Accepted Manuscript

Title: How strategic are children and adolescents? Experimental evidence from normal-form games

Author: Simon Czermak Francesco Feri Daniela Glätzle-Rützler Matthias Sutter

PII:	S0167-2681(16)30044-0
DOI:	http://dx.doi.org/doi:10.1016/j.jebo.2016.04.004
Reference:	JEBO 3779
To appear in:	*Journal of Economic Behavior & Organization*
Received date:	25-10-2012
Revised date:	8-1-2016
Accepted date:	5-4-2016

Please cite this article as: Czermak, Simon, Feri, Francesco, Glätzle-Rützler, Daniela, Sutter, Matthias, How strategic are children and adolescents? Experimental evidence from normal-form games.Journal of Economic Behavior and Organization http://dx.doi.org/10.1016/j.jebo.2016.04.004

How strategic are children and adolescents? Experimental evidence from normal-form games*

Simon Czermak
Management Center Innsbruck

Francesco Feri
Royal Holloway University of London

Daniela Glätzle-Rützler
University of Innsbruck

Matthias Sutter#
University of Cologne and University of Innsbruck and IZA Bonn

* We are indebted to Director Gerhard Sailer of the "Öffentliche Gymnasium der Franziskaner Hall" for making this study possible. Furthermore we would like to thank an associate editor and two referees for helpful comments. Financial support from the Hypo Tirol Bank AG (Forschungsförderungspreis der Hypo Tirol), the Austrian Central Bank (through Jubiläumsfonds grant 14680) and the Austrian Science Fund (through grant P22772) is gratefully acknowledged.

Corresponding author. Address: Department of Economics, University of Cologne, Albertus Magnus Platz, D-50923 Cologne, Germany. e-mail: matthias.sutter@wiso.uni-koeln.de

Highlights of the paper

How strategic are children and adolescents? Experimental evidence from normal-form games

- We examine the strategic sophistication of teenagers, aged 10 to 17 years.
- Subjects play two-person normal-form games.
- We elicit choices, first-order beliefs and second-order beliefs.
- Strategic sophistication does not change with age.
- We estimate that about 40% of subjects are strategic decision makers.

Abstract

We examine the strategic sophistication of 196 children and adolescents, aged 10 to 17 years, in experimental normal-form games. Besides choices, we also elicit first- and second-order beliefs. The share of subjects playing Nash or expecting opponents to play Nash is fairly stable across all age groups. The likelihood of playing best response to own beliefs increases in math skills. Using a mixture model, about 40% of subjects are classified as a strategic type, while the others are non-strategic. The distribution of types is somewhat changing with age. The estimated error rates also show some dependency on age and gender.

JEL-classification: C72, C91

Keywords: Strategic thinking, beliefs, experiment, age, adolescents

This version: 2 January 2016

1. Introduction

Standard game theory is an important pillar of research in economics, and more generally in the social sciences, since it provides a tool to analyze strategic interaction, like interaction in markets, bargaining, or in social dilemma situations. In this framework strategic sophistication refers to the extent to which a decision maker takes the partners' possible actions in a strategic situation into account and consequently chooses an optimal strategy for herself (Crawford et al., 2013). Strategic sophistication is therefore considered as advantageous for a decision maker when interacting with others.

Interestingly, numerous previous studies with adult decision makers have shown that the ability to act strategically is often limited (see, e.g., Stahl and Wilson, 1994, 1995; Haruvy et al., 1999; Costa-Gomes et al., 2001; Weizsäcker, 2003; Bhatt and Camerer, 2005; Crawford and Iriberri, 2007; Costa-Gomes and Weizsäcker, 2008; Rey-Biel, 2009; Danz et al., 2012; Sutter et al., 2013). The frequently observed failure to behave strategically in a sophisticated way is typically blamed on an ignorance of other players' incentives and their rationality (Costa-Gomes et al., 2001, Weizsäcker, 2003) or on a lack of consistency in best-responding with one's actions to own first-order-beliefs. For instance, Costa-Gomes and Weizsäcker (2008) report the latter failure to happen in about fifty percent of cases (in an experiment with university students). While the literature on strategic sophistication has developed models to explain this surprisingly large degree of bounded rationality (see Crawford et al., 2013, for a recent survey), it has largely remained silent on the development of strategic sophistication with age. Given the high degree of non-strategic behavior of adults in strategic interaction games, one might wonder how children and adolescents act in such situations.

In his textbook on behavioral game theory, Camerer (2003, p. 66) notes that the behavior of children is often "*closer* to the self-interest prediction of game theory than virtually any adult population". This would imply more equilibrium play of younger subjects. One should note, however, that equilibrium play need not be identical to strategic sophistication, especially in situations of a prisoners' dilemma in which cooperation of both players may result in a Pareto superior outcome. Moreover, it does not follow immediately from Camerer's statement that younger subjects also might have a better anticipation of their opponent's behavior and thus be more sophisticated in strategic interaction. From a developmental psychology point of view, one might actually favor the following conjecture: If one subscribes to the view that children and adolescents need to learn through education and socialization how to behave strategically in human interaction (a view suggested by developmental psychology; Kail and Cavanaugh, 2010), then one would expect to see a positive influence of age on the level of strategic sophistication. Given opposing conjectures, this issue is open for empirical investigation.

In this paper, we present an experiment with 196 children and adolescents, aged 10 to 17 years. They have to make decisions in 18 different normal-form games that have been designed by Costa-Gomes et al. (2001) to study strategic sophistication. These games have been designed with different levels of complexity (various patterns of iterated dominance and unique pure-strategy equilibria without dominance). This complexity allows a fine-grained estimation of strategic and non-strategic types, explained in more detail below. Hence, this design improves on previous papers on children's ability to play games in general. For instance, Brosig-Koch et al. (2015) study the development of the ability to use backward induction in simple games. They study the behavior of 6- to 23-year old subjects in a race game where backward induction is necessary to find out the winning strategy. In fact, before the teenage years, children improve on the ability to apply backward induction, an ability that is useful in playing the normal-form games used in this paper. However, Brosig-Koch et al. (2015) are not able – nor was it their intention – to examine the distribution of strategic and non-strategic types in playing normal-form games. We are also not so much interested in the question whether teenagers play these games as adults do (although we draw a few comparisons further down the line). This is something which, for example, work in psychology has addressed. Weller et al. (2012) show that 10 to 11 year olds make similar decisions as adults in a variety of contexts (like in sunk cost tasks or in the perception of risk), but their work does not focus on how choices, first- and second-order beliefs relate to each other and what follows from that for the level of strategic sophistication in playing interactive games. Moreover, they do not provide an estimation of the relative frequency of strategic and non-strategic types of decision makers in normal-form games.

We find that subjects play the Nash equilibrium strategy in about 45% of all cases, with no significant differences across the age range from 10- to 17-year olds. The relative frequency of equilibrium play is well in the range of 40% to 50% which is the typical finding in previous studies with adults (e.g., Costa-Gomes et al., 2001; Costa-Gomes and Weizsäcker, 2008; Sutter et al., 2013)[1], indicating that strategic decision making in our subject pool is largely similar to the level found for adults.

[1] In Sutter et al. (2013) the same games are used as in this study, but the sequence of eliciting choices, first- and second-order beliefs is slightly different from here. The focus in Sutter et al. (2013) is on comparing

In addition to eliciting choices in the normal-form games, we also elicit first-order beliefs (about the opponent's choice) and second-order beliefs (about the opponent's first-order belief). The elicitation of beliefs allows us to examine a decision maker's consistency in two ways: (i) We analyze whether a decision maker chooses a best response to the stated first-order belief. This happens in about 60% of cases. (ii) We examine whether a decision maker expects the opponent to act rationally by first-order beliefs being a best response to second-order beliefs. This happens at a lower level of around 50% of cases. We denote this as the opponent's expected consistency. The analysis of the determinants of consistency reveals that the probability of consistent choices (by best replying to first-order beliefs) is significantly lower when the opponent player has a dominant strategy. In general, students with better grades in mathematics are more likely to act consistently.

Applying the mixture model of Costa-Gomes et al. (2001), we classify subjects into eight different strategic or non-strategic types. While previous literature on the economic behavior of children and teenagers has studied game theoretic settings and whether subjects play equilibrium (see, among many others, Murnighan and Saxon, 1998, Sutter and Kocher, 2007, or Brosig-Koch et al., 2015), no previous paper has classified children's economic behavior as any of eight different types of strategic and non-strategic behavior. This means that we can provide a finer-grained picture of strategic (or non-strategic) sophistication of children and teenagers, allowing us to study in more detail their reasoning and thinking. We find that the majority of our subject pool (almost 60%) can be classified as non-strategic decision makers. The modal type is a non-strategic *Optimistic* type which plays the strategy that maximizes the maximum possible payoff in ignorance of the partner's payoffs. Roughly 40% of subjects are classified as strategic. We note a different type distribution between men and women. Men are more often (non-strategically) maximizing the total surplus, while women are more likely to be a type that eliminates dominated strategies. Age has some weak, gender specific effect on the probability to belong to a specific type. Eliminating dominated strategies becomes more frequent with age in the female sample, showing that this aspect of strategic thinking gets slightly more frequent with age, while maximizing the best possible own payoff – a non-strategic endeavour – decreases with age for males. The error rates, with which subjects are classified to be any of five different strategic or three non-strategic types also differ between men and women (with males having significantly lower error rates for two types), and depend somewhat on age. We also find that females with better math grades are more likely to be

individual choices to team choices (in an adult subject pool). One of the main results is that teams are more strategically sophisticated than individuals. When we refer to this study here, we refer to individual choices only.

classified as strategic, indicating that subjects with better analytical skills (in math) behave differently in strategic interaction.

The rest of the paper is organized as follows: In section 2 we present the experimental design. Section 3 reports the experimental results and section 4 concludes the paper.

