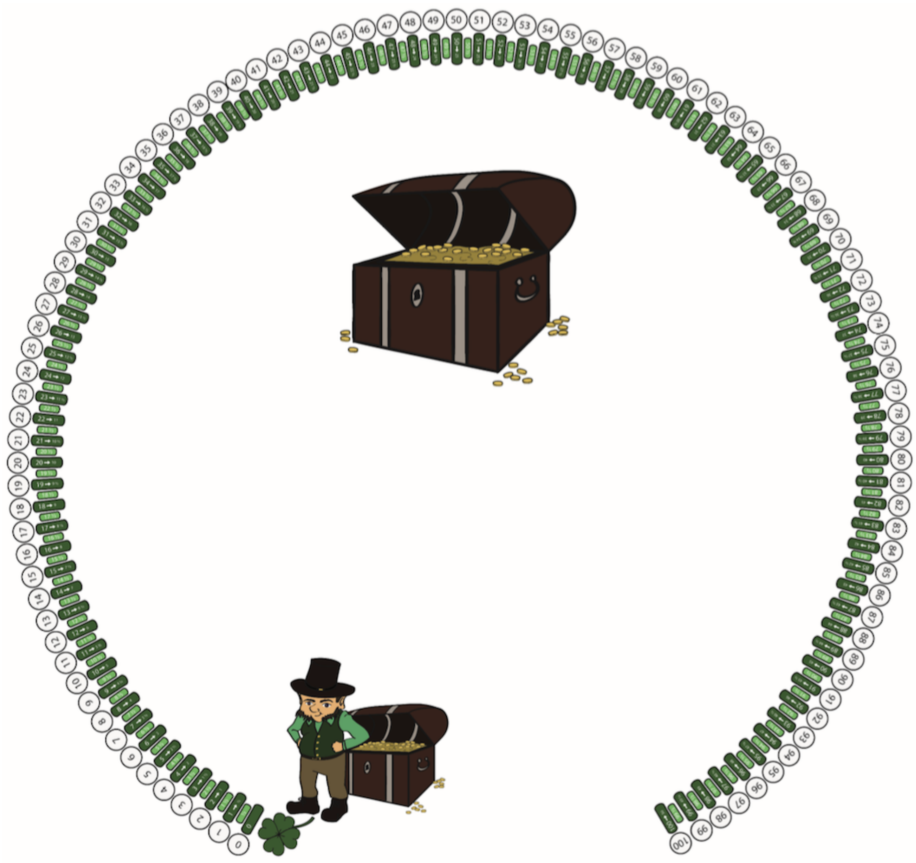
We conducted the study in March and April 2016 in three different schools in Germany. We contacted the three schools, asked for their participation (all agreed), and sent parental consent forms for data use to the teachers of the participating classes (we were not involved in the selection of classes). Six classes were tested, each within one day. Days were always structured in the same way. In the first lesson (45 minutes), the experimenter informed the children about the procedures of the day, including the incentives they could win by earning gold coins. Then, the experimenter guided the class through a workbook to collect data on a number of control outcomes using various questionnaires and tests. All questionnaires were read out loud to avoid problems with different levels of reading skills. In the subsequent lessons, we took randomly formed groups of five children out of the class and played the BCG with them in a separate room (again, each lesson lasted about 45 minutes). Groups were formed based on numbers on children's workbooks which were randomly distributed in the classroom in the first lesson. After the last lesson, children went to a separate room, where they could trade the gold coins they won during the game for toys.

2.2 Participants

Six classes in three different schools participated in the study; four classes were in the third grade (in total 77 children), two were in the fifth grade (in total 50 children). Overall, 113 of these children plus two children from another class played the BCG (a total of 115 children). The parents of one of these children did not provide consent for data use, resulting in a final sample of 114 children.[5] Children played the BCG in 23 groups of five children. Each group consisted of randomly selected children from one single class, with two exceptions: in one case, a child from class A played together with four children from class B because this child had not been able to play the BCG the day before due to time constraints. In the other case, we recruited two children from another class to ensure that we had a complete group of five

[5] We make no claim about the representativeness of our sample. Although we achieve a high participation rate (113 out of 127 children), it is possible that there is selective participation on the individual and/or the school level.

Figure 1: Design of the Beauty-Contest Board Game ("Goblin Game")



children playing the BCG. In the final sample, 63 children were female (55%), 64 were in the third grade (56%), and the mean age was 10 years (SD = 1, see Table A1 for details).

2.3 The Beauty-Contest Game

At the heart of the study was an incentivized experimental BCG in a new design. To facilitate understanding of the game, we (i) used a board game setup, so children had a visualization as well as a spatial mapping of the game in front of them, (ii) embedded the rules in a story-like design that was easy to recall and very simply structured, and (iii) ensured the highest possible degree of understanding by explaining the game in a one-to-one setting between each child and a trained experimenter, with each child asked to explain each step of the game back to the experimenter. We also simplified the game parameters to make the game easier to understand. Specifically, we used $p = 1/2$, the median (instead of the mean), five players (a group size for which strategic environment effects are triggered, see Hanaki et al. 2019), and integers in the range from 0 to 100.

The Board Game. We developed a board game mirroring the experimental BCG called the "Goblin Game" (see Figure 1). The board depicts the range of integers from 0 to 100 (in black and white) as the range of numbers from which to choose, and a second range of numbers (in green) from 0 to 100 in steps of $1/2$ to indicate the winner. In the center of the board, there was an open treasure chest containing gold coins (the prize that children could win in each round of the game). At zero, the goblin had his starting position (as he would always start from zero and would then "walk up" to the median player). Numbers were arranged in a circle to create an illusion of "equality of numbers", i.e., there was no "best" or "worst" number.

Children played the BCG at a large table in a separate room exclusively used for the study on the day of data collection. Each child was randomly assigned to one of the colors—yellow, blue, orange, white, or gray—and received a game piece (henceforth called a "pawn") in their respective color.[6] The color also determined each child's seating position at the table. One experimenter sat at the head of the table and led the game (see Figures A4–A6 in the Appendix for details on the experimental setup and the material used in the game).

The Rules of the Game. In order to make the experimental BCG as easy to understand as possible, we structured the rules of the game into a five-step procedure that was repeated for every round played. Each group played 10 rounds. The five steps within each round were as follows.

1. All players secretly write down a number between 0 and 100 on their game slip.
2. All players simultaneously place their pawn onto the number from step 1 on the board.
3. The goblin starts from zero and walks up to the third, i.e., the "middle" player.
4. Having reached the third player, the goblin jumps back by half the distance he has walked.
5. The player who is now closest to the goblin wins a gold coin.

No communication between children was allowed during the 10 rounds of the game. As practical measures, we told children to indicate that they had made their decision in step 1 by covering the number they had written on their game slip with their pawn. In this way, we avoided children writing or changing numbers after having seen choices of their peers. Of course, we also emphasized (and ensured) that in step 2 children had to put their pawn on the actual number they wrote down on their game slip. To help children avoid confusion between the numbers they chose themselves and the numbers on which the goblin was operating, we color-coded the two arrays of numbers: children used the larger, black-and-white numbers, whereas the (green) goblin used the green, inner circle of

[6] We tried to avoid using colors with strong meanings, such as red, green, pink, etc.

numbers. For children having problems calculating what "jumping back by half" meant, we printed the number that the goblin would jump to on each of the green fields (e.g., the green field for 30 would read "30 → 15"). For simplicity and because we could not split the gold coins, if several players were equally close to the goblin, each of them would receive a coin.[7]

At the same time, we embedded the rules in a story-like design. Children learned that the goblin had "hidden gold coins in the forest" but that he would be willing to help children find the coins and would "reveal the location of the gold coins to the middle player". Also, children were told that the goblin was "hexed"—for each step he would go forward, he would have to take half a step backward. At the end of each round, the goblin would give a gold coin to the player who was closest to him (see Section F.2 in the Appendix for the exact instructions).

Explaining the Rules. As this study was the first involving an experimental BCG played with children, a major challenge was to ensure that the game was properly understood by all children. Consequently, we had five trained experimenters present in each group to (i) ensure the best opportunities for each child to understand the game and (ii) to check the individual understanding of each child. Hence, every child individually received a (standardized) explanation from an experimenter and could ask the experimenter any questions that he or she had regarding the game. In this part of the lesson, each child sat at an individual small table distributed around the classroom (cf. Figure A6 in the Appendix). To explain the game, the experimenter used a written instruction, which he/she read to the child, a small version of the board game, and the relevant materials (coins and pawns, see Figure A5 in the Appendix for the setup of an individual table).

At the beginning of the explanation, children were told that they would now play a game in which they could earn gold coins. These coins could later be exchanged for a real toy and the more coins they earned, the larger would be the selection of toys from which they could choose. Children were encouraged to "try hard" to earn as many coins as possible. By starting the game this way, we wanted to (i) motivate the children and make incentives salient, and (ii) make sure that the children understood that they could actually "do something" about the number of coins won and that the game was by no means pure luck. Subsequently, the rules of the game were explained to the children. Then, they could ask questions in case anything was not clear. Finally, the experimenter asked each child to explain back to him/her each of the five steps of the game, one by one. For each

[7] We should point out that, in principle, this game feature could imply an incentive for children to coordinate. However, the reason for this design choice was to keep the game design as simple as possible.

step, the experimenter rated the child's understanding (see Section 2.4). Afterwards, the experimenter expressed appreciation for the child's explanations and "awarded" him or her a first gold coin (this was done to avoid children having zero coins while playing the game). At the end of the individual explanation session, the experimenter conducted a very short task with the child (see Section D in the Appendix) and subsequently sent him or her to the large table in the middle of the classroom to play the "Goblin Game" with his or her classmates.

All research assistants acting as experimenters participated in a half-day training session, which included information on general procedures, a discussion of the importance of standardized instructions, how to deal with potential questions by children, and how to organize the data collection. General hints for experimenters included not reacting to strategies or suggestions articulated by children on how to win the game, and to answer questions only by referring to the general rules of the game. Experimenters were explicitly instructed not to mention any potential strategies to the children. The exact instructions read out to each child can be found in Section F.2 in the Appendix.

Playing the Game. Once all children had finished their instructions with their individual experimenter, they sat down at the large table in the middle of the classroom (see Figure A4). Each seat was clearly marked with the color of each pawn; hence, seating positions were exogenous. Before starting the first round, the main experimenter once again repeated the five steps of the game. He then guided the children through all 10 rounds of the game, each time structured by the five steps. At the end of each round, the experimenter gave a gold coin to the winning child. After the last round, individual workbooks and game slips (containing the numbers for each round) were collected and each child received another gold coin for his or her workbook. To avoid any communication about the game between children who already played the game with children who had not yet played the game, we then explicitly instructed children not to tell their peers anything about the game. Finally, children were brought back to their classroom and the next group was picked up to play the BCG in a separate room.

2.4 Outcomes and Data

To account for children's choices and performance during the BCG, we collected the individual workbooks and game slips filled out during the game. In addition, we collected the experimenters' ratings on children's understanding of the game (see previous Section). To analyze to what extent children's cognitive skills are determinants of successful performance in strategic interaction games, we elicited children's fluid IQ. We used a well-established measure feasible for children, namely Raven's "Coloured Progressive Matrices" (Bulheller and Häcker 2010). Based on the collected data, we constructed the following outcome and control variables:

Numbers Chosen. We used the numbers each child wrote down on his or her individual game slip during the BCG.[8]

Number of Coins Won. A straightforward measure of successful performance in the game is the number of coins each child won during the game. This is equivalent to the number of rounds (out of 10) that a child won. Note that in the case of a tie, more than one child could win a coin. Thus, the total number of coins per group could be larger than 10.

Distance to Best Response. The number of coins won may not fully reflect the degree to which a child made successful choices during the game. For instance, a child could miss out on winning by only one step, whereas another child could miss out on winning by 30 steps, yet they would be equal with respect to the number of coins they won. Therefore, to analyze successful choices beyond the simple indicator of *winning* a round, we constructed the following outcome measure. We computed the best response for each player in each round (i.e., given the other players' choices in this round, the number that would be the ideal response to win this round—this number is equivalent to half of the second-smallest number of the other four players). We then calculate the distance between the player's actual choice and this best response, take the absolute value, and divide the distance by half the median number in this group in this round to account for scaling differences over the course of 10 rounds (if the median number was zero, we divide by one). Therefore, for each player, the measure of distance reports how far away (measured in fractions of half the median in this round) a player was from his or her ideal response to the other four players' choices.

Rank Based on Distance. Because the "distance to best response" measure is difficult to compare across (i) groups and (ii) rounds (despite the normalization with half the median in each round), we rank players within each group within each round based on their distance to the best response (i.e., the child with the shortest distance (the winning child) receives

[8] To double-check, one of the other experimenters wrote down the numbers for all children for all rounds during the game. In a very few cases (11 out of 1140), we used the number provided by the experimenter, mostly because of handwriting issues.

rank 1, the child who is closest to winning receives rank 2, and so on). Then, we can compute the average ranks across rounds. Note that better choices lead to *lower* average ranks.

Gender and Age. Gender is indicated by a dummy that is equal to one if the child is female. Age is measured in dimensions of years, i.e., increasing age by one implies an increase in age of one year (still, age is measured referring to the exact date of birth of each child).

Fluid IQ. To proxy children's cognitive skills, we measured fluid IQ with 18 out of 36 available items (plus two example items to explain the test) from Raven's Colored Progressive Matrix test (Bulheller and Häcker 2010). Each correctly solved item was weighted equally. The distribution of raw scores is shown in Figure A1 in the Appendix. To enable comparison of effect sizes, the variable was standardized to mean = 0 and SD = 1.