2. Experimental design

Our experimental design is based on the 18 normal-form games that were designed by Costa-Gomes et al. (2001) to study strategic sophistication (see Figure 1). There are 10 "**D**"-games in which one of the two players has a strictly dominant strategy and 8 "**ND**"-games in which no player has a dominant strategy. In all games there is a unique pure Nash equilibrium that is Pareto-dominated by another strategy combination of row and column players. The games can also be classified according to the number of rounds of iterated pure-strategy dominance that players need to identify the equilibrium strategy. This refers to a game's complexity. In **D**-games the number of rounds a player needs to reach his own equilibrium choice is either 1 or 2, while in **ND**-games the corresponding number of rounds may be 2, 3 or infinite. Figure 1 illustrates all games' types, complexities and the order of presentation of the specific games to the participants.[2] The total of 18 games includes eight pairs of isomorphic games that are identical for row and column players except for transformation of player roles and small, uniform payoff shifts.

Figure 1 about here

Each game was played only once and in each game subjects had to make three consecutive decisions:[3] A subject had to (i) choose one of the available strategies (**choice**), (ii) state a first-order belief about the opponent's action (**FOB**) and, finally, (iii) state a second-order belief, i.e., a belief about the opponent's first-order belief (**SOB**).[4]

[2] This information can be found in the upper left and right corner of each game in Figure 1. The order of games in the experiment is indicated by "game # x", with x $\in$ {1, ..., 18}. (D) refers to **D**-games and (ND) refers to **ND**-games. Numbers in brackets indicate a game's complexity for the row and column player, where [xR, yR] denotes the number of rounds needed for the row (x) and the column player (y). In italics we denote the strategy predicted by the eight different (strategic and non-strategic) types that we are estimating in section 3.4.

[3] Games were presented to all participants in the same order and in a way that they saw themselves as a row player (as in Costa-Gomes et al., 2001). The transformation of the column players' perspective didn't have an influence on the characteristics of the games and was intended to avoid any influence of the kind of presentation on behavior.

[4] We elicited point beliefs for two reasons. First, asking for a probability distribution (and paying in an incentive compatible way by using, for example, a quadratic scoring rule) was too complicated for our youngest

Our subjects didn't get any feedback until all participants had taken all 54 decisions. This was done in order to suppress learning and reputation formation. At the end of the experiment, one subject in each pair of players was asked to draw a card from a deck of cards showing numbers from 1 to 18. The drawn number determined the game that was relevant for payment (for both players). In a next step the subjects learnt about the own and the opponent's choice, first- and second-order beliefs in the particular game. A second card drawn by the other subject in each pair of players then determined whether choices (card A), first-order beliefs (card B) or second-order beliefs (card C) were paid.

The experiment was run at the "Öffentliche Gymnasium der Franziskaner Hall", a public high school located 5 km east of Innsbruck, the capital of the state of Tyrol in Austria. This school teaches children in 8 different grades, in Anglo-Saxon terminology grades 5 to 12. We conducted the experiment in grade 5 (10- to 11-year olds), grade 7 (12- to 13-year olds), grade 9 (14- to 15-year olds), and grade 11 (16- to 17-year olds), with two classes in each grade. Table 1 presents the number of subjects in each grade, broken down also by row- and column-player and by gender.

Table 1 about here

The sessions in both classes of a given grade were always run simultaneously in two separate rooms in order to avoid any potential dissemination of information. Students were randomly matched. However, matching never took place within the same class, but we matched two subjects from two different classes. Due to the circumstance that class sizes were very similar, most participants had exactly one matching partner. If class sizes were unequal, we matched a student of the larger class with two students of the smaller class. Of course, participants with two matching partners only received payment for the interaction with one (randomly determined) partner. Subjects were also informed that the identities would be kept strictly confidential toward other participants. In order to guarantee also anonymity within a class, we used sliding walls between subjects so that they could not observe other subjects' decisions.

participants. We pre-tested asking for a probability distribution and encountered difficulties in understanding. Point beliefs, however, were very easy for subjects to understand. Second, eliciting point beliefs requires considerably less time than asking for a probability distribution of beliefs. Given that we already had a considerable number of decisions (54) and that the experiment took about 2.5 hours, we opted for the shorter elicitation of point beliefs. It is correct that eliciting a probability distribution has some advantages from a theoretical point of view, but it was impractical to implement that with our subject pool. Furthermore, Sutter et al. (2013) show that asking for a probability distribution can also lead to theoretical problems in studying a decision maker's consistency.

Each session was started with an extensive description and training of the game (see the instructions in the appendix). Questions were answered in private. Before the start of the experiment, each participant had to fill in a control questionnaire that checked the understanding of how decisions or beliefs mapped into payoffs. Five subjects (out of 196) did not answer all questions correctly. They are excluded in the following, yielding 191 subjects for the subsequent analysis.

The experiment itself was run with paper and pen. We handed out the three decision sheets for each game (for choices, first- and second-order beliefs) game by game. That means that subjects had to make all three decisions for a particular game before proceeding to the next game. After the experiment all participants were asked in a questionnaire to provide information about the number of siblings, their current grades in mathematics and German, and their ability to play chess (either yes or no). Due to the circumstance that 11 participants in grade 7 did not provide all corresponding information, parts of our analysis (in Table 6 and Table 7) are based on 180 subjects.

In total, each session lasted approximately 2.5 hours. The average earnings were about 7 Euros for subjects in 5th and 7th grade, and 14 Euros for those in 9th and 11th grade, and subjects were paid in private after the experiment. These earnings were slightly higher than the average weekly pocket money of around 5 Euros for 10- to 13-year-olds, and 13 Euros for 14- to 17-year-olds. The experimental exchange rate (from points to Euros) was contingent on subjects' age. From a survey before running the experiment we knew that subjects in grades 9 and 11 received on average slightly more than twice as much weekly pocket money as subjects in grades 5 and 7. Accordingly, if a subject was paid for her choice, we paid 20 Euro-Cents per experimental point in grades 9 and 11, and 10 Euro-Cents per point in grades 5 and 7. If subjects were paid for their first-order or second-order belief, 9th-graders and 11th-graders received 10 Euros for a correct belief, and 5th-and 7th-graders got 5 Euros. Incorrect beliefs yielded no payoffs across all grades. Note that all subjects received a show up fee, which was 4 Euros in grades 9 and 11, and 2 Euros in grades 5 and 7.

3. Experimental results[5]

3.1. Frequency of choosing and expecting Nash and the complexity of decisions

Table 2 reports in the upper panel the relative frequencies of particular choices. The middle and lower panel illustrate in the same manner first-order and second-order beliefs. The first column considers all 18 games, while columns two and three present separate data for ***D***-games and ***ND***-games. We denote decisions that imply the strategy that leads to the Nash-equilibrium by "Nash", decisions that imply the strategy that would yield an outcome that Pareto-dominates the Nash-equilibrium by "Pareto" and decisions implying other strategies by "Other".

Table 2 about here

We note that on average Nash is played in 45% of all cases. As becomes clear from Table 2 the relative frequency of playing Nash is, in the aggregate, roughly the same across all four grades in which we ran the experiment. It ranges from 42% in 11th grade to 48% in 5th grade, and a Kruskal-Wallis test rejects a significant different across age groups.[6]

Not surprisingly, the relative frequency of choosing the Nash strategy is clearly higher in ***D***-games (62%) than in ***ND***-games (24%). This is also true for each single age group. Looking at first- and second-order beliefs we observe that the relative frequency of the Nash strategy is about 10 percentage points smaller compared to actual play.[7]

Table 3 about here

[5] All analyses presented in the results section are based on pooled data of row and column players. This is justified since there are 16 isomorphic games in the set of 18 games and all decision tasks were presented in a way that players saw themselves as row players. Note that all results reported here would also hold if we concentrated only on the 16 isomorphic games.

[6] In order to check whether the insignificant result might be due to sample size, we conducted several robustness checks. To start with, we note that the sample size would need to be multiplied by a factor of 4.65 to get a significant Kruskal-Wallis-test statistic if we assumed that the distribution of choices within each age group would remain unchanged. This means that only with a sample size of more than 900 subjects the actually observed relative frequencies would be statistically significant, albeit economically small. Second, we calculated the power of our test and found that a power of 0.80 for a significance level of $\alpha = 0.05$ could be achieved with our current sample size of N = 191 already if the range of the relative frequency of choosing Nash is less than 9 percentage points (assuming a mean of 45%). This means that our sample size is large enough to capture potential treatment (i.e., age) effects.

[7] This pattern is consistent with a model of noisy introspection by Goeree and Holt (2004) in which they predict more noise – and hence less equilibrium play – with higher-order beliefs than with actual play.

Table 3 shows how the degree of a game's complexity affects choices and beliefs. A game's complexity is defined by the type of the underlying game (*"ND"* or *"D"*) and the number of rounds of iterated pure-strategy dominance required in order to identify the equilibrium choice.[8] The table reports the frequencies of chosen strategies (as in Table 2) according to five different categories: ***D***-games with either one or two rounds of iterated pure-strategy dominance – 1R(D) and 2R(D) – and ***ND***-games with either two, three or an infinite number of rounds – 2R(ND), 3R(ND), and ∞R(ND). The first column in the upper panel illustrates that in games where the decision maker herself has a dominant strategy, she plays Nash in 85% of cases. The likelihood of playing Nash is drastically reduced once a player has no longer a dominant strategy (see the four right most columns in Table 3). The middle and lower panel of Table 3 refer to first- and second-order beliefs. While Nash remains on average at or above 80% in the first column, the other columns show that Nash is considerably less often expected (both in first- and second-order beliefs) than actually played in games without a dominant strategy for the decision maker. Table 3 shows only little evidence for any age effects, as the average choice frequencies are typically very similar across age groups and only in a few cells we note significant age differences (according to a Kruskal-Wallis test). This observation suggests that strategic decision making is fairly well developed at an age of 10 years and hardly changes in subsequent years of adolescence. We summarize the findings in this subsection as follows:

Result 1: Overall, the equilibrium strategy is chosen in about 45% of cases. First- and second-order beliefs of equilibrium play are significantly less frequent by roughly 10 percentage points. The relative frequency of playing Nash decreases with a game's complexity, i.e., the number of rounds of iterated pure-strategy dominance needed to identify the equilibrium. The share of equilibrium play and expected equilibrium play (Nash first-order and second-order beliefs) changes hardly between the age of 10 and 17 years.

3.2. Frequency of consistent choices and beliefs

We define a player's "*own consistency*" as the relative frequency of choices that are a best response to her own first-order belief in a particular game, and the "*opponent's expected*

[8] Very similar results are obtained if we consider the number of available strategies as an indicator for a game's complexity. In Table B.1 of the Appendix we present relative choice frequencies of strategies when taking the number of strategies in a game as the classification for complexity.

consistency" as the relative frequency with which a player's first-order beliefs are a best response to her second-order beliefs. As can be seen from Table 4, the relative frequency of "own consistency" ranges from 61% to 64% across all age groups, which is another indication that there is hardly any change in strategic decision making from the age of 10 to 17 years. The "own consistency" is higher in **D**-games than in **ND**-games across all age groups, stressing the importance of dominant strategies for consistent behavior. The "opponent's expected consistency" ranges from 51% to 56%. The latter frequency is significantly lower than the frequency of "own consistency" ($p < 0.05$, two-sided Wilcoxon-signed rank tests for all pairwise comparisons).

Inconsistent choices have an expected cost for a particular subject. To see this, consider for example that the row-player in game #3 at the top-left corner of Figure 1 plays strategy [1] and expects the column player to play strategy [2]. If the expectation was correct, the row-player's payoff would be 31, although he could have obtained 55 if he played his strategy [2] as a best-response to his expectation. In this way, one can calculate the expected costs of inconsistent choices (based on the assumption that first-order expectations would be correct). They are 10.12 units on average, and 10.11, 10.65, 9.47, 10.29 in the four different grades (in ascending order of grades). The numbers for the different grades are not significantly different from each other (Kolmogorov-Smirnov two-sample tests for equality of distribution functions, $p > 0.293$[9]).

The same approach can be applied to the opponent's expected consistency, yielding average expected costs of 12.99 units, and 13.49, 13.50, 12.15, 12.86 in the four different grades (in ascending order of grades). The numbers for the different grades are, again, not significantly different from each other (Kolmogorov-Smirnov two-sample tests for equality of distribution functions, $p > 0.162$[10]).

Table 4 and Table 5 about here

Of course, standard game theory does not only assume that subject best respond to their beliefs, but more precisely it predicts a combination of Nash equilibrium choices and a belief that the opponent also plays Nash. This game theoretically rational behavior is denoted Nash-

[9] The exact *p*-values for pairwise comparisons are as follows: 0.908 (5th grade vs. 7th grade), 0.602 (5th grade vs. 9th grade), 0.779 (5th grade vs. 11th grade), 0.293 (7th grade vs. 9th grade), 0.724 (7th grade vs. 11th grade), 0.440 (9th grade vs. 11th grade)

[10] The exact *p*-values for pairwise comparisons are as follows: 0.769 (5th grade vs. 7th grade), 0.162 (5th grade vs. 9th grade), 0.806 (5th grade vs. 11th grade), 0.196 (7th grade vs. 9th grade), 0.946 (7th grade vs. 11th grade), 0.408 (9th grade vs. 11th grade)

consistency (Nash-CON) in the following and analyzed in Table 5. There we also consider another type of consistency which appeared very frequently in our data. This rather naive type of consistency is called "Maximum-consistency" (Max-CON) and is defined as a combination of a player's choice and first-order belief that results in the maximum available payoff for the player (in the case of "own consistency") or in the maximum payoff for the opponent player (in the case of "opponent's expected consistency"). An important feature of the games we employ is that Nash-CON never overlaps with Max-CON since all games have the structure of a prisoner's dilemma, i.e., the Nash-equilibrium is Pareto-dominated.

Table 5 shows that Nash-CON is observed for own choices in roughly 16% of all cases. It is much more frequent in ***D***-games than in the ***ND***-games ($p < 0.05$, two-sided Wilcoxon signed rank tests). Moreover, the share of Nash-CON is higher in the category "own consistency" than in the category "opponent's expected consistency". At the same time, Max-CON is observed in roughly 40% of all cases, indicating that subjects are pretty optimistic (by expecting the highest possible payoffs from their combination of choices and first-order beliefs). We summarize the findings in this subsection as follows:

Result 2: Choices are a best reply to first-order beliefs in more than 60% of cases, while first-order beliefs are a best reply to second-order beliefs in a significantly smaller number of cases (less than 56%). Most frequently, consistent choices and beliefs are of the Max-CON type, meaning that subjects expect the maximum payoffs to occur. Nash-consistency (playing the Nash equilibrium strategy and expecting the opponent to play Nash) is less frequent with 16% on average (all age groups aggregated), and more likely in ***D****-games than in* ***ND****-games. We observe no significant differences across age groups.*

3.3. Determinants of consistent decisions

In this section we examine in more detail what are important determinants of consistent decision making in normal form games, i.e., of behavior where choices are best responses to first-order beliefs or where first-order beliefs are best responses to second-order beliefs. Of course, if a player has a dominant strategy – as one player always has in our D-games – then the optimal choice is independent of any belief, and hence we exclude all cases in which a player has a dominant strategy. Note, however, that we keep those cases for the analysis in which one player faces an opponent with a dominant strategy, because in this case expectations obviously matter.

Table 6 about here

Table 6 reports the results of a probit estimation[11] of a player's *"own consistency"* on the independent variables "ability to play chess", "math grade", "German grade"[12], "existence of siblings" (as a dummy variable) and five other factors which are explained in the following: The dummy variable "opponent dominant strategy" captures the cases where the opponent player has a dominant strategy. The next three independent variables ("7th grade", "9th grade", "11th grade") are dummies for the different grades that were included in our experiment (with 5th graders as the benchmark) and the variable "# of possible outcomes" is calculated by multiplying the number of strategies for player A by the number of strategies for player B to capture potentially more demanding games for cognitive workload (which might be negatively related to consistency of choices and beliefs).

The estimation shows that a player's "own consistency" is mainly influenced by the existence of a dominant strategy for the opponent and (marginally) by the person's performance in mathematics assessed by her corresponding grade. A dominant strategy for the opponent player leads to a significantly lower likelihood of "own consistency". Due to the circumstance that a dominant strategy reduces the complexity of a game to its lowest possible degree the negativity of the corresponding correlation is surprising at first sight. A further analysis of this result indicates a lack of basic strategic thinking which can be explained as follows: Experimental participants were expecting the opponent to play his dominant strategy in more than 86% of all cases where the opponent had a dominant strategy (see Table 3). Obviously, many participants were able to detect an opponent's dominant strategy when being asked for their first-order belief, but unable to integrate this insight into their own decision making process by choosing a best reply. Additionally our analysis in Table 6 reveals a significant influence of the math grade, i.e., children and adolescents with better grades have a higher probability to be consistent. Note that lower grades indicate better performance in Austrian schools. All other independent variables in Table 6 have no significant impact on "own consistency".

Table 7 about here

---

[11] In our probit estimations (Tables 6 and 7) standard errors are clustered on the decision maker, since each decision maker had to make decisions in all 18 games.

[12] Both the math grade and the German are coded such that lower grades indicate better performance.

In Table 7 we present a probit regression in which we use the same independent variables as in the previous regression (Table 6), but take the "*opponent's expected consistency*" as the dependent variable. We again find that the "opponent's expected consistency" is significantly influenced by the existence of a dominant strategy for the player and by the player's grade in mathematics. Our results show that a dominant strategy for the player himself reduces the likelihood of consistent expectations whereas a relatively good grade in math increases it. Additionally we observe some significant effects related to our age groups in which the probability of the "opponent's expected consistency" is significantly higher for 9th graders (compared to 5th graders). Finally, our analysis in Table 7 shows a significant and positive influence of the German grade, which means that children and adolescents with better grades state consistent expectations for their opponent with a lower probability.

Result 3: The likelihood of consistent decision making significantly increases with better grades in mathematics. Furthermore the probability of a player's "own consistency" ("opponent's expected consistency") is significantly lower when his opponent has a dominant strategy (the player has a dominant strategy). We find some influence of age and German grades on "opponent's expected consistency".

3.4. Non-randomness of decisions

Before turning to the estimation of strategic and non-strategic types, we address whether the choices analyzed so far might have been simply the result of random play of subjects. We start with the relative frequency of subjects playing the Nash-strategy. It is important to stress that random play would lead to a relative frequency of Nash-play of 43%, which is, at first sight, very close to the actually observed 45% (see Table 2). This is no indication of random play of our subjects, however. Indeed, we can compute the distribution of relative frequencies of playing Nash under the assumption of random play, thus calculating the relative frequencies of observing 0, 1, 2, …, 17, 18 Nash-choices in case of random play.[13] Figure 2 shows the results for this calculation (the bell-shaped thick curve) and compares it to the

[13] The theoretical distribution is computed as follow. We assume that in each game, each possible choice is taken with equal probability so that each possible combination of the 18 choices can happen with equal probability. Then the probability to observe an individual choosing x Nash choices (with x ranging from zero to 18) is given by the ratio between the number of all combinations with x choices with the total number of combinations. Given that individuals take the decisions independently, this distribution corresponds to the relative frequencies to take x Nash-choices under random behavior.

actually observed relative frequencies of playing Nash (on the individual level), separate for the four different age groups. We note that a Kolmogorov-Smirnov one sample test reveals that the actually observed distribution is significantly different from the theoretical distribution for the case of pure random play ($p < 0.01$), thus rejecting the notion that play is random. The same holds true if we look (in the aggregate) at choices, first- and second-order beliefs in Figure 3. Again, the theoretically predicted distribution of relative frequencies is markedly different from the actually observed distributions.