Understanding of the Game. After the instructions, each child had to explain the game back to the experimenter (see Section 2.3). The experimenter rated understanding for each of the five steps of the game on a 4-point scale, ranging from 4 = "Immediately completely and correctly explained" to 1 = "Not completely understood" (see Section F.2 in the Appendix). Understanding is calculated using the sum of the five steps, with a maximum value of 20.

We also collected some other measures that were not part of the present study. Finally, we checked whether parental consent for data use was available (which was the case for 114 out of 115 children), and analyzed the data using the statistical software Stata/SE 15.

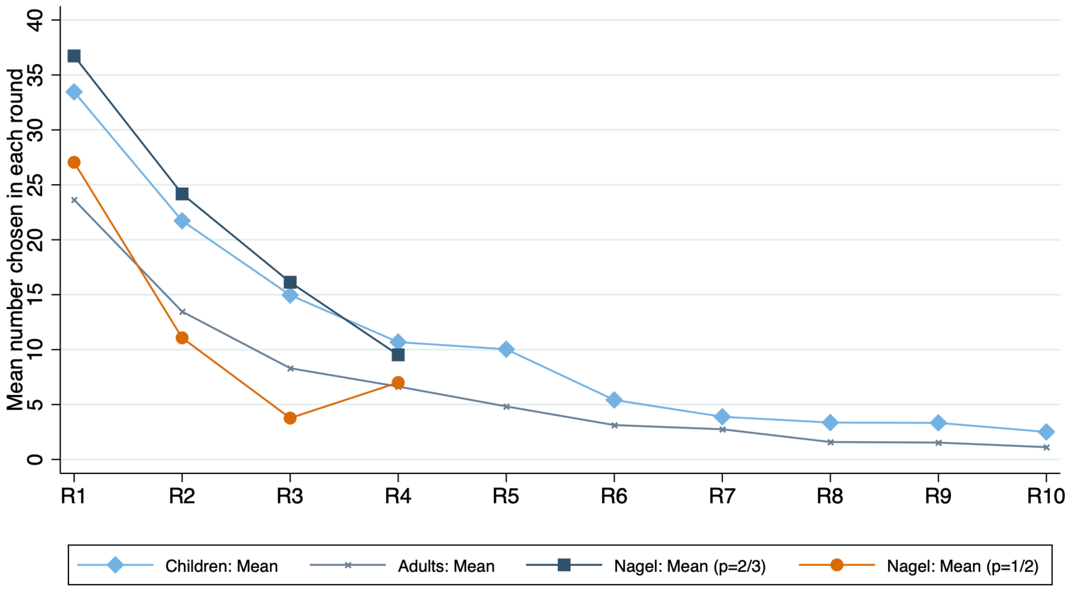
3 Results & Discussion

We first provide descriptive evidence to support that children successfully understood and played a strategic interaction game in the form of an experimental BCG. Then, we turn to the results regarding important determinants of performance in the BCG and discuss our findings. Finally, we report the results from our replication study with an adult sample.

3.1 Descriptive Results

As noted, we had 114 participants in our sample, 63 of whom were female. Distributions over the classes and grades as well as summary statistics for fluid IQ and understanding of the game can be found in Table A1 in the Appendix.

Figure 2: Development of Mean Numbers Chosen over the Rounds



Notes: This figure plots mean numbers chosen in each round for the present study (children and adult samples) compared with Nagel (1995), using averages of sessions with $p = 2/3$ and $p = 1/2$.

To assess how well children understood the game, each child had to explain the five steps of the game to an experimenter. For each step, a child's level of understanding was rated by the experimenter with a maximum of four points, i.e., a child could achieve at most 20 points (see Section 2.3). Overall, 90 children (80%) were rated with 19 or 20 points, only five children (4%) received less than 17 points (see Table A1 in the Appendix).[9] Hence, experimenter ratings indicate a very high level of understanding of the game.

To analyze children's choices in the BCG, we compare the results from the present study with other data from experimental BCG studies. First, in Figure 2 we plot mean choices (for median choices, see Tables 1 and A2 in the Appendix) over the 10 rounds and compare them with the choices from Nagel's seminal paper (1995, using sessions with $p = 2/3$ and sessions with $p = 1/2$) as well as with our own replication with an adult sample (for details on the replication with adults, see Section 3.4).[10] Generally, we find that the average number chosen by children is decreasing over the 10 rounds, with the bulk of the decrease occurring in rounds 1–6. Thus, children's choices seem to converge toward the game-theoretic equilibrium over time, a finding that is well established in other studies with adult samples (e.g., Nagel 1995; Duffy and Nagel 1997; Ho, Camerer, and Weigelt 1998).

[9] Excluding the children with ratings of understanding below 19 points does not affect our findings (see Section C in the Appendix).

[10] Note that for several reasons (see Section 2.3) we used the median, whereas Nagel (1995) used the mean to identify the winner in their experimental BCG.

Table 1: Comparison of Mean and Median Numbers with Nagel (1995)

	Children, $p = 1/2$		Adults, $p = 1/2$		Nagel, $p = 2/3$		Nagel, $p = 1/2$	
Round	Mean	Median	Mean	Median	Mean	Median	Mean	Median
1	33.5	28	23.6	21	36.7	33	27.1	17
2	21.7	20	13.5	11	24.2	18	11.1	7.25
3	15	11.5	8.3	7	16.1	12	3.8	3
4	10.7	8	6.6	4	9.5	8	7	0.755
RoD 1–4	0.66	0.69	0.72	0.75	0.74	0.76	0.76	0.95

Notes: Comparison of mean and median numbers chosen in our study (children and adult samples) with those in Nagel (1995); average values for sessions with $p = 2/3$ and $p = 1/2$, based on raw data provided by the author. RoD refers to the "Rate of Decrease" of the mean or median within a group over rounds 1–4. RoD = $(mean_{t=1} - mean_{t=4})/mean_{t=1}$, or the median, respectively. Reported rates of decrease are averages weighted by group size.

In addition, children's choices mimic both the level as well as the rate of decrease found in Nagel (1995)—however, the children sample is very close to the sessions with $p = 2/3$ in the Nagel study, indicating that children's choices start on a higher level (but develop with a similar rate of decrease). In other words, children seem to play a $p = 1/2$ BCG using the new design proposed in our study in a very similar manner to the way that adults play the classical $p = 2/3$ experimental BCG.[11] Table 1 reports the detailed values comparing the first four rounds of the game across the four samples. Mean and median numbers for the children are largely comparable with the numbers from Nagel (1995) in the sessions using $p = 2/3$ (testing children's choices in rounds 1–4 against choices in the corresponding rounds in the Nagel $p = 2/3$ sample reveals no significant difference, Mann-Whitney U tests for each round, all $p > .164$).[12]

Second, to further support the notion that children played the new design of the experimental BCG in a way that is largely comparable with adults, we benchmark our data with results on the average "depth of reasoning" from Duffy and Nagel (1997). We apply Duffy and Nagel's definition of the depth of reasoning by calculating the depth of reasoning d for individual i in group j for round t solving:

$$number_{i,t} = median_{j,t-1} * p^{d_{i,t}}$$

Table 2 reports the values for our children sample as well as our replication with an adult sample and the comparable numbers from Duffy and Nagel (1997). Note that for this comparison, the game parameters of both designs match exactly, i.e., the data from Duffy and Nagel (1997) are also based on $p = 1/2$ and the median (subjects in their study were undergraduate university students). Table 2 clearly shows that, in all three samples, the

[11] Also, the distribution of first-round choices is, by and large, in line with adult studies; see Figure A2 and, e.g., Nagel (1995) or Bosch-Domènech et al. (2002).

[12] Results for the adult sample are discussed in Section 3.4.

Table 2: Comparison of Depth of Reasoning with Duffy and Nagel (1997)

	$d > 3$	$d = 3$	$d = 2$	$d = 1$	$d = 0$	$d < 0$
Round	Children, $p = 1/2$, median					
1	.04	.02	.16	<u>.39</u>	.31	.08
2	.00	.01	.12	<u>.41</u>	.38	.08
3	.01	.03	.11	<u>.42</u>	.35	.08
4	.03	.02	.11	<u>.38</u>	.36	.11
	Adults, $p = 1/2$, median					
1	.04	.06	.27	<u>.50</u>	.13	.01
2	.02	.03	.22	<u>.43</u>	.23	.08
3	.03	.04	.20	<u>.34</u>	.28	.11
4	.06	.06	.15	<u>.37</u>	.20	.17
	Duffy & Nagel (1997), $p = 1/2$, median					
1	.07	.04	.30	<u>.43</u>	.11	.05
2	.02	.09	<u>.29</u>	<u>.29</u>	.16	.16
3	.02	.11	.32	<u>.38</u>	.09	.09
4	.07	.18	<u>.41</u>	.14	.05	.14

Notes: Comparison of the distribution of the average depth of reasoning in our study (children and adult sample) with the study by Duffy and Nagel (1997, p. 9). $d_{i,t}$ solves the equation of $number_{i,t} = median_{j,t-1} * p^{d_{i,t}}$ for individual i in group j for round t. We used 50 as the initial reference point. An individual i is categorized as, e.g., $d_{i,t} = 2$ if his or her $d_{i,t}$ lies within the interval with the boundaries $median_{j,t-1} * p^{2+1/2}$ and $median_{j,t-1} * p^{2-1/2}$. Modal categories are <u>underlined</u>.

majority of players chose numbers in the range of $d = 1$ (except for round 4 in the Duffy & Nagel sample), but in Duffy & Nagel slightly more mass lies on higher values of d compared with our design. Testing whether the distributions for levels of depth of reasoning are equal across samples reveals that children choose significantly different from the Duffy & Nagel sample in round 1 ($\chi^2 = 11.66, p = .040$) and in the subsequent rounds 2–4 (all $\chi^2 > 24.51$, all $p < .001$), with more probability mass in lower levels of depth of reasoning. Overall, adult university students in the Duffy & Nagel sample played with a slightly higher depth of reasoning than the children in our sample; however, considering the fact that we played the BCG with children aged 9–11 years, the distributions of d are surprisingly comparable.[13]

Combining the similarity of choices over the rounds with the fact that these choices were taken based on a correct understanding of the rules of the BCG, we derive our first main result:

[13] We also performed some additional analyses on determinants of game descriptives, both on the individual as well as on the group level (e.g., whether fluid IQ relates to choosing zero early in the game or other patterns of play). None of these tests yielded significant results.

Result 1:
Using a new design of the experimental BCG, children aged 9–11 years are able to understand and play a strategic interaction game like the experimental BCG used in many other studies with adults. While children start with slightly higher numbers than adults, the rate of decrease and depth of reasoning are, by and large, comparable with values for adults.

3.2 Determinants of Successful Performance

Now, we turn to the question of whether fluid IQ, an important part of cognitive skills, is associated with high strategic interaction skills, i.e., successful performance in a BCG.[14] Importantly, all analyses presented are of a correlational nature; based on our study design we cannot make any causal claim. Nevertheless, we believe that the analyses provide interesting insights into the importance of fluid IQ for choices in an experimental BCG, once the abstract version of the BCG is transformed into an easy-to-understand board game.

We present results from ordinary least squares (OLS) regressions; all models control for group fixed effects (i.e., we only look at differences within a group of five children playing the BCG) as well as gender and age. All standard errors are clustered at the group level. Gender is a binary variable; age is measured in years (but varies with a day-to-day precision between children). Fluid IQ is standardized to mean = 0 and SD = 1 to make interpretation easier (details on how we formed our outcome variables can be found in Section 2.4.).

Fluid IQ was measured using a Raven's CPM test. We collected fluid IQ scores for all 114 children in our sample. Raw scores range from 5 to 18 and show large variation. The average score is 14.7, median score is 15, and the standard deviation is 2.8 (the full distribution of scores can be found in Figure A1 in the Appendix). Therefore, we believe that we have a meaningful measure of fluid IQ to proxy an important part of general cognitive skills.

In a first step, we analyze the link between fluid IQ and the probability of choosing a weakly dominated number, i.e., a number larger than 50.[15] We estimate the likelihood of choosing a weakly dominated number in the first round (in which no other behavior has been observed prior to the choice) and for all 10 rounds of the game, both using linear probability and probit models. As documented in Table A3 in the Appendix, there is a negative correlation between fluid IQ and the likelihood of choosing a weakly dominated

[14] We also collected data on a second set of abilities relevant for successful strategic interaction, namely perspective-taking abilities. We present some first, exploratory findings for these measures in Section D in the Appendix.

[15] We use a median BCG with $p = 1/2$. For all numbers larger than 50, a player would be equally or better off when choosing 50 instead; hence, numbers larger than 50 are weakly dominated.

Table 3: Determinants of Successful Performance in the Game

	DV: Coins (R2–10)	DV: Rank (R2–10)	DV: Distance (R2–10)
	(1)	(2)	(3)
Female	0.072	0.033	-0.305
	(0.370)	(0.177)	(0.426)
Age in years	-0.200	0.120	-0.156
	(0.506)	(0.203)	(0.490)
Fluid IQ	0.251	-0.068	0.027
	(0.236)	(0.101)	(0.220)
Group FE	Yes	Yes	Yes
Observations	114	114	114

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. * p<.10, ** p<.05, *** p<.01

strategy (statistically significant in three out of four specifications). The effect of fluid IQ is also large in size: moving from an average fluid IQ to a fluid IQ one SD above the average score substantially decreases the likelihood of choosing a weakly dominated strategy in the first round (from 27.1% to 16.7%, see column (2) in Table A3).[16]

Next, we investigate the relationship between the number chosen in the first round and fluid IQ. Previous studies show a negative relationship between cognitive skills and first-round choices in BCGs, i.e., individuals with higher cognitive skills choose closer to the game-theoretic equilibrium of zero (Burnham et al. 2009; Carpenter, Graham, and Wolf 2013). However, in our setting there is no significant relationship between fluid IQ and and the number chosen in the first round (for more detailed findings and a short discussion of first-round choices, see Section B and Table A4 in the Appendix).

To analyze the predictive power of fluid IQ for strong strategic interaction skills, we first examine the number of coins a child has won, i.e., how many rounds the child won during the experimental BCG. In all subsequent analyses, we exclude results from the first round because here no prior interaction has taken place and the children cannot condition their choices on the observed behavior of their peers. Table 3, column (1), presents our findings for regressing the number of coins won on the dispositional characteristics of the children. It shows that neither gender, age, nor fluid IQ significantly explain variation. Hence, in this setting, cognitive skills, measured as fluid IQ, are not related to successful performance in the experimental BCG.

Using the number of coins (or rounds) won is easy and straightforward but there is a caveat: this measure disregards any difference in children's performance apart from being "the best" in a given round. For example, in these analyses, a child who fails to win a coin by only one step is considered equal to a child who misses half the median by 30 or

[16] Note, however, that we can only use a subsample for this analysis because we control for group FEs, and in eight out of the 23 groups no child chose a weakly dominated strategy in the first round.

more steps. Obviously, the latter child can be thought to have performed much worse than the first child. Thus, we also want to present an analysis accounting for the variations in performance between all five children in a group, not only an analysis of the winning child vs. the non-winning children. To do this, we calculate the "distance to the best response", that is, how far a child is from the choice that would make him or her win the round, given the other children's choices (see Section 2.4 for details). Because this measure is very heterogeneous both across rounds as well as across groups, we *rank* children within groups and rounds based on their distance to the best response (i.e., as noted above in Section 2.4, the child with the shortest distance—the winning child—receives rank 1, the child with the second-shortest distance receives rank 2, and so on). We can then calculate an average rank for each child over rounds 2–10, with a "good performance" corresponding to a low average rank. We report the results from regressing average rank on the personal characteristics of the child in Table 3, column (2). Our findings confirm results from column (1): neither gender, nor age, nor fluid IQ are related to average rank over the rounds. In column (3), we analyze the average distance instead of the rank—results are very comparable.[17]

Next, we analyze children's behavior over the rounds to understand how children adapt their choices over time. To do so, we use the panel structure of our data setting the child as the panel variable and rounds 1–10 as the time variable. We cannot use conventional linear panel models in our setting because the unobserved panel-level effects will very likely be correlated with the lag of the dependent variable, in our case the number chosen in the previous round. We therefore use the Arellano-Bond estimator (Arellano and Bond 1991) which basically instruments the first difference of interest (here choice in round t and $t-1$) with the second (and higher order) lag (e.g., choices in round 5 are instrumented by choices in rounds 3, 2, and 1).[18]

Table 4 presents results for the relationship between the number in round t and the following variables: numbers chosen in round $t-1$, winning number in round $t-1$, and position of the goblin in $t-1$. We see that (i) children on average choose lower numbers across rounds (the coefficient of number in round $t-1$ is < 1 in all models, confirming the results from our previous analyses), (ii) children do—to some degree—"stick" to their number chosen in a previous round (the influence of choice in round $t-1$ is significant in all specifications), but (iii) are more strongly influenced by the position of the goblin

[17] Estimating the models in Table 3 using school or class FE instead of group FE does not change the results.

[18] Note that in order to yield consistent estimates the Arellano-Bond estimator requires that the differenced unobserved time-invariant component is not correlated with the second (and higher order) lag of our outcome variable. We test this assumption in the full model (column (3) in Table 4) and find (weak) support: an Arellano-Bond test yields $p = .173$ for the test for autocorrelation of order 2; thus, we cannot reject the hypothesis that there is a zero second-order autocorrelation in our data set.

Table 4: Prior Choices and Learning over Rounds

	DV: Number Chosen in Round t				
	(1)	(2)	(3)	(4) Below-median IQ	(5) Above-median IQ
Number Chosen in $t-1$	0.435***	0.231***	0.189***	0.277***	0.147**
	(0.036)	(0.060)	(0.057)	(0.078)	(0.060)
Winning Number in $t-1$		0.623***	0.167	-0.094	0.295
		(0.141)	(0.280)	(0.504)	(0.330)
Goblin Position in $t-1$			0.662*	0.784	0.560
			(0.358)	(0.650)	(0.406)
Observations	912	912	912	384	528

Notes: The results are based on a linear dynamic panel models using the Arellano-Bond estimator (Arellano and Bond 1991). Models (1)–(3) are based on the full sample of n = 114 children. Models (4) and (5) are based on a median split in fluid IQ scores, with n = 48 children in the below-median fluid IQ group and n = 66 in the at- or above-median fluid IQ group. Standard errors in parentheses are clustered at the individual level. * p<.10, ** p<.05, *** p<.01

in the previous round (column (3) clearly shows that the position of the goblin in round $t-1$ has the strongest influence on the number chosen in round t). Because Arellano-Bond estimators do not allow for time-invariant controls, we conduct a median sample split in order to at least qualitatively analyze differences in learning behavior with respect to fluid IQ. Column (4) is based on the n = 48 children in our sample with a below-median score in the fluid IQ test, and column (5) is using the remaining n = 66 children (at or above median IQ score). Results from these two models point to two potential mechanisms how fluid IQ might affect learning behavior: First, it seems that children with lower fluid IQ tend to more strongly "stick" to their choices from the previous rounds (comparing the coefficients of Number Chosen in $t-1$). Second, the findings seem to suggest that both groups pay attention to the position of the goblin in round $t-1$ when choosing a number in round t, but that only children with a higher fluid IQ do potentially account for the position of the winning child (although this effect is statistically not significant). The latter strategy, however, seems important for successful game performance because when forming beliefs about the other players' next round choices it is potentially important whether the winning child was above or below the goblin's position (i.e., 50% of the median number). Of course, given sample size and power, these findings have to be interpreted cautiously and can only point to interesting mechanisms for further research.

Finally, we also estimate how a child's depth reasoning is related to individual characteristics. Note that a high level of depth of reasoning in itself is not necessarily a predictor of good performance in the game—children could display an excessively high depth of reasoning and "outsmart themselves" by choosing numbers that are too low (cf. Kocher and Sutter 2005). Table A5 in the Appendix reports that there is no relationship of gender, age, or fluid IQ with a child's average depth of reasoning across rounds. Thus, we can conclude our second main result:

Result 2:
In our new design of the experimental BCG, cognitive skills—measured as fluid IQ—are neither related to first round choices nor to successful performance but only predict the choice of weakly dominated strategies. Similarly, fluid IQ is not associated with a child's average depth of reasoning.

Taken together, our findings indicate that strategic interaction skills in the new design of the experimental BCG are not linked to gender or age (within our age range of 9–11 years). Fluid IQ is only relevant in predicting whether children choose weakly dominated strategies but is not associated with first round choices, successful performance, or higher depth of reasoning in the game. To support the stability of our findings, we conducted several robustness checks (excluding children with low understanding, excluding weakly dominated choices, estimations without group-fixed effects). In all versions, our findings remain stable (for details, see Section C in the Appendix).

3.3 Discussion

Previous studies have generally demonstrated positive links between cognitive skills and lower entries in the experimental BCG (Burnham et al. 2009; Brañas-Garza, García-Muñoz, and González 2012; Carpenter, Graham, and Wolf 2013). We also find a link between cognitive skills, measured as fluid IQ, and choosing weakly dominated strategies, replicating findings from Burnham et al. (2009, p. 172). Yet, Burnham et al. (2009) also document a relationship between higher cognitive skills and lower numbers chosen in a one-shot BCG which we cannot replicate (first-round choices are not linked to fluid IQ, see Table A4 in the Appendix). We believe that the specific measure for cognitive skills could be an aspect that helps explain this: Brañas-Garza, García-Muñoz, and González (2012) use a Raven's IQ test, as we do in our study, and also find no relationship between fluid IQ and choices in an experimental BCG (yet, they do report a significant link to the Cognitive Reflection Task). Moreover, our sample consists of children aged 9–11 years and it is possible that, for this age group, other abilities simply matter more than IQ; however, this explanation is made less likely by the fact that, in our replication using an adult sample, there is no significant relationship between fluid IQ and successful performance (see Section 3.4). This leads us to the explanation that we consider most plausible. The difference between our results and previous findings could be driven by the fact that in all these studies, the instructions for the BCG were abstract and, therefore, cognitive skills were more important (or even a prerequisite) for understanding the mechanisms of the game. In our setting, the instructions are far more concrete and the game itself has a visual and spatial representation. In other words, choices and their consequences are mapped into concrete and observable operations. Thus, our speculative hypothesis is that the new design of the experimental BCG lowers the importance of cognitive skills for successful performance in the game. Note that by removing the requirement to translate abstract instructions into concrete operations, we can study actual behavior in strategic interaction settings in a much more focused way. Indeed, real-world strategic interaction is often characterized by repetition, observable behavior, and concrete outcomes, as well as possibilities to learn from one's choices. Hence, removing (or lowering the demand for) this abstract component from the experimental BCG might actually increase external validity for real-world strategic interaction.

There are two methodological challenges with the lack of significant relationships between fluid IQ and successful performance. First, this could be due to restricted or limited variance. If, for example, classes or groups were very homogeneous with respect to fluid IQ levels, this could (partially) explain why there is no significant link between fluid IQ and performance. To analyze this concern, we checked the variance of the results from the Raven's Matrices task within our sample. For the whole sample, the variance in raw scores for fluid IQ amounts to 8.0 points. Calculating the within-class variance and then averaging over these within-class variances for all classes (weighted by class size) results in an average variance at the class level of 6.4 points.[19] Because we assigned children randomly to groups (within a class), we expect that the average variance at the group level would not differ from that at the class level. Indeed, the average within-group variance amounts to 6.6 points. Finally, we can compare the variance in our study with figures from a different study using the same test for fluid IQ (Berger et al. 2020). In four different testing waves with a sample of more than 500 German primary schoolchildren aged 7–9 years, this study finds variances of 8.3, 7.1, 8.1, and 6.3 points, respectively. Thus, the variance of fluid IQ in our sample (and within our groups) is substantial and in line with other, much larger samples of schoolchildren. An alternative way of testing for within-group-restricted variance as a potential explanatory factor is to estimate the OLS models from Tables 3 without group-fixed effects. In doing so, we exploit the full distribution of IQ scores within our sample (but also lose control over other factors varying between groups). We report

[19] Fluid IQ scores are positively correlated with age. Given that we cover an age range of 9–11 years and that classes are homogeneous in terms of age, it is not surprising that the variance for the whole sample is somewhat larger than the average variance at the class level.

these estimations in Table A9 in the Appendix; there is no significant link between fluid IQ and any of our measures of successful performance when excluding group-fixed effects.

Second, the lack of a significant relationship between fluid IQ and successful strategic interaction could be due to limited statistical power of our study. However, comparing our results for the first round choices with findings from Burnham et al. (2009, they use a BCG with $p = 1/2$ and choices between 0 and 100), we see that they report an effect of -9.67 in first-round choices for a one standard deviation increase in cognitive skills (p. 173). In contrast, if we regress the number chosen in round 1 on gender, age and fluid IQ (clustering standard errors at the group level), the 95% confidence interval for a one standard deviation increase in fluid IQ on the number chosen in round 1 ranges from -7.26 to 4.29; this suggests that we can basically rule out effects of fluid IQ in the size found by Burnham et al. (2009). In addition, our exploratory analysis on the link between perspective-taking abilities and successful performance in the BCG (see Section D in the Appendix) indicates that for other individual characteristics (even a binary one), our study seems to be sufficiently powered to identify statistically significant relationships.

Finally, we do not identify any significant relationship between age and successful strategic interaction. In principle, an increase in strategic interaction skills with age could be anticipated (e.g., Brosig-Koch, Heinrich, and Helbach (2015) show that children's ability to reason backward clearly improves from the age of 6 years onward, and Charness et al. (2019) show with a sample of children between 3 and 11 years old that Theory of Mind increases considerably with age). On the other hand, Czermak et al. (2016) find no substantial effects of age in strategy games and conclude that their results suggest that "strategic decision-making is fairly well developed at an age of 10 years and hardly changes in subsequent years" (p. 270). Potentially, this conclusion could already apply at the age of 9 years, which is the lower bound of age in our sample of children. However, our results regarding age must be interpreted with caution because (i) we study a rather small age range of only two years, (ii) we only compare children within groups who are even more homogeneous in age than the full sample (because groups were randomly drawn from the same class), and (iii) our distribution of age is not linear because we study a cohort of third and fifth graders (i.e., there are no fourth graders in the sample). Hence, identifying age effects in such a setting is challenging and the absence of significant age differences in successful strategic interaction in our study should not be interpreted as evidence of the absence of development in strategic interaction skills within this age range.

3.4 Replication with an Adult Sample

When we tested the new design of the experimental BCG with our sample of children and compared it with the previous findings in the literature, we changed two factors simultaneously: the design of the game and the sample of participants. To provide the "missing piece", we replicated our study using the new design of the experimental BCG but with an adult sample of university students.

Experimental Design. We recruited $n = 120$ participants, 60% of whom were female, with a mean age of 22.6 years (see Table A12 in the Appendix for details). The experiment was conducted in the MABELLA (Mainz Behavioral and Experimental Laboratory). The sessions were combined with another experiment but mirrored the basic structure of the study with children: first, all participants within a session (10) were seated at separate tables in a large room and filled out questionnaires and tests (including a (short) version of Raven's Matrices for adults). Subsequently, two groups of five adults (randomly assigned) each went to a separate room with an experimenter to play the new design of the experimental BCG. The only difference with the study with children was that adults did not receive one-to-one instructions but were instructed as a group (also, they did not have to explain the game back to the experimenter). After the BCG, participants were paid anonymously in a separate room. We conducted 12 sessions with two groups each, sessions lasted for 70-90 min in total. Average payoff was EUR 15.45, including a show-up fee of EUR 5. The experimental BCG was incentivized, with the winner of a randomly drawn round receiving EUR 20. The fluid IQ test was not incentivized.