Figure 2 to 5 about here

In Figure 4 and 5 we repeat the same exercise to show that also the likelihood of consistent choices (where choices are best responses to first-order beliefs; see Figure 4; or where we also consider the opponent's consistency; see "fob" in Figure 5) cannot be explained by random behavior of our participants. The distribution of expected relative frequencies of consistent choices (from zero to 18) is significantly different from the distribution that has been actually observed in single age groups (Figure 4) or in the aggregate (Figure 5).

Table 8 about here

Table 8 shows the *p*-values of Kolmogorov-Smirnov one-sample tests about the differences between the two distributions (the theoretical one vs either the observed of Nash choices or Best reply or Nash consistent choices). All tests are conducted for the aggregate of observations and for each age group separately. The results show that the null-hypothesis of random choice of our participants can be rejected. Further evidence that behavior is not random will come from the estimation of the error rate model that is presented in the next subsection.

Result 4: Actual behavior does not conform to the pattern that would result if subjects played randomly.

3.5. Estimation of strategic and non-strategic types

In the following we present a maximum likelihood error-rate analysis of players' choices following the framework of Costa-Gomes et al. (2001). It is a mixture model in which each player's type is drawn from a common prior distribution over eight types and where the type is assumed to be constant for all 18 games. The eight different types can be classified into non-strategic and strategic types and are defined as follows:[14]

Non-strategic types: (1) An *Altruistic* type makes an attempt to maximize the sum of both players' payoffs, implicitly assuming that the opponent is also *Altruistic* (see Costa-Gomes et al., 2001). Note that *Efficiency-loving* would probably be a more appropriate term for such a type. Hence, we will call this type *Altruistic/Efficiency-loving*. (2) A *Pessimistic* type plays maximin, thus is taking choices that secure him the best of all worst outcomes. (3) An *Optimistic* type chooses the strategy that maximizes the best possible own payoff, thus ignoring the incentives of the opponent player. As noted by Costa-Gomes et al. (2001), it is impossible to distinguish an *Optimistic* type from a *Naïve* type in the 18 games used here. A *Naïve* type assigns equal probabilities to the opponent's strategies and best responds to this naïve belief. While a *Naïve* type might reflect strategic decision making with diffuse beliefs, Costa-Gomes et al. (2001) describe *Naive* types as non-strategic. We follow their approach, but talk about *Optimistic* types, which are non-strategic for sure.

Strategic types: (4) Type *L2* is choosing a best response to *Optimistic* types. (5) Type *D1* plays best reply to a uniform prior over the opponent's remaining strategies after applying one step of deleting strategies that are dominated by pure strategies. (6) Type *D2* goes one step further in deleting dominated strategies. After applying two steps of deleting dominated strategies he chooses a best reply to the opponent's remaining strategies. (7) An *Equilibrium* type takes equilibrium choices (which are unique in our games). (8) Choices of a *Sophisticated* type are based on the actually observed distribution of strategies in the experiment's subject pool (i.e., his age group). A player of this type takes this distribution as a probability distribution for his opponent's choice and plays best reply to this information.

Figure 1 includes in italic font information on which decisions should be taken (or expected in the case of the belief determination) by row and column players if they act according to a certain theoretical type (as explained in the previous two paragraphs).

For the estimation of the mixture model let $i = 1, \ldots, N$ index the different players, let $k = 1, \ldots, K$ index our types, and let $c = 2$, 3, or 4 be the number of a player's possible

[14] We follow Costa-Gomes et al. (2001) in the selection of types to be considered. As they indicate, the definition of types is largely based on earlier work by Stahl and Wilson (1994, 1995).

decisions in a given game. We assume that player *i* of type-*k* normally makes type *k*'s decision, but in each game he makes an error with probability $\varepsilon_{ik} \in [0,1]$, type *k*'s *error rate*, in which case he takes each of his *c* decisions with probability 1/*c*. The probability of type *k*'s decision is then $1 - \frac{c-1}{c}\varepsilon_{ik}$. So the probability of any single non-type *k* decision is $\frac{\varepsilon_{ik}}{c}$. We assume errors are independently and identically distributed across games and are independently distributed across players. Moreover we assume that individual error rates depend on a set $\{y_j\}$ of observable exogenous variables, following the specification:

$$\varepsilon_{ik} = \frac{e^{\beta_{k0}+\sum_{j=1}^{m}\beta_{kj}\cdot y_{ji}}}{1+e^{\beta_{k0}+\sum_{j=1}^{m}\beta_{kj}\cdot y_{ji}}}\ for\ k \in \{1,\dots,8\}$$

where y_{ji} denotes the observed value of variable y_j for player *i* and β_{kj} are exogenous parameters that take into account the effect of variable y_j on the error rate ε_{ik}.[15]

Let p_{ik} denote player *i*'s common prior *k*-type probability where $\sum_{k=1}^{K} p_{ik} = 1$. We assume that player *i*'s common prior *k*-type probabilities depend on a set $\{y_j\}$ of observable exogenous variables, following the specification:

$$p_{ik} = \frac{e^{\alpha_{k,0}+\sum_{j=1}^{m}\alpha_{kj}\cdot y_{ji}}}{1+\sum_{h=1}^{7} e^{\alpha_{h,0}+\sum_{j=1}^{m}\alpha_{h,j}\cdot y_{ji}}}\ for\ k \in \{1,\dots,7\}\ \text{and}\ p_8 = \frac{1}{1+\sum_{h=1}^{7} e^{\alpha_{h,0}+\sum_{j=1}^{m}\alpha_{h,j}\cdot y_{ji}}}$$

where α_{kj} are exogenous parameters that take into account the effect of variable y_j on the probability p_{ik}.

The likelihood function can be constructed as follows. Let T^c denote the total number of games in which players have *c* decisions. In our design we have T^2 =11, T^3 =6, and T^4 =1. Then let x_k^{ic} denote the number of player *i*'s decisions that equal type *k*'s in games in which he has *c* decisions, with $x_k^i = (x_k^{i2}, x_k^{i3}, x_k^{i4})$, $x^i = (x_1^i, \dots, x_K^i)$, and $x = (x^1, \dots, x^n)$. Let $p_i = (p_{i1}, \dots, p_{iK})$, $p = (p_1, \dots, p_N)$, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iK})$, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$. Given that a game has one type-*k* decision and *c*-1 non-type-*k* decisions, the probability of observing a particular sample with x_k^i type-*k* decisions when player *i* is type *k* can be written as:

$$\mathrm{L}_k^i(\varepsilon_{ik}|x_k^i) = \prod_{c=2,3,4}\left[1-\frac{1-c}{c}\varepsilon_{ik}\right]^{x_k^{ic}}\left[\frac{\varepsilon_{ik}}{c}\right]^{T^c-x_k^{ic}} \quad (1)$$

Weighting the right-hand side by p_{ik}, summing over *k*, taking logarithms, and summing over *i* yields the log-likelihood function for the entire sample:

[15] Checking the influence of observable exogenous variables on the error rates is motivated by the fact that looking at errors can be informative for understanding human behavior (see Jacobson and Petrie, 2009, for instance, who show that in a risky choice task, those making mistakes – by being inconsistent – are also more likely to make suboptimal decisions) and that previous research has shown that error rates in these kinds of mixed-type models are often not uniformly distributed (see, von Gaudecker et al., 2011, for example).

$$\ln L(p,\varepsilon|x) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} p_{ik} L_k^i\left(\varepsilon_{ik}|x_k^i\right) \quad (2)$$

We estimate this model using the following exogenous variables: age, gender, an interaction variable between gender and age and the grades in German and math. By this specification the model has 90 parameters[16] that we can use to estimate the marginal effects of age, gender and grades in German and math on the prior type-*k* probabilities and on the error rates. Results are summarized in Tables 9 and 10 that report, separated by gender, the estimated error rates and prior type-*k* probabilities at the average of the exogenous variables (Table 9) and the marginal effects of the exogenous variables on prior type-*k* probabilities and the error rates (Table 10).[17]

Table 9 and Table 10 about here

Table 9 reveals that the modal type for both groups is the Optimistic with a share of 41% for males and 54% for females and that 41% of males and 38% of females are of one of the strategic types, but these differences across gender are not significant.

We note a different type distribution between males and females, though. Three types are significantly represented in both groups and they are Pessimistic, Optimistic and D1. The types Altruistic/Efficiency-loving, Sophisticated and L2 are significant only in the male sample, while Equilibrium is significant only in the female group. Furthermore males are significantly more probable to be Altruistic/Efficiency-loving or L2 and less probable to be D1 than females. Finally even if there are no significant differences between males and females in the estimated probability to be one of strategic types, we note that the most frequent strategic type is L2 for males, but D1 for females.

The estimated error rates are almost all significantly different from one (that corresponds to random choice) with the exception of type D2 (that is not significant in the whole sample,

[16] We estimated a more parsimonius model where the error rate is the same across types (48 parameters). Both Akaike information criterion and Bayesian information criterion indicate that the model with more parameters is preferred.

[17] Note that subjects have to choose one of the 5,971,968 (= $2^{11}\cdot3^{6}\cdot4$) possible combinations of choices in the 18 games. Our eight behavioral types imply eight out of almost 6 million combinations. It is noteworthy that 38 out of 191 subjects have all 18 decisions consistent with one of these eight behavioral types. This is a remarkable fraction of 20% of subjects. Yet, given that 80% of subjects do not conform in all 18 decisions to one specific type, this justifies the use of our mixture model that allows for errors and different types, and in this subsection we estimate the likelihood of different behavioral types. Another approach would be to estimate each behavioral model separately and then test which has the best fit with our data. Such an approach would assume that we look for one specific type that best describes our full sample (in fact, this would be the Naïve type). Yet, we do not report the results of such an exercise because – like the literature following Costa-Gomes et al. (2001) – we believe that the mixture model presented here is the more appropriate approach to capture the heterogeneity of behavioral types.

however) and the Equilibrium type for females (see symbol + in Table 9). There is some difference between males and females in the error rates of Optimistic and Equilibrium types, with males characterized by lower error rates.