Results. Figure 2 and Table 1 show that choices by adults start at a lower level than those by children and remain consistently lower in subsequent rounds (Mann-Whitney U test for rounds 1–4 for each round, all $p < .0001$). More detailed information on adults' choices can be found in Table A13 and Figure A3. Benchmarking adults' choices with choices in the study by Nagel (1995) suggests that our adult sample playing the BCG in the new design behaves very similarly to the adult sample in Nagel's study playing the classical BCG with $p = 1/2$. Comparing the numbers chosen in the first round for our adult sample and for the Nagel $p = 1/2$ sample reveals no significant difference (Mann-Whitney U test, $p = .899$). However, in rounds 2–4 the Mann-Whitney U tests are significant, suggesting that participants in the Nagel $p = 1/2$ sample choose lower numbers than those in our adults sample (Mann-Whitney U tests for each round, all $p < .032$), which is in line with the notion that median choices decrease somewhat faster in the classical BCG, see Table 1.[20]

When comparing the distributions of depth of reasoning in rounds 1–4 in Table 2, the majority of adults display a $d = 1$. If we compare the distributions of levels of d, we see that our adult sample is significantly different from the children sample in round 1 ($\chi^2 = 23.10, p < .001$), with adults having more probability mass in higher levels of reasoning, as one would expect. In later rounds, however, the picture is mixed.[21] Comparing the depth of reasoning in our adult sample with the adult sample in Duffy and Nagel (1997), we find that for round 1 the distributions of d are not significantly different ($\chi^2 = 4.95, p = .422$). In subsequent rounds, adults in our sample show lower levels of depth of reasoning than the adults in Duffy & Nagel (for rounds 2–4, all $\chi^2 > 9.82$, all $p < .080$). Overall, levels of d in our adult sample are slightly higher than in our children sample and slightly lower than in the adult sample by Duffy and Nagel (1997), which places the distribution of our adult sample playing the new design of the BCG *between* the children sample playing the new design of the BCG and the adult sample by Duffy and Nagel (1997) playing the classical design.

In total, the replication of the new design of the experimental BCG with an adult sample shows that this design can be used successfully to study strategic interaction with adults:

Result 3:
Adults playing the new design of the experimental BCG behave in a way that is largely comparable with adult behavior in the classical BCG. The average numbers chosen and the rate of decrease are very similar to those chosen by other adult samples. The average depth of reasoning is higher than that in our children sample and slightly lower than that in the classical BCG.

Although this was not the focus of our replication study, we also conducted a parallel analysis of the link between cognitive skills, measured as fluid IQ, and successful performance for adults. Results can be found in Table A14. Similar to the children sample, successful performance is not related to fluid IQ. This indicates that the new design might indeed place lower demands on cognitive skills by making the game less abstract and easier to understand (see Section 3.3). In addition, we document a substantial gender difference in the adult sample. Women perform significantly worse than men when looking at the

[20] A potential difference driving this could be the fact that Nagel (1995) used the mean, whereas we used the median in our study to identify the winner.

[21] The distributions of d are (marginally) significantly different in round 2 ($\chi^2 = 9.87, p < .079$), not significantly different in round 3 ($\chi^2 = 6.38, p < .271$), and significantly different in round 4 ($\chi^2 = 11.37, p < .044$).

number of coins won and the average rank.[22] Taken together, the replication study using an adult student sample generally confirms that the new design of the experimental BCG can also be used with adults. Further investigating determinants of successful performance and gender differences in strategic interactions for adults as well as skills that have a *causal* effect on successful performance appear to be promising avenues for further research.

4 Conclusion

This paper introduces a new design of the experimental BCG. We use this new design for the first-ever study conducting a BCG with groups of children and demonstrate that children are capable of understanding and playing this BCG. This allows for a wide range of applications in studying the skill formation process for strategic interaction. Moreover, our findings demonstrate that an important part of cognitive skills, namely fluid IQ, is not significantly related to successful performance in this strategic interaction setting, opening up the question which skills are important to succeed in strategic interaction. At the same time, our new design allows for research designs focusing on the development of strategic interaction skills (and, potentially, its determinants) starting already at young age. Finally, in the implementation of the new BCG design with adults we find results largely in line with behavior in the classical BCG, suggesting that our new BCG design can also used with adult samples.

In future research, it would be promising to extend the age range studied, for example, to children entering school or adolescents. This could improve our understanding of how the ability to strategically interact in groups develops with age. Moreover, in order to advance our understanding of important determinants of successful strategic interaction it seems promising to include measures of abilities in the area of perspective-taking and empathy (see Section D in the Appendix for some first, exploratory evidence). A different perspective could use longitudinal data to analyze predictors of successful strategic interaction, i.e., which skills and abilities are the building blocks of this complex ability? Relatedly, which background characteristics are linked to strategic interaction skills? For example, the detection of an early gap in strategic interaction skills based on socioeconomic background (controlling for cognitive skills and other important abilities) would contribute to our understanding of the intergenerational transmission of strategic interaction skills as well as

[22] When analyzing how the adult sample adapts choices over time (compared to Table 4), we find qualitatively very similar results. This similarity is largely in line with results of testing for differences in the rates of decrease from Table 1: Testing for equality in central tendencies between the rates of decrease for children vs. adults reveals no significant difference for the decrease in mean numbers (Mann-Whitney U test, $p = .349$) but a significant difference for the decrease in median numbers (Mann-Whitney U test, $p = .045$).

the origins of socioeconomic inequalities. Finally, causal evidence, e.g., using priming and/or cognitive load paradigms, or even targeted interventions could provide further insights into the importance of various skills as building blocks of strategic interaction skills.

References

Arellano, Manuel, and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *The Review of Economic Studies* 58 (2): 277–297.

Berger, Eva M., Ernst Fehr, Henning Hermes, Daniel Schunk, and Kirsten Winkel. 2020. *The Impact of Working Memory Training on Children’s Cognitive and Noncognitive Skills.* IZA Discussion Paper No. 13338. Bonn/Germany: IZA.

Birch, S., V. Li, T. Haddock, S. Ghrear, P. Brosseau-Liard, A. Baimel, and M. Whyte. 2017. “Chapter Six - Perspectives on Perspective Taking: How Children Think About the Minds of Others,” edited by Janette B. Benson, 52:185–226. Advances in Child Development and Behavior. JAI.

Bosch-Domènech, Antoni, José G. Montalvo, Rosemarie Nagel, and Albert Satorra. 2002. “One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments.” *The American Economic Review* 92 (5): 1687–1701.

Bosch-Rosa, Ciril, and Thomas Meissner. 2020. “The one player guessing game: a diagnosis on the relationship between equilibrium play, beliefs, and best responses.” *Experimental Economics.*

Brañas-Garza, Pablo, Teresa García-Muñoz, and Roberto Hernán González. 2012. “Cognitive effort in the Beauty Contest Game.” *Journal of Economic Behavior & Organization* 83 (2): 254–260.

Brandts, Jordi, and David J. Cooper. 2006. “A Change Would Do You Good .... An Experimental Study on How to Overcome Coordination Failure in Organizations.” *American Economic Review* 96 (3): 669–693.

Braun, Sebastian, Nadja Dwenger, Dorothea Kübler, and Alexander Westkamp. 2014. “Implementing quotas in university admissions: An experimental analysis.” *Games and Economic Behavior* 85:232–251.

Brocas, Isabelle, and Juan D. Carrillo. 2018a. “Iterative dominance in young children: Experimental evidence in simple two-person games.” *Journal of Economic Behavior & Organization.*

———. 2018b. “The determinants of strategic thinking in preschool children.” *PLOS ONE* 13 (5): 1–14.

———. 2019. *Steps of reasoning in children and adolescents.* Technical report.

———. 2020. “The evolution of choice and learning in the two-person beauty contest game from kindergarten to adulthood.” *Games and Economic Behavior* 120:132–143.

Brosig-Koch, Jeannette, Timo Heinrich, and Christoph Helbach. 2015. “Exploring the capability to reason backwards: An experimental study with children, adolescents, and young adults.” *European Economic Review* 74:286–302.

Bulheller, S., and H. O. Häcker. 2010. *Coloured Progressive Matrices (CPM). Deutsche Bearbeitung und Normierung nach J. C. Raven.* Frankfurt: Pearson Assessment.

Burnham, Terence C., David Cesarini, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2009. “Higher cognitive ability is associated with lower entries in a p-beauty contest.” *Journal of Economic Behavior & Organization* 72 (1): 171–175.

Carpenter, Jeffrey, Michael Graham, and Jesse Wolf. 2013. “Cognitive ability and strategic sophistication.” *Games and Economic Behavior* 80:115–130.

Castillo, Marco, Jeffrey L. Jordan, and Ragan Petrie. 2018. “Children’s rationality, risk attitudes and field behavior.” *European Economic Review* 102:62–81.

Charness, Gary, John A. List, Aldo Rustichini, Anya Samek, and Jeroen Van De Ven. 2019. "Theory of mind among disadvantaged children: Evidence from a field experiment." *Journal of Economic Behavior & Organization* 166:174–194.

Cooper, David J., and John H. Kagel. 2005. "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games." *American Economic Review* 95 (3): 477–509.

Czermak, Simon, Francesco Feri, Daniela Glätzle-Rützler, and Matthias Sutter. 2016. "How strategic are children and adolescents? Experimental evidence from normal-form games." *Journal of Economic Behavior & Organization* 128:265–285.

Duffy, John, and Rosemarie Nagel. 1997. "On the Robustness of Behaviour in Experimental 'Beauty Contest' Games." *The Economic Journal* 107 (445): 1684–1700.

Egan, Daniel, Christoph Merkle, and Martin Weber. 2014. "Second-order beliefs and the individual investor." *Journal of Economic Behavior & Organization,* Empirical Behavioral Finance, 107, Part B:652–666.

Fe, Eduardo, David Gill, and Victoria Prowse. 2019. *Cognitive skills, strategic sophistication, and life outcomes.* Technical report. Competitive Advantage in the Global Economy (CAGE).

Grosskopf, Brit, and Rosemarie Nagel. 2008. "The two-person beauty contest." *Games and Economic Behavior* 62 (1): 93–99.

Hanaki, Nobuyuki, Yukio Koriyama, Angela Sutan, and Marc Willinger. 2019. "The strategic environment effect in beauty contest games." *Games and Economic Behavior* 113:587–610.

Harbaugh, William T., Kate Krause, Steven G. Liday, and Lise Vesterlund. 2003. "Trust in Children." In *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research,* 302–322. Russell Sage Foundation.

Harbaugh, William T., Kate Krause, and Steven J. Liday. 2003. *Bargaining by Children.* SSRN Scholarly Paper ID 436504. Rochester, NY: Social Science Research Network.

Harbaugh, WT, and K Krause. 2000. "Children's altruism in public good and dictator experiments." *Economic Inquiry* 38 (1): 95–109.

Ho, Teck-Hua, Colin Camerer, and Keith Weigelt. 1998. "Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests"." *The American Economic Review* 88 (4): 947–969.

Kautz, Tim, James Heckman, Ron Diris, Bas t. Weel, and Lex Borghans. 2014. *Fostering and Measuring Skills.* OECD Education Working Papers. Paris: Organisation for Economic Co-operation and Development.

Kocher, Martin G., and Matthias Sutter. 2005. "The decision maker matters: Individual versus group behaviour in experimental beauty-contest games." *The Economic Journal* 115 (500): 200–223.

———. 2006. "Time is money—Time pressure, incentives, and the quality of decision-making." *Journal of Economic Behavior & Organization* 61 (3): 375–392.

Murnighan, J.Keith, and Michael Scott Saxon. 1998. "Ultimatum bargaining by children and adults." *Journal of Economic Psychology* 19 (4): 415–445.

Nagel, R. 1995. "Unraveling in guessing games: An experimental study." *American Economic Review* 85 (5): 1313–1326.

Piaget, Jean. 1962. "The Stages of the Intellectual Development of the Child." *Bulletin of the Menninger Clinic* 26 (3): 120–128.

Rangvid, Jesper, Maik Schmeling, and Andreas Schrimpf. 2013. "What do professional forecasters' stock market expectations tell us about herding, information extraction and beauty contests?" *Journal of Empirical Finance* 20:109–129.

Sher, Itai, Melissa Koenig, and Aldo Rustichini. 2014. “Children’s strategic theory of mind.” *Proceedings of the National Academy of Sciences* 111 (37): 13307–13312.

Sutter, Matthias. 2005. “Are four heads better than two? An experimental beauty-contest game with teams of different size.” *Economics Letters* 88 (1): 41–46.

Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler. 2019. “Economic behavior of children and adolescents – A first survey of experimental economics results.” *European Economic Review* 111:98–121.

Weber, Roberto A. 2003. “‘Learning’ with no feedback in a competitive guessing game.” *Games and Economic Behavior* 44 (1): 134–144.

Wellman, Henry M., David Cross, and Julanne Watson. 2001. “Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief.” *Child Development* 72 (3): 655–684.

Online Appendix

for

If You Could Read My Mind—An Experimental

Beauty-Contest Game with Children

Henning Hermes and Daniel Schunk

A Descriptive Statistics

Table A1: Sample Descriptives

	Mean	SD	N	Min	Max
Female	0.55	0.50	114	0	1
Age in years	9.98	1.04	114	8	12
Class 1	0.22	0.42	114	0	1
Class 2	0.22	0.42	114	0	1
Class 3	0.13	0.34	114	0	1
Class 4	0.13	0.34	114	0	1
Class 5	0.13	0.34	114	0	1
Class 6	0.17	0.37	114	0	1
Fifth Grade	0.44	0.50	114	0	1
Fluid IQ	0.01	1.00	114	-3.4	1.2
Understanding of the Game	19.13	1.59	113	9	20
Perspective-taking	0.22	0.42	112	0	1
Social Appropriateness	0.00	1.00	114	-6	1.2
Interpersonal Reactivity	0.00	1.00	114	-3.1	2.4

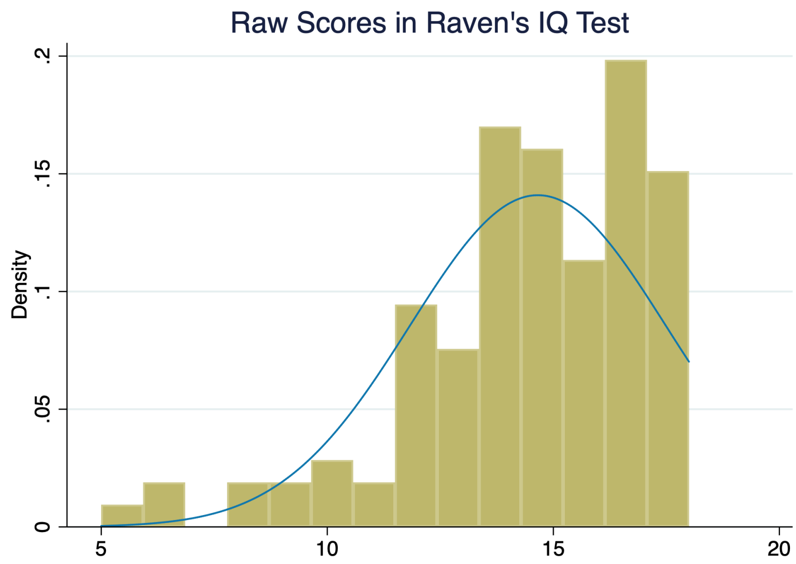
Notes: Sample descriptives for the children sample. We collected data for 23 groups of five children. For one child, we did not obtain parental consent for data use. Therefore, the final number of observations is 114. For one child, we are missing data on the ratings for understanding of the game because the experimenter did not record responses for these children in the protocol sheet.

Table A2: Game Descriptives

	Mean	Median	SD	N	Min	Max
Number Round 1	33.46	28	20.64	114	1	91
Number Round 2	21.73	20	12.17	114	4	56
Number Round 3	14.96	12	11.45	114	1	75
Number Round 4	10.68	8	11.18	114	0	100
Number Round 5	10.04	6.5	14.47	114	0	100
Number Round 6	5.42	3	5.38	114	0	30
Number Round 7	3.89	2.5	4.81	114	0	27
Number Round 8	3.36	2	5.30	114	0	44
Number Round 9	3.33	1	9.77	114	0	100
Number Round 10	2.51	1	5.83	114	0	50
Mean Number 1–10	10.94	10	4.91	114	2.6	26

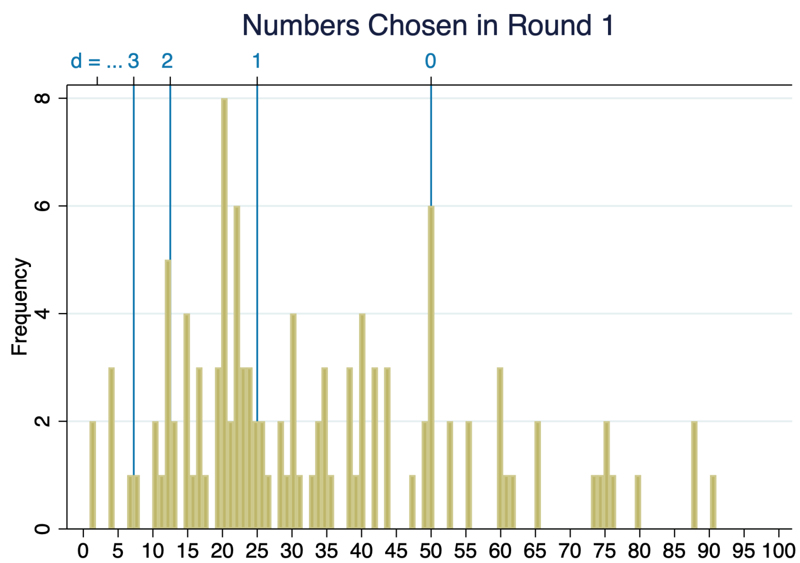
Notes: Descriptives for the numbers chosen in each round of the game for the children sample. We collected data for 23 groups of five children. As we could not obtain parental consent for data use for one child, the final number of observations is 114.

Figure A1: Histogram of Raven's IQ Scores



Notes: This figure shows a histogram of the raw scores in the Raven's IQ test. In principle, scores range from 0–18; in the sample, the minimum score is 5, the maximum score is 18. The average score is 14.7 and the median score is 15. The number of observations is 114.

Figure A2: Histogram of Numbers Chosen in Round 1



Notes: This figure shows a histogram for the numbers chosen in round 1 in the children sample. The blue vertical lines display resulting values for a depth of reasoning (d) of 0, 1, 2, and 3 (starting from an initial reference point of 50). See Section 3.1 for details on depth of reasoning.

B Further Analyses

Table A3: Fluid IQ and Weakly Dominated Choices in the BCG

	DV: Dominated Choice in R1		DV: Dominated Choice in R1–10	
	R1 (OLS) (1)	R1 (Probit) (2)	R1–R10 (OLS) (3)	R1–R10 (Probit) (4)
Female	0.169	0.531	0.163	0.504
	(0.122)	(0.380)	(0.131)	(0.389)
Age in years	0.056	0.198	0.005	-0.002
	(0.137)	(0.446)	(0.108)	(0.345)
Fluid IQ	-0.118	-0.401*	-0.106*	-0.364*
	(0.068)	(0.221)	(0.056)	(0.206)
Group FE	Yes	Yes	Yes	Yes
Observations	74	74	89	89

Notes: The results are based on OLS regressions (columns 1 and 3) and probit models (columns 2 and 4). Standard errors in parentheses are clustered at the group level. We cannot use all observations in these models because in round 1, there are eight groups in which no child chose a weakly dominated number, and over all 10 rounds, there are such five groups. Therefore, the variable for weakly dominated choices is perfectly collinear with the group dummy and these groups have to be dropped. * p<.10, ** p<.05, *** p<.01

Table A4: OLS Regressions for Round 1

	DV: Nr. R1 (1)	DV: Nr. R1 $\leq$ 50 (2)	DV: Nr. R1 Age (3)	DV: Coins R1 (4)	DV: Rank R1 (5)	DV: Dist. R1 (6)
Female	4.561	-0.330	5.721	0.107	0.208	0.383
	(3.571)	(2.980)	(4.291)	(0.081)	(0.268)	(0.232)
Age in years	8.639**	7.020**	12.036*	-0.103	0.513	0.606*
	(3.962)	(2.524)	(6.942)	(0.118)	(0.340)	(0.295)
Fluid IQ	-1.878	3.345*	-2.370	0.078	-0.083	-0.275
	(3.762)	(1.841)	(4.198)	(0.065)	(0.227)	(0.299)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114	94	100	114	114	114

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. Column (2) excludes all weakly dominated choices, i.e., numbers larger than 50. Column (3) excludes age outliers, i.e., all children below the 5th and above the 95th percentile within their grade. For details on the outcomes in columns (4)–(6), see Section 2.4.
* p<.10, ** p<.05, *** p<.01

Table A5: OLS Regressions for Depth of Reasoning in R2–10

	DV: Depth of Reasoning R2–R10 (1)
Female	-0.000
	(0.205)
Age in years	0.023
	(0.167)
Fluid IQ	0.059
	(0.086)
Group FE	Yes
Observations	114

Notes: The results are based on OLS regressions. The dependent variable is the average depth of reasoning in rounds 2–10 (see Section 3.1 for details), capped at values -6 and +6. Standard errors in parentheses are clustered at the group level. * p<.10, ** p<.05, *** p<.01

Choices in the First Round. Table A4 reports our analysis of numbers chosen in the first round. In this round, no prior behavior from other players has been observed. There is no significant effect of fluid IQ on first-round choices. The coefficient in column (1) is slightly negative; however, once weakly dominated choices (> 50) are excluded—see column (2)—the coefficient even turns positive (and is significant at the 10% level). Also, for all measures of successful performance, there is no significant link to fluid IQ, as shown in columns (4)–(6). Hence, even for first-round choices, cognitive skills (measured as fluid IQ) are not related to lower entries or successful performance in the new design of the experimental BCG.

Table A4 also reports a significant and large effect of age: older children choose higher numbers in the first round of the game. This is not driven by weakly dominated choices, as column (2) indicates. To rule out potential effects of age outliers, we exclude children below the 5th and above the 95th age percentile within each cohort (third grade and fifth grade). Results in column (3) indicate that the positive link between age and first-round choices persists (the coefficient even increases); therefore, it appears that age outliers are not driving this finding. Columns (4)–(6) demonstrate that these higher numbers translate into worse performance for older children in the first round (but statistically significant only for the distance measure). However, the effect does not transfer to the next rounds (see results from Tables 3; also, when looking at the numbers chosen in round 2, the relationship fades out and becomes insignificant). At first sight, a reasonable explanation would be that older children are (more) familiar with higher numbers. However, we only compare children within groups coming from the same class. Therefore, all of them learn together and have the same experiences in mathematics. We argue that this makes it very unlikely that there are any age-related differences with respect to familiarity with numbers within class (and within groups).[23] Overall, there seems to be a strong link between older children choosing higher numbers that cannot be explained by weakly dominated choices, age outliers, or familiarity with higher numbers.

[23] One could, in principle, think of children repeating a grade but when excluding these age outliers, the effect persists, see column (3).

C Robustness Checks

To support the stability of our findings, we conduct a number of robustness checks. Our tables for robustness checks always report results from re-estimating the models in Table 3.

First, we exclude all children with an experimenter-rated understanding of the game below 19 points (out of 20 points). Results in Table A6 show that our findings are not affected by this sample restriction. Second, we wanted to check whether outliers or extreme choices are driving our results. Therefore, we reproduced our estimations excluding children who chose weakly dominated numbers (see Table A7). Third, in light of the (non-existing) age effects discussed above (see Section 3.3), we wanted to analyze whether the presence of very young or very old children within their grade (e.g., because of grade retention) could influence the results obtained. Therefore, we excluded children below the 5th and above the 95th age percentile (within their cohort) and present the results for these models in Table A8. The coefficients for age remain insignificant, confirming that there is no systematic effect of age on successful performance in the BCG in our age range (however, coefficients now at least point into the hypothesized direction, indicating that in a larger age range one might be able to detect an effect of age). Fourth, all our main findings are conservatively estimated using group-fixed effects, i.e., we only compare children within a group of five kids. The upshot of this approach is that we control for any unobserved heterogeneity across groups. However, we also report estimations using OLS without group-fixed effects in Table A9 in the Appendix. When excluding group-fixed effects, there is still no significant relationship between individual characteristics and performance in the BCG. Finally, to support the notion that age is not substantially related to successful performance and is only linked to choices in the first round, we also present estimations with and without age in Table A10 in the Appendix. Removing age as a control variable essentially does not alter the coefficients for any of our determinants of success in the game.

In addition, we also checked whether choices and performance in the game were related to two external factors, using ANOVAs. First, the color (and also the seating position at the table while playing the game, as this was determined by the color, see Figure A4 in the Appendix) did not influence performance throughout the game. Second, experimenters rotated between colors across groups so that they would explain the game to players with different colors each round.[24] Therefore, we can identify a separate effect of the person who explained the game to the child. There was no significant difference in choices or

[24] Experimenter A explained the game to the child with white in group 1, then to the child with yellow in group 2, and so on, and experimenter B explained the game the child with yellow in group 1, blue in group 2, and so on.

performance with respect to the experimenter who explained the game to the child. Because we use group-fixed effects in all main estimations, other differences such as time of testing, class-, teacher- or school-fixed effects are captured in the group dummy.