Table 10 shows that age has some weak impact on the probability to be a certain type that is gender specific: the probability to be D1 is increasing with the age in the female group while that to be Optimistic is decreasing with age in the male group. Finally, we also note that the probability to be one of the strategic types has a positive coefficient for age, but it is not significant. Concerning the influence of grades we find that females with better math grades are more likely to be classified as strategic.

Age does have an impact on some of the error rates. For non-strategic types, for males this effect is significantly positive for Altruistic/Efficiency-loving and Pessimistic types, while for females it is negative for the Pessimistic type only. For strategic types, age has a negative impact on the error rate of L2, but this effect is significant only for males.*Result 5: The likelihood of subjects to be of any of five different strategic types (Equilibrium, Sophisticated, D1, D2, L2) is about 40%. The modal type, however, is a non-strategic Optimistic type. Males are significantly more often an Altruistic/Efficiency-loving or L2 type and less often a D1 type than females. Age has a positive effect on the probability of a female to be a D1 type and a negative effect on the probability of a male to be an Optimistic type. Error rates depend on age for some types, but there is no overall compelling pattern in the error rates.*

4. Conclusion

In this paper we have studied strategic sophistication of 191 adolescents, aged 10 to 17 years, in 18 different normal-form games taken from Costa-Gomes et al. (2001). Overall, we have found that about 40% of subjects can be classified as strategic, thus taking into account their opponent's incentives and strategies when making decisions.

Interestingly, we have found at best a weak influence of age on the likelihood to be a strategic type. The estimated coefficient, while positive, is insignificant. In fact, many of the stylized facts found in our experiment with teenagers are very similar to what is known from experiments with university students in their early 20ies. It is particularly noteworthy that in Sutter et al. (2013) it has been estimated for adults that 40% of subjects are classified as strategic types, a fraction that is obviously not significantly different from the share of children and teenagers classified as strategic in the experiment presented here. Moreover, the relative frequency of playing Nash in the 18 different games is approximately 45% in our subject pool, which is well in the range of 40% to 50% of Nash-play reported in other studies (e.g., Costa-Gomes et al., 2001; Costa-Gomes and Weizsäcker, 2008; Danz et al., 2012; Sutter et al., 2013). The share of consistent decisions – as best response to one's own beliefs – is around 60% in our sample, while it is around 55% both in Costa-Gomes and Weizsäcker (2008) and Sutter et al. (2013) who have run their experiments with university students. Hence, teenagers and young adults in their early 20ies play normal form games in a very similar manner. This finding is also consistent with evidence from Brosig-Koch et al. (2015) about the development of the ability to use backward induction in a race game. While they have found an effect of age below the age of 10, in the teenage years they have not observed any further development with age, which is in line with what we observe about strategic sophistication. We do find a slight effect of age on the error rates with which experimental participants are classified into different (strategic and non-strategic) types. However, the estimations about the error rates do not provide a compelling pattern. For instance, for some types we find an increase in error rates with age, for others a decrease.

While our paper contributes to the literature on strategic sophistication by showing that adolescents are able to play complex games in a way that is very similar to the behavior of adults (university students) and that age has no marked impact on strategic behavior, another contribution to this literature refers to the analysis of consistent decision making. Our corresponding analyses include both game characteristics (such as the existence of dominant strategies or the number of available strategies) and players' individual characteristics (such as grades or the existence of siblings). In this context, our results (in section 3.3) have revealed that students with better grades in math have a significantly higher probability to state consistent choices and beliefs.

Given our results, one straightforward extension of our study would be to study also subjects who are younger than 10 years of age in order to determine whether and when noticeable changes in the ability to act and think strategically develop. Our experimental design is most likely not the optimal choice for younger children below the age of 10 years, because the games used in our study and the decisions on first- and second-order beliefs might easily overburden younger children. While we have chosen our design due to its suitability for studying different types of strategic and non-strategic behavior and because a lot of evidence from college or university students in their early 20ies is available for comparison, we regard it as an interesting avenue for future research to design simpler games that could provide insights into how strategic sophistication, and the interaction of choices and beliefs, develops in the first ten years of human life. Such an endeavor will shed further light on the origins of game-theoretic thinking in humans.

References

Almas, I., Cappelen, A., Sorensen, E., Tungodden, B. (2010), Fairness and the development of inequality acceptance. *Science* 328(28 May): 1176-1178.

Bhatt, M., Camerer, C. F. (2005), Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior* 52: 424-459.

Brosig-Koch, J., Heinrich, T., Helbach, C. (2015), Exploring the capability to backward induct: An experimental study with children, adolescents, and young adults. *European Economic Review* 74: 286-302.

Camerer, C. F. (2003), *Behavioral Game Theory*. *Experiments in Strategic Interaction*. Princeton, Princeton University Press.

Costa-Gomes, M., Crawford, V., Broseta, B. (2001), Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69: 1193-1235.

Costa-Gomes, M., Weizsäcker, G. (2008), Stated beliefs and play in normal-form games. *The Review of Economic Studies* 75: 729-762.

Crawford, V., Costa-Gomes, M., Iriberri, N. (2013), Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51: 5-62.

Crawford, V., Iriberri, N. (2007), Fatal attraction: Salience, naiveté, and sophistication in experimental "Hide-and-Seek" games. *American Economic Review* 97: 1731-1750.

Danz, D., Fehr, D., Kübler, D. (2012), Information and beliefs in a repeated normal-form game. *Experimental Economics* 15: 622-640.

Goeree, J., Holt, C. (2004), A model of noisy introspection. *Games and Economic Behavior* 46: 365-382.

Haruvy, E., Stahl, D., Wilson, P. (1999), Evidence for optimistic and pessimistic behavior in normal-form games. *Economic Letters* 63: 255-259.

Jacobson, S., Petrie, R. (2009), Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of Risk and Uncertainty* 38: 143-158.

Kail, R. V., Cavanaugh, J. C. (2010.), *Human Development: A Life-Span View*, 5th Edition. Belmont, CA: Wadsworth Publishing.

Murnighan, J.K., Saxon, M.S. (1998), Ultimatum bargaining by children and adults. *Journal of Economic Psychology* 19: 415-445.

Rey-Biel, P. (2009), Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior* 65: 572-585.

Stahl, D., Wilson, P. (1994), Experimental evidence of players' models of other players. *Journal of Economic Behavior and Organziation* 25: 309-327.

Stahl, D., Wilson, P. (1995), On players' models of other players: theory and experimental evidence. *Games and Economic Behavior* 10: 213-254.

Sutter, M., Czermak, S., Feri, F. (2013), Strategic sophistication of individuals and teams. Experimental evidence. *European Economic Review* 64: 395-410.

Sutter, M., Kocher, M. (2007), Trust and trustworthiness across different age groups. *Games and Economic Behavior* 59: 364-382.

von Gaudecker, H.-M., van Soest, A., Wengström, E. (2011), Heterogeneity in risky choice behavior in a broad population. *American Economic Review* 101: 664-694.

Weizsäcker, G. (2003), Ignoring the rationality of others: evidence from experimental normal-form games. *Games and Economic Behavior* 44: 145-171.

Weller, J. A., Levin, I. P., Rose, J. P., Bossard, E. (2012), Assessment of decision-making competence in preadolescence. *Journal of Behavioral Decision Making* 25: 414-426.

Tables and Figures

Table 1: Absolute numbers of subjects across age groups

		Players' Types		*Gender*		
	Total	***Row***	***Column***	***Male***	***Female***	***not reported***
5th grade	45	23	22	24	21	0
7th grade	52	28	24	20	21	11
9th grade	52	28	24	22	30	0
11th grade	42	23	19	25	17	0
Sum	191	102	89	91	89	11

This table excludes 5 subjects who had difficulties in understanding the game.

Table 2: Choices and beliefs of adolescents (relative frequencies in %)

	All games	***D-games***	***ND-games***
		CHOICES	
Nash **5th grade**	47.90	64.67	26.94
Nash **7th grade**	45.62	62.12	25.00
Nash **9th grade**	45.83	63.08	24.28
Nash **11th grade**	41.53	59.05	19.64
Nash overall	45.31	62.31	24.08
Pareto **5th grade**	43.58	31.56	58.61
Pareto **7th grade**	46.80	35.38	61.06
Pareto **9th grade**	46.05	33.85	61.30
Pareto **11th grade**	50.40	38.33	65.48
Pareto overall	46.63	34.71	61.52
Other **5th grade**	8.52	3.78	14.44
Other **7th grade**	7.59	2.50	13.94
Other **9th grade**	8.12	3.08	14.42
Other **11th grade**	8.07	2.62	14.88
Other overall	8.06	2.99	14.40
		FIRST-ORDER BELIEFS	
Nash **5th grade**	34.81	51.33	14.17
Nash **7th grade**	33.01	49.23	12.74
Nash **9th grade**	34.83	51.92	13.46
Nash **11th grade**	34.26	51.67	12.50
Nash overall	34.20	50.99	13.22
Pareto **5th grade**	57.16	46.44	70.56
Pareto **7th grade**	59.94	48.46	74.28
Pareto **9th grade**	57.05	45.19	71.88
Pareto **11th grade**	57.14	44.76	72.62
Pareto overall	57.88	46.28	72.39
Other **5th grade**	8.03	2.22	15.28
Other **7th grade**	7.05	2.31	12.98
Other **9th grade**	8.12	2.88	14.66
Other **11th grade**	8.60	3.57	14.88
Other overall	7.91	2.72	14.40
		SECOND-ORDER BELIEFS	
Nash **5th grade**	37.90	51.33	21.11*
Nash **7th grade**	32.59	48.27	12.98
Nash **9th grade**	34.19	49.62	14.90
Nash **11th grade**	34.26	50.48	13.99
Nash overall	34.64	49.84	15.64
Pareto **5th grade**	55.31	46.00	66.94
Pareto **7th grade**	59.30	49.04	72.12
Pareto **9th grade**	56.52	47.50	67.79
Pareto **11th grade**	57.01	46.67	69.94
Pareto overall	57.10	47.38	69.24
Other **5th grade**	6.79	2.67	11.94*
Other **7th grade**	8.12	2.69	14.90
Other **9th grade**	9.30	2.88	17.31
Other **11th grade**	8.73	2.86	16.07
Other overall	8.26	2.77	15.12

** (*) significant difference at $p < 0.05$ ($p < 0.1$) across all age groups for a particular set of games and a given strategy according to a two-sided Kruskal-Wallis test.