Table A6: OLS Regressions with Coins Won, Rank, and Distance Excluding Low Understanding

	DV: Coins R2–10 (1)	DV: Rank R2–10 (2)	DV: Distance R2–10 (3)
Female	-0.090	0.081	-0.392
	(0.464)	(0.205)	(0.496)
Age in years	-0.126	0.060	-0.346
	(0.692)	(0.259)	(0.620)
Fluid IQ	0.305	-0.041	0.203
	(0.306)	(0.155)	(0.287)
Group FE	Yes	Yes	Yes
Observations	90	90	90

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. In this table, children with an experimenter-rated understanding of the game below 19 points (out of 20) are excluded from the analysis. For details on the outcomes, see Section 2.4.
* p<.10, ** p<.05, *** p<.01

Table A7: OLS Regressions with Coins Won, Rank, and Distance Excluding Weakly Dominated Choices

	DV: Coins R2–10 (1)	DV: Rank R2–10 (2)	DV: Distance R2–10 (3)
Female	0.007	0.040	0.067
	(0.478)	(0.204)	(0.447)
Age in years	-0.261	0.111	-0.145
	(0.612)	(0.215)	(0.505)
Fluid IQ	0.436	-0.141	-0.029
	(0.277)	(0.121)	(0.212)
Group FE	Yes	Yes	Yes
Observations	87	87	87

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. In this table, children who chose a weakly dominated number in any of the 10 rounds are excluded from the analysis. For details on the outcomes, see Section 2.4.
* p<.10, ** p<.05, *** p<.01

Table A8: OLS Regressions with Coins Won, Rank, and Distance Excluding Age Outliers

	DV: Coins R2–10 (1)	DV: Rank R2–10 (2)	DV: Distance R2–10 (3)
Female	-0.228	0.123	-0.073
	(0.349)	(0.183)	(0.447)
Age in years	0.456	-0.101	-0.835
	(0.370)	(0.198)	(0.843)
Fluid IQ	0.365	-0.134	-0.076
	(0.268)	(0.106)	(0.246)
Group FE	Yes	Yes	Yes
Observations	100	100	100

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. In this table, children aged below the 5th or above the 95th percentile within their respective grade (third or fifth grade) are excluded from the analysis. For details on the outcomes, see Section 2.4. * p<.10, ** p<.05, *** p<.01

Table A9: OLS Regressions with Coins won, Rank, and Distance (No Group FEs)

	DV: Coins R2–10		DV: Rank R2–10		DV: Distance R2–10	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.072	-0.148	0.033	0.095	-0.305	-0.239
	(0.370)	(0.363)	(0.177)	(0.147)	(0.426)	(0.373)
Age in years	-0.200	0.171	0.120	-0.009	-0.156	0.010
	(0.506)	(0.219)	(0.203)	(0.062)	(0.490)	(0.250)
Fluid IQ	0.251	0.194	-0.068	-0.067	0.027	0.099
	(0.236)	(0.194)	(0.101)	(0.078)	(0.220)	(0.197)
Group FE	Yes	No	Yes	No	Yes	No
Observations	114	114	114	114	114	114

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. For details on the outcomes, see Section 2.4. * p<.10, ** p<.05, *** p<.01

Table A10: OLS Regressions with Coins Won, Rank, and Distance (Not Controlling for Age)

	DV: Coins R2–10		DV: Rank R2–10		DV: Distance R2–10	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.072	0.063	0.033	0.038	-0.305	-0.312
	(0.370)	(0.363)	(0.177)	(0.175)	(0.426)	(0.413)
Age in years	-0.200		0.120		-0.156	
	(0.506)		(0.203)		(0.490)	
Fluid IQ	0.251	0.257	-0.068	-0.071	0.027	0.031
	(0.236)	(0.232)	(0.101)	(0.100)	(0.220)	(0.229)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114	114	114	114	114	114

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. Columns (2), (4), and (6) present the main results without controlling for age. For details on the outcomes, see Section 2.4. * p<.10, ** p<.05, *** p<.01

D Exploratory Analysis of Perspective-Taking Abilities in Strategic Interaction

In an attempt to provide some first, exploratory evidence on the importance of *other* abilities than cognitive skills for successful strategic interaction, we also collected data on children's perspective-taking abilities. In contrast to cognitive skills, which we measured as fluid IQ, using an established test (Raven's Matrices), for perspective-taking abilities there is no such an established measure, at least not for children in our age range. Therefore, we decided to measure perspective-taking abilities with different instruments, covering several methods as well as several aspects of perspective-taking abilities. Specifically, we decided to use (i) an easy and quick-to-implement behavioral task measuring perspective-taking ("E on the forehead", Glen 1984), and to use two different self-reported measures, (ii) one focusing on understanding social situations as well as emotions and behavior in third-person situations (Meindl 1998), and (iii) another focusing more on an individual difference perspective along the dimension of cognitive empathy, i.e., the ability to understand and process other people's emotions and perspectives (Garton and Gringart 2005). The latter questionnaire was developed as a child-friendly version of a frequently used tool to measure perspective-taking and empathy in adults, namely the Interpersonal Reactivity Index (IRI, Davis 1983).

Answers for the two questionnaire measures were collected in the first lesson of each school day, together with the whole class (see Section 2.1). The behavioral task was conducted at the end of the explanation session which each child received individually with an experimenter, right before children went to the large table to play the game together with the other children (see Section 2.3 for details). Table A1 provides descriptive statistics for our measures of perspective-taking abilities; the exact instructions for the behavioral perspective-taking task and the wording of the items of the questionnaires are provided in Sections F.2 and G. Specifically, we use the following variables for our subsequent analysis:

Perspective-taking. To provide a behavioral measure of perspective-taking abilities, we adapted the "E on the forehead" task (Glen 1984). This task was designed as a measure of perspective-taking based on self-awareness or self-consciousness and has frequently been used in psychological experiments as a quick and intuitive measure of perspective-taking behavior vs. egocentric responses (see, for example, Steins and Wicklund 1996; Galinsky et al. 2006). In the present study, we asked each child to "trace a capital 'E' with your forefinger on your forehead" and recorded, whether the 'E' was readable from the child's or the experimenter's perspective (or something different was traced). The dummy variable is one if the child traces an 'E' that is readable from the experimenter's perspective and zero if she traces something different (i.e., an 'E' readable from the child's perspective or something different). We were able to collect this task from n = 112 children in our sample; 25 children (22%) traced an 'E' that was readable from the experimenter's perspective.

Social Appropriateness Scale. As one of our two self-reported measures of perspective-taking, we selected six small stories from a questionnaire measuring empathy and appropriate behavior in a social situation that was developed by Meindl (1998). Children receive a point for each correctly solved question. To enable comparison of effect sizes, the variable was standardized to mean = 0 and SD = 1. All children in our sample (n = 114) answered this questionnaire.

Interpersonal Reactivity Index. As a second self-reported measure for perspective-taking abilities, we use a questionnaire adapted for school-aged children developed by Garton and Gringart (2005). This questionnaire is based on one of the most frequently used self-reported measures for perspective-taking and empathy in adults, namely the Interpersonal Reactivity Index (IRI, Davis 1983). We selected four items focusing on cognitive aspects of empathy from the children's version of the questionnaire. In contrast to the Social Appropriateness scale, there is no correct answer, but children rate how much each item applies to them personally. We use the sum of all four items (no reverse-coded item) and standardize the resulting variable to mean = 0 and SD = 1 to enable comparison of effect sizes. All children in our sample (n = 114) answered this questionnaire.

To analyze the relationship between perspective-taking abilities and successful performance in the experimental BCG, we simply add the three measures of perspective-taking abilities to the models we estimated in Table 3. Results are presented in Table A11. Note that our measures for fluid IQ, Social Appropriateness, and Interpersonal Reactivity are standardized to mean = 0 and SD = 1; thus, one can easily compare the size of their coefficients.

As columns (1) and (2) in Table A11 show, all three measures of perspective-taking significantly predict the number of rounds a child wins during the game. When including group-fixed effects, children displaying high perspective-taking abilities in the "E on the forehead" task (labeled "perspective-taking" in the tables) win more than one additional coin (on average, children win 3.8 coins, SD = 1.9). A one standard deviation increase in

Table A11: Exploratory Analysis on the Role of Perspective-Taking Abilities

	DV: Coins R2–10		DV: Rank R2–10		DV: Distance R2–10		DV: Depth of Reas.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.173	-0.068	-0.008	0.068	-0.251	-0.221	0.001
	(0.376)	(0.360)	(0.170)	(0.139)	(0.476)	(0.410)	(2.170)
Age in years	-0.327	0.056	0.175	0.034	-0.066	0.065	0.552
	(0.487)	(0.230)	(0.202)	(0.072)	(0.446)	(0.253)	(1.697)
Fluid IQ	0.137	0.141	-0.013	-0.051	0.179	0.117	-0.297
	(0.260)	(0.211)	(0.110)	(0.084)	(0.283)	(0.219)	(0.843)
Perspective-taking	1.105**	1.149***	-0.532**	-0.388***	-1.652*	-0.219	5.940*
	(0.416)	(0.336)	(0.192)	(0.118)	(0.900)	(0.495)	(3.435)
Social Appropriateness	0.258*	0.345**	-0.112*	-0.102**	-0.292	-0.246	1.418**
	(0.130)	(0.133)	(0.064)	(0.039)	(0.317)	(0.257)	(0.542)
Interpersonal Reactivity	-0.450***	-0.404**	0.200***	0.157**	0.306	0.173*	-2.080**
	(0.155)	(0.182)	(0.065)	(0.057)	(0.214)	(0.091)	(0.982)
Group FE	Yes	No	Yes	No	Yes	No	Yes
Observations	112	112	112	112	112	112	112

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level.
* p<.10, ** p<.05, *** p<.01

scores on the Social Appropriateness scale is linked to 0.26 more coins. Surprisingly, while scores from the Interpersonal Reactivity Index (IRI) significantly predict performance in the experimental BCG, they are *negatively* linked to game performance: a one standard deviation increase in scores from the IRI is related to 0.45 fewer coins during rounds 2–10.

Similarly, perspective-taking abilities are strongly linked to successful performance measured as the rank in distance to the best response (see Section 2.4 for details on the dependent variables), as presented in columns (3) and (4) in Table A11. Controlling for group-fixed effects, children with high perspective-taking abilities in the "E on the forehead task" perform half a rank better (mean rank over rounds 2–10 is 2.5, SD = 0.70). Social Appropriateness is significantly linked to better performance (one SD increase improves the average rank by around 0.1), while the IRI is related to worse performance (a one SD increase worsens the average rank by around 0.2). Columns (5) and (6) show a very similar pattern for the average distance over rounds 2–10 as an outcome, although most relationships are no longer statistically significant. Finally, similar to our findings for measures of actual game *performance*, higher depth of reasoning is systematically associated with perspective-taking abilities (see column (7) in Table A11).

All in all—in contrast to fluid IQ as an important part of cognitive skills—measures of perspective-taking abilities predict successful performance in an experimental BCG: Children showing high perspective-taking abilities in the "E on the forehead" task and scoring high on Social Appropriateness demonstrate better success (and higher depth of reasoning), whereas children with higher scores on the Interpersonal Reactivity Index show worse performance in the game.

But why are children who traced the 'E' so that the experimenter could read it on average so much better in strategic interaction (controlling for other characteristics such as gender, age, and fluid IQ)? Our preferred hypothesis is that the ability to form accurate beliefs about other players' choices and to "put yourself into the other players' shoes" is a key ability to succeed in strategic interaction games. Thus, while we are not aware of any prior study using the "E on the forehead" task to predict strategic interaction behavior, tracing the 'E' in this way seems to be indicative of a very important ability in the area of perspective-taking that makes children successful in strategic interactions. Further studies should shed light on the exact underlying skills or dispositional characteristics that are indicated using this simple behavioral task. Likewise, the Social Appropriateness score, which has mainly been used to assess school-aged children's social and emotional competencies with a focus on empathy (e.g., Schick and Cierpka 2005), is positively associated with successful game performance. However, the relationship is smaller in size. Still, this score seems to correlate reasonably strong with successful performance and could, thus, be of interest for further research with children in this area. The Interpersonal Reactivity Index (IRI) is associated with successful performance to a greater extent and it is highly significant in most of our specifications. Yet, the relationship for the IRI is negative, indicating that high levels of interpersonal reactivity can harm successful performance in the BCG. Why is this the case? In contrast to the other two perspective-taking measures, the IRI score is already significantly linked to choices in the first round and also first-round performance. However, this relationship seems strongly driven by choices above 50: excluding weakly dominated choices eliminates the link between IRI and higher numbers in the first round. In addition, raw correlations between IRI scores and choosing a weakly dominated number in round 1 ($\rho = .25, p = .007$) confirm that children scoring high on the IRI have a higher probability of choosing a number above 50 in the first round. However, this is not the case for higher numbers in general: for example, the correlation between IRI scores with a dummy for choosing numbers higher than 30 is zero. We can only speculate on why the relationship between IRI scores and performance is negative. One potential explanation could be that the IRI is indicative of children who focus on other aspects of the game (or on the other children's behavior in the game) that do not improve performance and occupy cognitive resources. Another hypothesis is that children scoring high on the IRI focus too much on *past* behavior instead of focusing on adjusting the number to be chosen in the right way.

Taken together, our measures of perspective-taking abilities are strongly linked to successful performance in an experimental BCG, although in different directions. Further studies should look more closely at the relationship between this set of abilities and performance in strategic interaction settings, as well as potential interaction effects of cognitive skills and perspective-taking abilities.

E Adult Sample

Table A12: Sample Descriptives (Adults)

	Mean	SD	N	Min	Max
Female	0.60	0.49	120	0	1
Age in years	22.63	4.18	120	18	46
Fluid IQ	0.00	1.00	120	-3.12	1.78

Notes: Sample descriptives for the adult sample. The study was conducted in 12 sessions with 10 participants each (two groups of five adults per session).

Table A13: Game Descriptives (Adults)

	Mean	Median	SD	N	Min	Max
Number Round 1	23.64	21	14.92	120	1	99
Number Round 2	13.47	11	9.00	120	0	47
Number Round 3	8.31	7	7.33	120	0	50
Number Round 4	6.64	4	9.55	120	0	58
Number Round 5	4.83	3	9.21	120	0	90
Number Round 6	3.13	2	6.70	120	0	60
Number Round 7	2.75	1	8.42	120	0	87
Number Round 8	1.59	1	2.74	120	0	17
Number Round 9	1.54	1	5.20	120	0	50
Number Round 10	1.12	0	4.42	120	0	46
Mean Number 1–10	6.70	5.9	4.08	120	1.2	25

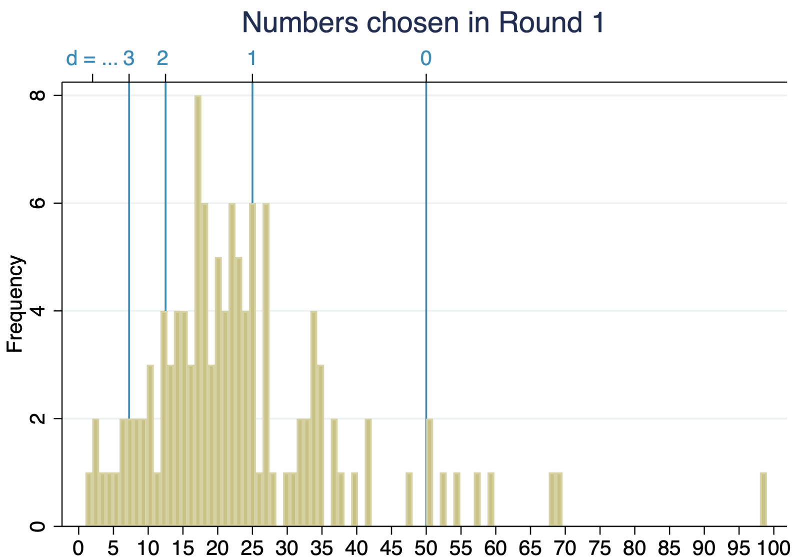
Notes: Descriptives for the numbers chosen in each round of the game for the adult sample.

Table A14: OLS Regressions with Coins Won, Rank, and Distance (Adults)

	Coins R2–10	Rank R2–10	Distance R2–10
Female	-1.282***	0.444**	0.158
	(0.440)	(0.203)	(0.750)
Age in years	-0.040	0.009	0.048
	(0.049)	(0.018)	(0.090)
Fluid IQ	0.183	-0.083	0.026
	(0.276)	(0.122)	(0.295)
Group FE	Yes	Yes	Yes
Observations	120	120	120
R-squared	0.441	0.221	0.180

Notes: The results are based on OLS regressions. Standard errors in parentheses are clustered at the group level. * $p<.10$, ** $p<.05$, *** $p<.01$

Figure A3: Histogram of Numbers Chosen in Round 1 (Adults)



Notes: This figure shows a histogram for the numbers chosen in round 1 in the adult sample. The blue vertical lines display resulting values for a depth of reasoning (d) of 0, 1, 2, and 3 (starting from an initial reference point of 50). See Section 3.1 for details on depth of reasoning.

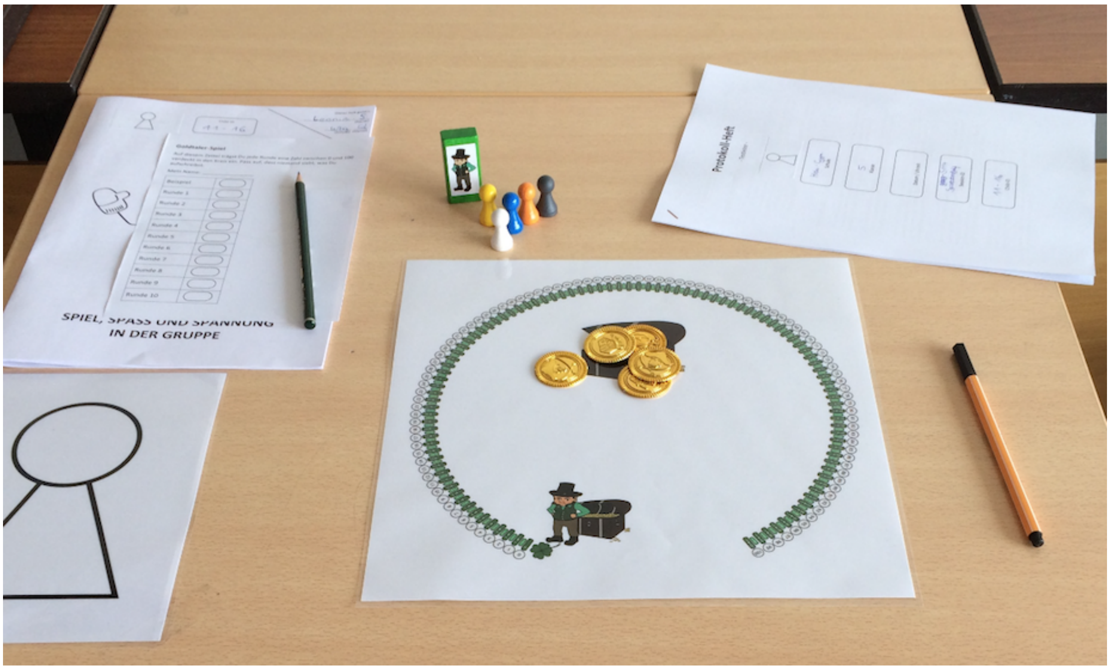
F The Beauty-Contest Game

F.1 Setup of the Game

Figure A4: Experimental Setup of the "Goblin Game"

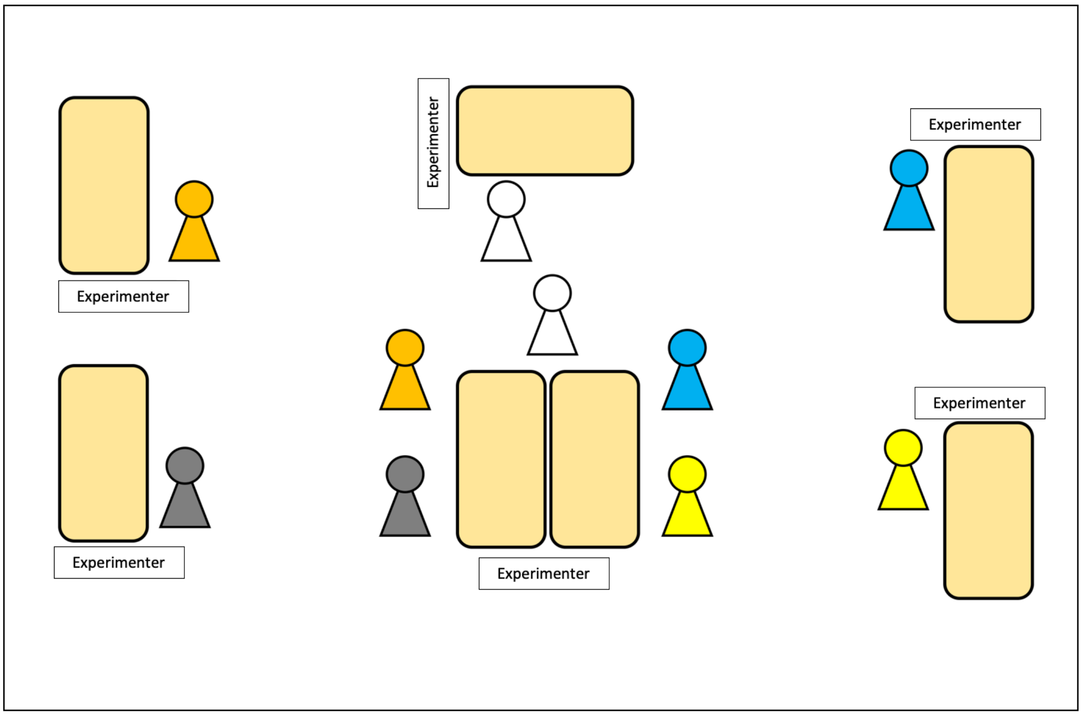
Notes: This is the setup for the main table at which the board game was played. The experimenter guiding the children through the 10 rounds of the game sat at the bottom center of the table. The five other seating positions are marked with one of the five colors (gray, orange, white, blue, and yellow). A note with a reminder about the five steps of the game is placed at every seating position. In the center, there is the actual board game with the goblin, the five colored pawns, and the treasure box filled with gold coins. The board measured 59x59 cm (about 23x23 in), so children could read everything but could also reach everything on the board. In the background, there is one of the five small tables used to explain the game to children in a one-to-one setting (see Figure A5 for details).

Figure A5: Material Used in the "Goblin Game"



Notes: This is the setup for one of the five tables at which the game was explained to the children in a one-to-one setting with an experimenter. In the center, there is a mini-version of the board game used for the explanation, including some gold coins, the green goblin figure, and the five colored pawns. At the bottom left is the indicator that this is the table for the white player (this ensured that the right child would come to the right experimenter—experimenters rotated over colors, so that they would explain the game to a different color for every group). At the top left is the game slip for the child on which he/she wrote down the numbers for each round. Below that is the workbook containing the questionnaires. At the top right is the script used by the experimenter to explain the game to the child (see Section F.2)

Figure A6: Plan for the Room in which the "Goblin Game" Was Played



Notes: This is the plan used to prepare the separate room in which the experimental BCG was played with groups of five children. The separate rooms in all schools were set up following this plan to standardize the setting for all children across schools. There were five small individual tables used to explain the game to children in a one-to-one setting with an experimenter. The main experimenter always sat at the marked position on the large table to guide the children through the game (see Section 2.3 for details about the procedures).

F.2 Instructions for Experimenters

(The following instructions were used as a script by each experimenter during the one-to-one explanation of the game. Each experimenter had these instructions in print in front of him or her and followed the script exactly, step by step; see Figure A5.)

Information for the Experimenter/Instruction Reader

- Show no reactions to the strategies/ideas/suggestions of the children. Only answer specific questions by referring to the rules.
- Example
  Child: "I should never set a figure on numbers greater than 50, right?"
  You: "You can set your figure on any number between 0 and 100. The player who is closest to the goblin at the end of the round will win."
- No explicit suggestions, including nonverbal suggestions, should be given regarding potential game strategies.
- All bold solution parts of the rules should be mentioned, missing aspects should be recorded.

Game Instructions: The Goblin Game

We are going to play a game, in which the goal is to win as many gold coins as possible. At the end of the game you can trade the gold coins for toys. The more gold coins you have at the end of the game, the wider choice of toys you will have. So, try hard! ☺

We will play in groups of five. Every player will get a pawn, you have the color [*NAME THE COLOR and point to the figure*]. There are also the colors [*name the other colors and point to the pawns*].

Other than the five players, there is the goblin [*point to the figure*]. The goblin has hidden gold coins in the forest. But he is a nice goblin and will help you find the gold coins. He always reveals the location of the gold coins to the middle player. However, the goblin is hexed: for every step he takes, he must go back half a step.

The goblin will show you the way to the gold coins. Therefore, you must be **as close as possible to the goblin** after every round. The player who is closest to the goblin after the round [*point to the figure closest to the goblin for clarification*], will win a gold coin [*point to the gold coin in the treasure box*]. If two players are equally far away from the goblin, both of these players will get a gold coin.

In total, we will play 10 rounds, so it is possible to win 10 gold coins. Look, here on your game slip you can see fields for the 10 rounds [*give the players the game slip and show them the 10 rounds*]. Write your name on the top of the page [*let them write their name*].

Good job! Every round will work in the same way, in other words, every round will contain five steps [*accordingly count the steps with your fingers*]:

Step 1:

Every player will write down a secret number between 0 and 100 on his or her paper.

For the first round, this number will be written in the circle after round 1, in the second round in the circle after round 2, and so on. When you are finished writing down your number, you should simply put your playing figure over the number. This way, the other players cannot see what you wrote down.

It is important that the others do not see what number you have written down. Maybe you can hold your hand over what you are writing. Now, write down a number between 0 and

100 in the example field! [*let the child write down a number in the example field of the game slip and write down a number yourself*].

Step 2:

When all players are done writing down a number, everybody can set their figure on the respective field. Of course, you should set your figure on the number you wrote down.

Therefore, we will use the fields on the board game [*point to the fields, so that they can distinguish the goblin fields from the player fields*]. There are fields between 0 and 100 [*point to the fields*]. Therefore, the number that you wrote down in step 1 must be between 0 and 100. Every player will set his/her figure on the respective number that he/she has written down. Let's go! [*now let the child set his / her figure on the field and set the other figures on the fields, 7, 24, 45, and 79—do not let any two figures be set on the same field; if this happens, set them aside from each other*].

Step 3:

Then, the goblin comes. He always runs from his field along the **goblin fields** [*show which ones these are*]. In order to reveal where the gold coins are hidden, he runs to the **middle player**. That is also the third player, if you count from the beginning. Now you can move the goblin! [*count out loud (1, 2, and 3), while the child moves the goblin along the fields*].

Step 4:

Now comes the goblin jump. Because the goblin is hexed, he has to **go back half a step** for every step he takes. So the goblin always jumps back to the field that is **half the number** of the field on which he originally stood. Half of each number is written down on the goblin fields [*point to the respective field, the arrow and the half value*], so that you do not have to calculate. So, the **goblin jumps back by half of the middle player's number** [*move the goblin figure and read the numbers out loud while doing so*].

Step 5:

Now we look: **Who is closest to the goblin?** This player **wins a gold coin**. If two players are the same distance from the goblin, **both** these players win a gold coin. Who wins a gold coin this time? [*let the child identify the winner*].

Very good! Then let's set all the figures back and the next round can begin. Altogether, we will play 10 rounds.

So, again to summarize:

The point of the game is to win as many gold coins as possible. The player who is closest to the goblin after each round wins a gold coin. Each round has five steps [*again count the steps with your fingers while explaining*]: First, each player secretly writes down a number. Then, each player sets his/her figure on that number. Next, the goblin runs to the third player. The goblin then jumps back by half of the number chosen by the third player. The player who is now closest to the goblin gets a gold coin. Then, the next round starts.

Do you have any questions on the rules of the game?

TESTING UNDERSTANDING OF THE GAME

We will go through the game together one more time. You get to explain the steps of the game to me. If you cannot remember something, no problem—I will gladly explain it to you again. So: [*Read out loud (without the headlines "Question X"). If necessary, explain the step* ***one more time****, otherwise don't mind and encourage the child by saying "you will surely see how it works during the game!"*]

Question 1: Can you please explain one more time what happens first in every round?

Answer: All players **secretly** write down, without letting anybody else see, a **number** between **0 and 100** on their game slip. The players **set their pawn on the secret number** that they have written down in order to cover it. This means that the players are done writing down their number.

☐ Immediately completely and correctly explained

☐ With a hint completely and correctly explained

☐ After a repeated explanation, completely and correctly explained

☐ Not completely understood, the following is missing: ...

Question 2: What happens after all players have written down their secret number?

Answer: When all players have written down their number (not before!), **the players simultaneously set their figures on the field** (on the outer fields on the board game) according to **the number written on the game slip**.

☐ Immediately completely and correctly explained

☐ With a hint completely and correctly explained

☐ After a repeated explanation, completely and correctly explained

☐ Not completely understood, the following is missing: ...

Question 3: What happens after all players have set their figures on the board?

Answer: Then the **goblin** runs over the **goblin fields** up to the **third/middle player**.