Table 3: Complexity of the game, choices and beliefs (relative frequencies in %)

Complexity (game type)[a]	*1R (D)*	*2R (D)*	*2R (ND)*	*3R (ND)*	*∞R (ND)*
			CHOICES		
Nash **5th grade**	89.78	39.56	38.89	8.88	30.00
Nash **7th grade**	86.15	38.08	35.58	7.69	28.37
Nash **9th grade**	83.46	42.69	36.54	7.69	26.44
Nash **11th grade**	81.90	36.19	28.57	8.33	20.83
Nash overall	85.34	39.27	35.08	8.11	26.57
Pareto **5th grade**	6.22	56.89	61.11	42.22**	65.56
Pareto **7th grade**	11.15	59.62	64.42	45.19	67.31
Pareto **9th grade**	15.00	52.69	63.46	55.77	62.98
Pareto **11th grade**	16.19	60.48	71.43	47.62	71.43
Pareto overall	12.14	57.28	64.92	47.91	66.62
Other **5th grade**	4.00	3.56	0.00	48.89**	4.44
Other **7th grade**	2.69	2.31	0.00	47.12	4.33
Other **9th grade**	1.54	4.62	0.00	36.54	10.58
Other **11th grade**	1.90	3.33	0.00	44.05	7.74
Other overall	2.51	3.46	0.00	43.98	6.81
			FIRST-ORDER BELIEFS		
Nash **5th grade**	87.11	15.56	14.44	13.33	14.44
Nash **7th grade**	87.69	10.77	15.38	10.58	12.50
Nash **9th grade**	86.54	17.31	15.38	7.69	15.38
Nash **11th grade**	86.67	16.67	11.90	5.95	16.07
Nash overall	87.02	14.98	14.39	9.42	14.53
Pareto **5th grade**	12.00	80.89	85.56	42.22*	77.22
Pareto **7th grade**	10.00	86.92	84.62	46.15	83.17
Pareto **9th grade**	11.54	78.85	84.62	55.77	73.56
Pareto **11th grade**	9.52	80.00	88.10	47.62	77.38
Pareto overall	10.78	81.78	85.61	48.17	77.88
Other **5th grade**	0.89	3.56	0.00	44.44	8.33*
Other **7th grade**	2.31	2.31	0.00	43.27	4.33
Other **9th grade**	1.92	3.85	0.00	36.54	11.06
Other **11th grade**	3.81	3.33	0.00	46.43	6.55
Other overall	2.20	3.25	0.00	42.41	7.59
			SECOND-ORDER BELIEFS		
Nash **5th grade**	77.78	24.89*	20.00	22.22**	21.11
Nash **7th grade**	82.69	13.85	13.46	8.65	14.91
Nash **9th grade**	77.69	21.54	11.54	8.65	19.71
Nash **11th grade**	85.71	15.24	9.52	8.33	19.05
Nash overall	80.84	18.85	13.61	11.78	18.59
Pareto **5th grade**	18.67	73.33	80.00	44.44	71.67
Pareto **7th grade**	15.00	83.08	86.54	46.15	77.88
Pareto **9th grade**	20.38	74.62	88.46	50.00	66.35
Pareto **11th grade**	11.43	81.90	90.48	45.24	72.02
Pareto overall	16.54	78.22	86.39	46.60	71.99
Other **5th grade**	3.56	1.78	0.00	33.33*	7.22
Other **7th grade**	2.31	3.08	0.00	45.19	7.21
Other **9th grade**	1.92	3.85	0.00	41.35	13.94
Other **11th grade**	2.86	2.86	0.00	46.43	8.93
Other overall	2.62	2.93	0.00	41.62	9.42

** (*) significant difference at $p < 0.05$ ($p < 0.1$) across all age groups for a particular set of games and a given strategy according to a two-sided Kruskal-Wallis test.

[a] The columns separate behavior according to (i) the different number of rounds (***R***) of iterated pure-strategy dominance a player needs to identify the own equilibrium choice, and (ii) the presence (***D***) or absence (***ND***) of a dominant strategy in the game.

Table 4: Consistency of decisions (relative frequency of best reply)

		All games	*D-games*	*ND-games*
Own consistency	**5th grade**	64.44	67.78	60.28
Choice is best reply to first-	**7th grade**	61.47	64.32	57.93
order belief	**9th grade**	64.42	68.27	59.62
	11th grade	60.85	64.05	56.85
	Overall	62.84	66.15	58.71
Opponent's expected	**5th grade**	50.99	55.78	45.00
consistency	**7th grade**	51.99	53.42	50.24
First-order belief is best	**9th grade**	55.98	60.00	50.96
reply to second-order belief	**11th grade**	51.59	55.24	47.02
	Overall	52.75	56.17	48.49

** (*) significant difference at $p < 0.05$ ($p < 0.1$) across all age groups for a particular set of games and a given strategy according to a two-sided Kruskal-Wallis test.

Table 5: Relative frequency of consistency-types Nash-CON and Max-CON

		All games	*D-games*	*ND-games*
	Nash-CON **5th grade**	15.56	23.78	5.28
	Nash-CON **7th grade**	14.55	21.60	5.77
	Nash-CON **9th grade**	18.38	27.69	6.73
	Nash-CON **11th grade**	15.48	24.76	3.87
Player's own consistency	Nash-CON **overall**	16.04	24.47	5.50
	Max-CON **5th grade**	40.99	42.00	39.72
	Max-CON **7th grade**	39.77	41.37	37.74
	Max-CON **9th grade**	38.03	37.69	38.46
	Max-CON **11th grade**	37.70	36.43	39.29
	Max.CON **overall**	39.13	39.43	38.74
	Nash-CON **5th grade**	10.13**	14.67**	4.44
	Nash-CON **7th grade**	6.22	9.32	2.40
	Nash-CON **9th grade**	11.22	16.92	4.09
	Nash-CON **11th grade**	8.99	12.86	4.17
Expected consistency of opponent	Nash-CON **overall**	9.11	13.43	3.73
	Max-CON **5th grade**	37.90	39.56	35.83
	Max-CON **7th grade**	42.34	42.54	42.07
	Max-CON **9th grade**	40.92	40.58	41.35
	Max-CON **11th grade**	39.29	40.71	37.50
	Max.CON **overall**	40.24	40.90	39.40

** (*) significant difference at $p < 0.05$ ($p < 0.1$) across all age groups for a particular set of games and a given strategy according to a two-sided Kruskal-Wallis test.

Table 6: Determinants of a player's "own consistency"

Variable	***Coefficient***	***Std. error***	***p-value***
opponent dominant strategy	-0.296	0.056	0.000
dummy for **7th grade**	0.095	0.168	0.571
dummy for **9th grade**	0.204	0.165	0.214
dummy for **11th grade**	0.019	0.158	0.907
# of possible outcomes	0.024	0.022	0.266
ability to play chess	-0.101	0.110	0.356
math grade	-0.108	0.061	0.078
German grade	0.053	0.063	0.401
existence of siblings	0.007	0.048	0.880
Constant	0.214	0.195	0.273

Probit model with a player's "own consistency" as dependent variable.
N = 2.340; standard errors clustered for the 180 decision makers (45 in 5th grade, 41 in 7th grade, 52 in 9th grade, 42 in 11th grade).

Marginal effects of independent variables

	Marginal effect	***Std. error***	***p-value***
opponent dominant strategy	-0.117	0.022	0.000
dummy for **7th grade**	0.038	0.066	0.571
dummy for **9th grade**	0.080	0.064	0.214
dummy for **11th grade**	0.007	0.062	0.907
# of possible outcomes	0.010	0.009	0.266
ability to play chess	-0.040	0.043	0.356
math grade	-0.043	0.024	0.078
German grade	0.021	0.025	0.401
existence of siblings	0.003	0.019	0.880

N =2.340; standard errors clustered for the 180 decision makers (45 in 5th grade, 41 in 7th grade, 52 in 9th grade, 42 in 11th grade).
Note that in this estimation we have excluded cases where a player has a dominant strategy.

Table 7: Determinants of a player's "opponent's expected consistency"

Variable	***Coefficient***	***Std. error***	***p-value***
dominant strategy	-0.628	0.068	0.000
dummy for **7th grade**	0.068	0.094	0.471
dummy for **9th grade**	0.304	0.120	0.011
dummy for **11th grade**	0.062	0.104	0.554
# of possible outcomes	0.010	0.028	0.724
ability to play chess	-0.050	0.073	0.490
math grade	-0.118	0.049	0.017
German grade	0.095	0.049	0.052
existence of siblings	0.040	0.029	0.168
Constant	-0.163	0.190	0.389

Probit model with a player's "opponent's expected consistency" as dependent variable.
N = 2.340; standard errors clustered for the 180 decision makers (45 in 5th grade, 41 in 7th grade, 52 in 9th grade, 42 in 11th grade).

Marginal effects of independent variables

	Marginal effect	***Std. error***	***p-value***
dominant strategy	-0.232	0.022	0.000
dummy for **7th grade**	0.026	0.036	0.471
dummy for **9th grade**	0.118	0.047	0.011
dummy for **11th grade**	0.024	0.040	0.554
# of possible outcomes	0.004	0.011	0.724
ability to play chess	-0.019	0.028	0.490
math grade	-0.045	0.019	0.017
German grade	0.037	0.019	0.052
existence of siblings	0.015	0.011	0.168

N = 2.340; standard errors clustered for the 180 decision makers (45 in 5th grade, 41 in 7th grade, 52 in 9th grade, 42 in 11th grade).
Note that in this estimation we have excluded cases where the opponent (the player about which expectations are formed) has a dominant strategy.