☐ Immediately completely and correctly explained

☐ With a hint completely and correctly explained

☐ After a repeated explanation, completely and correctly explained

☐ Not completely understood, the following is missing: ...

Question 4: What happens when the goblin is at the middle player?

Answer: The **goblin jumps**. This means, that he jumps back **by half of the number on the field** (the number to which the goblin must jump is indicated on his original field, behind the arrow).

☐ Immediately completely and correctly explained

☐ With a hint completely and correctly explained

☐ After a repeated explanation, completely and correctly explained

☐ Not completely understood, the following is missing: ...

Question 5: What happens after the goblin has jumped back?

Answer: The player who **is closest to the goblin** wins **a gold coin** (if two players are the same distance from the goblin, both these players win a gold coin). Then, **all figures are returned** to the players and **a new round** begins.

☐ Immediately completely and correctly explained

☐ With a hint completely and correctly explained

☐ After a repeated explanation, completely and correctly explained

☐ Not completely understood, the following is missing: ...

In Conclusion: You did a great job! The game is about to start—but first, I will give you **a gold coin** as a thank you!

G Measures for Perspective-taking Abilities

G.1 Perspective-taking Task

(This is the script used by the experimenter to explain the perspective-taking task "E on the forehead". The task is conducted right after the child has received the gold coin for the explanation of the rules of the game; see previous page. The experimenter reads the following text out loud and records the observed behavior right away.)

"Before we start, I have one small task for you.

Please trace, as fast as possible, with your forefinger, the capital letter 'E' on your forehead."

[*Repeat the instructions only once more if necessary. If the child does not understand, encourage him/her to take his/her seat at the group table.*]

The 'E' is ...

☐ From my (experimenter) perspective reversed

☐ From my (experimenter) perspective legible

☐ Neither, the child traced something else

☐ The child did not understand the task, other

"Very good job—now you can go to your spot at the group table. Have fun playing the game!"

G.2 Social Appropriateness Scale

(This questionnaire was adapted from Meindl (1998). It was filled out during the first lesson by all children in the classroom. Questions were read out loud by the experimenter and children filled out the questionnaire in an individual workbook.)

<u>Situation 1: The Camera</u>

Fritz met his friend Jochen on the street and showed him the new camera his parents gave him. Jochen asked Fritz if he could try the camera. While trying to take a picture with the camera, Jochen tripped. The camera fell down and broke.

Question A: How does Fritz feel when he sees that the camera is broken?

☐ 1. He is mad because his camera is broken.

☐ 2. He does not care because he will surely get another camera from his parents.

Question B: How does Jochen feel?

☐ 1. He feels guiltless, because he did not mean to break the camera on purpose.

☐ 2. He is embarrassed that he broke the camera.

Question C: How would you react if you were Fritz?

☐ 1. I would yell at Jochen because he should have been more careful with the camera.

☐ 2. I would tell Jochen that I am upset, but not mad at him, because he did not break the camera on purpose.

Question D: How would you react if you were Jochen?

☐ 1. I would apologize.

☐ 2. I would tell Fritz that he should not be mad because I did not break the camera on purpose.

Situation 2: The Computer

Jürgen wants for a computer for his birthday. However, his parents do not have enough money and give him something else instead.

Question A: How does Jürgen feel when he sees that he did not get a computer?

☐ 1. He does not mind because he received something else instead.

☐ 2. He is disappointed.

Question B: How does not fulfilling Jürgen‘s wish feel to his parents?

☐ 1. They feel sorry that they cannot fulfill his wish.

☐ 2. They do not care because they would not have had enough money to buy the computer anyway.

Question C: How would you react if you were Jürgen?

☐ 1. I would complain loudly to my parents that I would have rather had a computer.

☐ 2. I would try not to show my disappointment and rejoice over the other gift I got.

Situation 3: The Horror Film

Susanne would like to watch a horror film later in the evening. However, her father does not allow this and sends her to bed, with the reasoning that she is still too young.

Question A: What does her father think, when he says that Susanne is still too young?

☐ 1. He thinks that if Susanne watches the film, she would get very scared.

☐ 2. He wants to upset Susanne.

Question B: How does not being allowed to watch the film make Susanne feel?

☐ 1. She feels sad.

☐ 2. She feels mad.

Question C: What would you do if you were Susanne?

☐ 1. I would yell at my father because he is so mean to me.

☐ 2. I would try to talk to him about it again.

Situation 4: The Dishes

Sebastian is about to leave the house because he has arranged to play soccer with his friends. However, his mother asks him to wash the dishes as she still has a lot to do.

Question A: How does Sebastian feel when he hears that he should wash the dishes?

☐ 1. He is sad.

☐ 2. He is mad.

Question B: How does his mother feel?

☐ 1. She is stressed from all the work.

☐ 2. She just does not want Sebastian to go and play soccer.

Question C: What would you do if you were Sebastian?

☐ 1. I would go and play soccer because it was already arranged.

☐ 2. I would wash the dishes and then go to soccer later.

Situation 5: Teasing

Markus constantly gets teased by his classmates because he stutters. Doris joins the class as a new student. She notices that Markus gets teased by everybody and also joins in the teasing.

Question A: Why does Doris tease Markus?

☐ 1. Because Markus stutters.

☐ 2. Because she wants to be accepted by the others.

Question B: How does Markus feel when he gets teased?

☐ 1. He is sad and feels excluded.

☐ 2. He does not take it seriously.

Question C: How would you react if you were Doris?

☐ 1. I also would have joined in the teasing.

☐ 2. I would have refrained from teasing.

Situation 6: The Best Grade

After school, Michael tells his friend Peter that he got an A in math. Peter, who got a C, says to Michael: “You’re a stupid nerd”.

Question A: Why does Peter say this?

☐ 1. He is envious of Michael.

☐ 2. He does not like Michael.

Question B: How does Michael feel thereafter?

☐ 1. He is hurt, because Peter offended him.

☐ 2. He finds Peter’s behavior ridiculous.

Question C: How would you react if you were Michael?

☐ 1. I would tell Peter that he should not exaggerate like that.

☐ 2. I would tell Peter that I am sorry that he only got a C, but that he does not need to offend me.

G.3 Interpersonal Reactivity Index

(This questionnaire uses the items for cognitive empathy from Garton and Gringart (2005). It was filled out during the first lesson by all children in the classroom. Questions were read out loud by the experimenter and children filled out the questionnaire in an individual workbook.)

1. I think people can have different opinions about the same thing.

☐ **does not** apply **at all**

☐ **does not generally** apply

☐ **sometimes** applies

☐ **generally** applies

☐ **fully** applies

2. When I am angry or upset at someone, I usually try to imagine what he or she is thinking or feeling.

☐ **does not** apply **at all**

☐ **does not generally** apply

☐ **sometimes** applies

☐ **generally** applies

☐ **fully** applies

3. When I am arguing with my friends about what we are going to do, I think carefully about what they are saying before I decide whose idea is best.

☐ **does not** apply **at all**

☐ **does not generally** apply

☐ **sometimes** applies

☐ **generally** applies

☐ **fully** applies

4. I sometimes try to better understand my friends by pretending I am them.

☐ **does not** apply **at all**

☐ **does not generally** apply

☐ **sometimes** applies

☐ **generally** applies

☐ **fully** applies

H References Online Appendix

Davis, Mark H. 1983. "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social Psychology* 44 (1): 113–126.

Galinsky, Adam D., Joe C. Magee, M. Ena Inesi, and Deborah H Gruenfeld. 2006. "Power and Perspectives Not Taken." *Psychological Science (0956-7976)* 17 (12): 1068–1074.

Garton, Alison F., and Eyal Gringart. 2005. "The Development of a Scale to Measure Empathy in 8- and 9-Year Old Children." *Australian Journal of Educational & Developmental Psychology* 5:17–25.

Glen, R. 1984. "Perspective taking and self-awareness: Drawing an E on your forehead." *Journal of Personality and Social Psychology* 46 (4): 788–798.

Meindl, C. 1998. "Entwicklung und Validierung eines Fragebogens zur Erfassung von Empathie und seine Anwendung im Rahmen eines Empathietrainings." *mimeo.*

Schick, Andreas, and Manfred Cierpka. 2005. "Faustlos: Evaluation of a curriculum to prevent violence in elementary schools." *Applied and Preventive Psychology* 11 (3): 157–165.

Steins, Gisela, and Robert A. Wicklund. 1996. "Perspective-Taking, Conflict, and Press: Drawing an E on Your Forehead." *Basic and Applied Social Psychology* 18 (3): 319–346.

Issued in the series Discussion Papers 2020

2020

01/20 January, **Laura Khoury**, Clément Brébion and Simon Briole. "Entitled to Leave: the impact of Unenployment Insurance Eligibility on Employment Duration and Job Quality"

02/20 January, Thomas Buser, **Alexander Cappelen**, Uri Gneezy, Moshe Hoffman and **Bertil Tungodden**. "Competitiveness, gender and handedness: a large-sample intercultural study"

03/20 February, **Patrick Bennett**, Chiara Ravetti and Po Yin Wong. "Losing in a Boom: Long-term Consequences of a Local Economic Shock for Female Labour Market Outcomes"

04/20 April, **Øivind A. Nilsen**. "The Labor Market in Norway: 2000-2018"

05/20 April, **Simen A. Ulsaker**. "Exclusionary contracts and incentives to innovate"

06/20 May, **Alexander W. Cappelen**, **Ranveig Falch**, **Erik Ø. Sørensen** and **Bertil Tungodden**. "Solidarity and Fairness in Times of Crisis"

07/20 May, Gozde Corekcioglu, Marco Francesconi and **Astrid Kunze**. "Do Generous Parental Leave Policies Help Top Female Earners?"

08/20 June, **Ola Honningdal Grytten**. "Weber revisited: A literature review on the possible Link between Protestantism, Entrepreneurship and Economic Growth"

09/20 June, Eva M. Berger , Ernst Fehr, **Henning Hermes** , Daniel Schunk and Kirsten Winkel. "The Impact of Working Memory Training on Children's Cognitive and Noncognitive Skills"

10/20 June, **Ola Honningdal Grytten**. "Two centuries of economic growth: Norwegian GDP 1816-2020"

11/20 July, **Ola Honningdal Grytten**, Magnus Lindmark and Kjell Bjørn Minde. "Energy Intensity and the Environmental Kuznets Curve"

12/20 August, **Ola Honningdal Grytten**. "Puritan Motivation for Serial Entrepreneurship: The Haugean Example"

13/20 August, Julian Johnsen, Hyejin Ku and **Kjell G. Salvanes**. "Competition and Career Advancement: The Hidden Costs of Paid Leave"

14/20 August, Patrick Bennett, Richard Blundell and **Kjell G. Salvanes**. "A Second Chance? Labor Market Returns to Adult Education Using School Reforms"

15/20 August, Paul Brandily, Clément Brébion, Simon Briole and **Laura Khoury**. "A Poorly Understood Disease? The Unequal Distribution of Excess Mortality Due to COVID-19 Across French Municipalities"

16/20 September, **Ingvild Almås**, **Vincent Somville** and Lore Vandewalle. "The Effect of Gender-Targeted Transfers: Experimental Evidence From India"

17/20 September, **Ola Honningdal Grytten.** "The Wealth of a Nation: Norways Road to Prosperity"

18/20 September, Asbjørn G. Andersen, Simon Franklin, Tigabu Getahun, Andreas Kotsadam, **Vincent Somville** and Espen Villanger. "Does Wealth Reduce Support for Redistribution? Evidence from an Ethiopian Housing Lottery"

19/20 September, **Ingvild Almås**, Lars Ivar Berge, **Kjetil Bjorvatn**, **Vincent Somville** and **Bertil Tungodden**. "Adverse selection into competition: Evidence from a large-scale field experiment in Tanzania"

20/20 September, Julian Vedeler Johnsen, Kjell Vaage and **Alexander Willén**. "Interactions in Public Policies: Spousal Responses and Program Spillovers of Welfare Reforms"

21/20 October, **Aline Bütikofer**, Rita Ginja, **Fanny Landaud** and **Katrine Løken**. "School Selectivity, Peers, and Mental Health"

22/20 November, Barton Willage and **Alexander Willén.** "Postpartum Job Loss: Transitory Effect on Mothers, Long-run Damage to Children"

2021

01/21 January, Sebastian Axbard and **Zichen Deng**. "Informed Enforcement: Lessons from Pollution Monitoring in China"

02/21 January, Richard Baldwin, **Jan I. Haaland** and Anthony J. Venables. " Jobs and technology in general equilibrium: A three-elasticities approach"

03/21 January, **Thor Andreas Aursland**, **Frode Steen** and **Simen A. Ulsaker**. "Unemployment shocks, cyclical prices and shopping behavior"

04/21 January, **Aline Bütikofer**, **René Karadakic** and **Kjell Gunnar Salvanes**. "Income Inequality and Mortality: A Norwegian Perspective"

05/21 February, **Ranveig Falch**. "How Do People Trade Off Resources Between Quick and Slow Learners?"

06/21 March, **Charlotte B. Evensen**, Øystein Foros, Atle Haugen and **Hans Jarle Kind**. "Size-based input price discrimination under endogenous inside options"

07/21 March, Samuel Dodini, Michael Lovenheim and **Alexander Willén.** "Understanding the Decline in Private Sector Unionization: A Skill-based Approach"

08/21 March, Thomas Buser, **Alexander Cappelen** and **Bertil Tungodden**. "Fairness and Willingness to Compete"

09/21 March, **Marc Goñi** and David de la Croix. "Nepotism vs. Intergenerational Transmission of Human Capital in Academia (1088-1800)"

NHH

NORGES HANDELSHØYSKOLE
Norwegian School of Economics

Helleveien 30
NO-5045 Bergen
Norway

T +47 55 95 90 00
E nhh.postmottak@nhh.no
W www.nhh.no