Table 8: Non-Randomness of decision making (significance-levels)

	CHOICES	***FIRST-ORDER BELIEFS***	***SECOND-ORDER BELIEFS***
Nash **5th grade**	0.10	0.01	0.01
Nash **7th grade**	0.05	0.01	0.01
Nash **9th grade**	0.01	0.01	0.01
Nash **11th grade**	0.10	0.01	0.01
Nash **overall**	0.01	0.01	0.01
	OWN	OPPONENT'S	
Consistency **5th grade**	0.01	0.01	
Consistency **7th grade**	0.01	0.01	
Consistency **9th grade**	0.01	0.01	
Consistency **11th grade**	0.01	0.01	
Consistency **overall**	0.01	0.01	
	OWN	OPPONENT'S	
Nash-CON **5th grade**	0.05	0.01	
Nash-CON **7th grade**	0.01	0.01	
Nash-CON **9th grade**	0.01	0.01	
Nash-CON **11th grade**	0.05	0.01	
Nash-CON **overall**	0.01	0.01	

N=191.

Results of Kolmogorov-Smirnov one sample tests, comparing the differences between the observed distributions and theoretical distributions in case of random decision making

0.01 – significant at 1% level; 0.05 – 5% level; 0.1 – 10% level)

Table 9: Estimated probability p_k and error rates ε_k of types
(**at the mean of the exogenous variable**)

	Probabilities			***Error rates***		
	Male	***Female***	***Difference***	***Male***	***Female***	***Difference***
Altruistic/Efficiency-loving	0.077**	0.004	0.073*	0.053	0.337	-0.284
Pessimistic	0.104**	0.079**	0.024	0.377	0.258	0.119
Optimistic	0.411***	0.541***	-0.130	0.078	0.251	-0.173***
Equilibrium	0.026	0.058*	-0.032	0.129	0.875^{+}	-0.747***
Sophisticated	0.118*	0.031	0.087	0.052	0.000	0.052
D1	0.082*	0.233***	-0.151**	0.315	0.377	-0.062
D2	0.000	0.000	0.000	1.000^{+}	1.000^{+}	0.000
L2	0.182**	0.054	0.128*	0.416	0.334	0.082
Sum Strategic types	0.409***	0.376***	0.033	-	-	-

N=179, Log likelihood = -1635.1764.
In columns “Difference” we test for gender differences (in a particular row)
*** (**) [*] significant at 1% (5%) [10%] level
$^{+}$ coefficient not significantly different from 1 (i.e., not significantly different from random)

Table 10: Marginal effect on type probabilities and error rates

		Probabilities		*Error rates*	
		Male (Std. Error)	*Female (Std. Error)*	*Male (Std. Error)*	*Female (Std. Error)*
Altruistic/Efficiency-loving	*age*	0.032 (0.021)	0.004 (0.007)	0.058* (0.035)	0.005 (0.014)
	math grade	-0.008 (0.031)	0.001 (0.002)	0.062* (0.037)	0.275 (0.256)
	German grade	0.008 (0.041)	-0.001 (0.002)	-0.028 (0.026)	-0.125 (0.170)
Pessimistic	*age*	0.028 (0.027)	-0.020 (0.018)	0.097** (0.044)	-0.212*** (0.055)
	math grade	0.071* (0.041)	0.073** (0.030)	0.235* (0.099)	0.192** (0.053)
	German grade	-0.071 (0.052)	-0.078** (0.038)	0.050 (0.102)	0.041 (0.092)
Optimistic	*age*	-0.072** (0.036)	0.005 (0.029)	0.000 (0.010)	0.006 (0.012)
	math grade	-0.040 (0.060)	0.073 (0.066)	-0.017 (0.011)	-0.044 (0.029)
	German grade	0.103 (0.076)	0.042 (0.040)	0.023* (0.013)	0.061* (0.035)
Equilibrium	*age*	0.010 (0.010)	-0.002 (0.013)	0.056 (0.104)	-0.001 (0.037)
	math grade	-0.033 (0.021)	-0.060 (0.037)	0.085 (0.099)	0.083 (0.078)
	German grade	0.027 (0.019)	0.042 (0.039)	0.088 (0.100)	0.086 (0.097)
Sophisticated	*age*	0.008 (0.026)	-0.011 (0.011)	-0.019 (0.025)	0.000*** (0.000)
	math grade	-0.025 (0.051)	0.001 (0.014)	0.095 (0.061)	0.000*** (0.000)
	German grade	0.074 (0.050)	0.010 (0.018)	0.038 (0.025)	0.000 (0.000)
D1-type	*age*	0.014 (0.023)	0.043* (0.025)	0.031 (0.034)	-0.014 (0.025)
	math grade	-0.065* (0.035)	-0.130** (0.065)	0.070 (0.068)	0.076 (0.064)
	German grade	0.071* (0.038)	0.131* (0.070)	-0.054 (0.071)	-0.059 (0.070)
D2-type	*age*	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	math grade	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.008)
	German grade	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.003)
L2-type	*Age*	-0.020 (0.031)	-0.019 (0.013)	-0.159*** (0.049)	-0.059 (0.072)
	math grade	0.099 (0.061)	0.042* (0.024)	-0.023 (0.079)	-0.021 (0.074)
	German grade	-0.212** (0.091)	-0.079** (0.035)	0.531*** (0.204)	0.487* (0.254)
Sum Strategic types	*Age*	0.012 (0.033)	0.011 (0.028)	-	-
	math grade	-0.023 (0.064)	-0.147** (0.068)	-	-
	German grade	-0.041 (0.080)	0.104 (0.079)	-	-

N=179, Log likelihood = -1635.1764.

*** (**) [*] significant at 1% (5%) [10%] level. Note that lower math grades denote higher skills.

Figure 1: The 18 normal-form games

game # 3 (*D*) **[1R, 2R]**

	A, P, O	*D12, L2, E, S*
A	72, 93	31, 46
D	84, 52	55, 79

game # 13 (*D*) **[1R, 2R]**

	A, P, O	*D12, L2, E, S*
D	94, 23	38, 57
A	45, 89	14, 18

game # 1 (*D*) **[2R, 1R]**

	D	*A*
D12, L2, E, S	75, 51	42, 27
A, P, O	48, 80	89, 68

game # 12 (*D*) **[2R, 1R]**

	D	*A*
A, P, O	21, 92	87, 43
D12, L2, E, S	55, 36	16, 12

game # 7 (*D*) **[2R, 1R]**

		D	*A*
A, P, O	59, 58	46, 83	85, 61
D12, L2, E, S	38, 29	70, 52	37, 23

game # 11 (*D*) **[2R, 1R]**

	A	*D*
D12, L2, E, S	31, 32	68, 46
P	72, 43	47, 61
A, O	91, 65	43, 84

game # 9 (*D*) **[1R, 2R]**

	D12, L2, E, S	*A, P, O*
	28, 37	57, 58
A	22, 36	60, 84
D	51, 69	82, 45

game # 16 (*D*) **[1R, 2R]**

	D12, L2, E	*P*	*A, O, S*
D	42, 64	57, 43	80, 39
A	28, 27	39, 68	61, 87

game # 6 (*ND*) **[3R, 2R]**

	A, P, O, S^{16}	*D12, L2, E, $S^{10,12,14}$*
A	53, 86	24, 19
P, O, D1, L2, S	79, 57	42, 73
D2, E	28, 23	71, 50

game # 15 (*ND*) **[3R, 2R]**

	A, P, O	*D12, L2, E, S*
A	76, 93	25, 12
D2, E	43, 40	74, 62
P, O, D1, L2, S	94, 16	59, 37

game # 14 (*ND*) **[2R, 3R]**

	A	*D2, E*	*P, O, D1, D2, S*
D12, L2, E, S	21, 26	52, 73	75, 44
A, P, O	88, 55	25, 30	59, 81

game # 2 (*ND*) **[2R, 3R]**

	D2, E	*A*	*P, O, D1, D2, S*
A, P, O, S^{14}	42, 45	95, 78	18, 96
D12, L2, E, $S^{10,12,16}$	64, 76	14, 27	39, 61

game # 8 (*ND*) **[∞R, ∞R]**

	O, D12, L2, S	*A, P*	*E*
L2, E, $S^{10,12,16}$	87, 32	18, 37	63, 76
A, P, O, D12, S^{14}	65, 89	96, 63	24, 30

game # 10 (*ND*) **[∞R, ∞R]**

	O, D12, L2, S	*A, P*	*E*
A, P, O, D12, S^{14}	67, 91	95, 64	31, 35
L2, E, $S^{10,12,16}$	89, 49	23, 53	56, 78

game # 5 (*ND*) **[∞R, ∞R]**

	L2, E, S	*A, P, O, D12*
E	72, 59	26, 20
A, P	33, 14	59, 92
O, D12, L2, S	28, 83	85, 61

game # 4 (*ND*) **[∞R, ∞R]**

	L2, E, S	*A, P, O, D12*
A, P	46, 16	57, 88
E	71, 49	28, 24
O, D12, L2, S	42, 82	84, 60

game # 17 (*D*) **[1R, 2R]**

	D12, L2, E, S	*A, P, O*
	22, 14	57, 55
	30, 42	28, 37
A	15, 60	61, 88
D	45, 66	82, 31

game # 18 (*D*) **[2R, 1R]**

			A	*D*
A, P, O	56, 58	38, 29	89, 62	32, 86
D12, L2, E, S	15, 23	43, 31	61, 16	67, 46

In italics we denote the strategies of the different strategic or non-strategic types. A … *altruistic*; P … *pessimistic*; O … *optimistic*; *L2, D1, D2* … strategic types (see section 3.4 in paper); E … *equilibrium*; S … *sophisticated* (S^{14} … sophisticated against students in 9th grade; S^{16} … sophisticated against students in 11th grade; $S^{10,\ 12,\ 16}$ … sophisticated against students in 5th, 7th, and 11th grade; $S^{10,\ 12,\ 14}$ … sophisticated against students in 5th, 7th, and 9th grade); D12 = D1 and D2; D = dominant strategy.

Figure 2: Relative frequency distribution of choosing the Nash-strategy for each grade

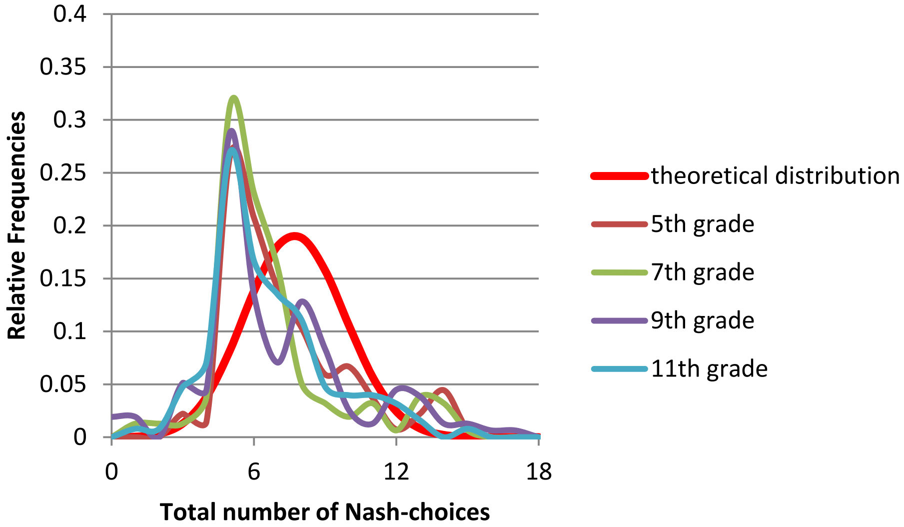


Figure 3: Relative frequency distribution of choosing the Nash-strategy for choices, first- and second-order beliefs

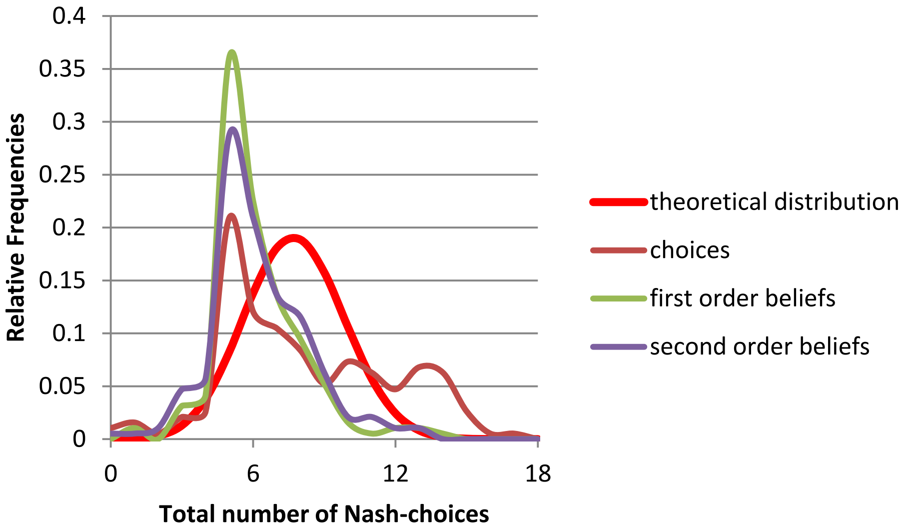


Figure 4: Relative frequency of consistent decisions for each grade

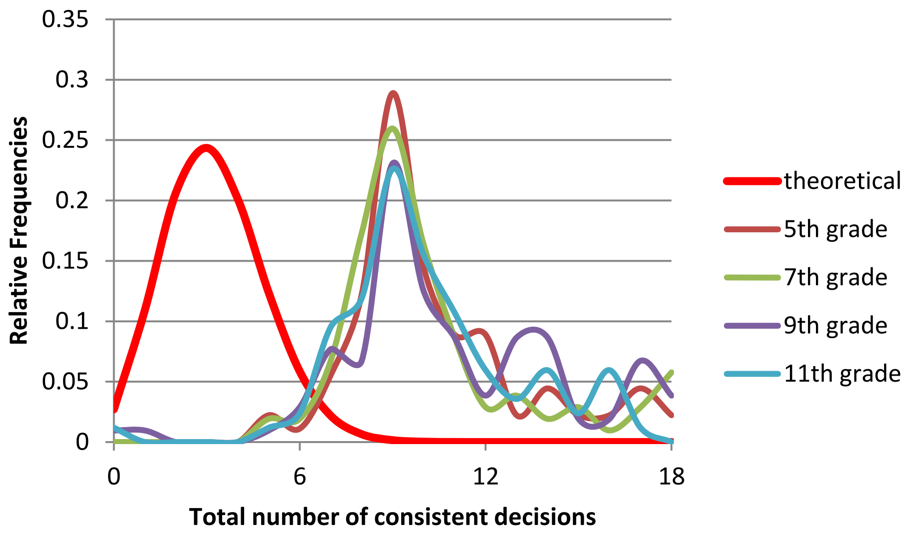
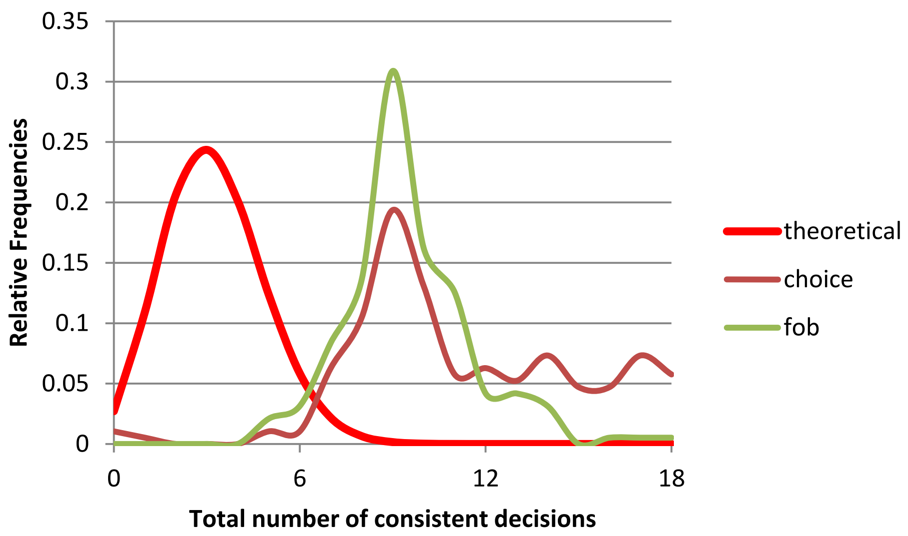


Figure 5: Relative frequency of consistent decisions for choices first- and second-order beliefs



Appendix

A. Experimental instructions

Instructions were read aloud at the beginning of each session. Before the experiment started, all participants had to answer control questions in order to make sure that they understood the instructions. Instructions primarily served as a basic guideline for a very detailed explanation of the game. As we wanted to be sure that our participants understood the rules of the game we took a lot of effort to explain them the instructions in a very detailed way and if necessary even personally. Originally all instructions were in German. In the following we present an English translation of the instructions and control questions used.

Instructions for 5th and 7th graders differ from instructions for 9th and 11th graders with respect to payoffs. In the following we present instructions for 5th and 7th graders as a baseline and indicate payoffs which were used for 9th and 11th graders in brackets and underlined letters.

Welcome! In this game it is very important that you do not communicate with any of your class mates for the whole duration of the game. Students who break this rule will be excluded from the game. You will earn some money by playing this game which will be paid to you at the end of the game. The amount of money you earn strongly depends on your decisions during the game. Thus it is very important that you understand the rules of the game. As soon as you have any questions, please raise your hand and an instructor of the game will come to you in order to answer your questions.

Today we will play a game consisting of 18 sub-games. Each sub-game is printed on a separate sheet of paper and consists of three decisions each player needs to make. The rules for those three decisions will be explained to you soon. Each player gets 18 sheets on which he needs to make his decisions.

You will play in groups of two students. Each student of this room will be matched with a student from another room. In this other room students from your parallel-class play the same game right now. The student with the number one in this room will be matched with student number one in the other room, student number two in this room will be matched with student number two in the other room etc. Same as you did, students in the other room have drawn numbers randomly. We take care that each student in this room is paired with a student in the other room. In the following we will call the student who is matched with you as “the other player” or “your interaction partner”.

We would like to explain the rules of the game based on the following examples:

Example for decision 1:

Which row do you want to choose?

		§§	%%	@@	µµ
O	§	65	47		
O	%	23	93		
	@				
	µ				

Your own payments

	§§	%%	@@	µµ
§	57	32		
%	85	71		
@				
µ				

Payments for the other player

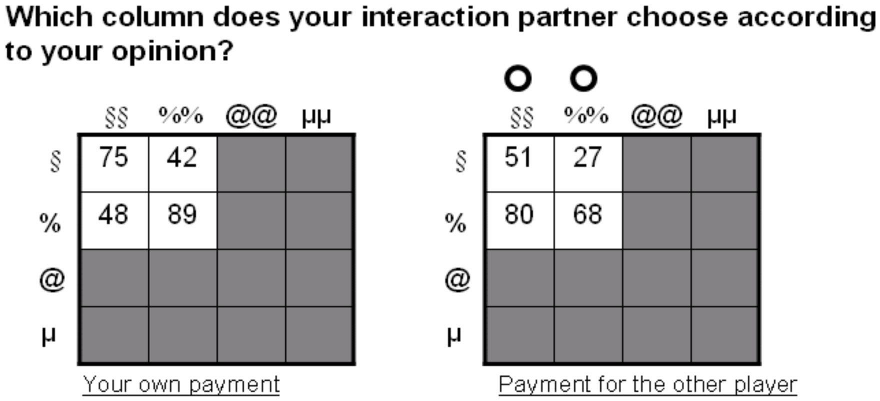
In the upper part of each sheet you need to make a decision. In this example you need to choose one of two options. The first option (first row) is indicated with the sign § and the second option (second row) is indicated with the sign %. Your interaction partner also has two options in this example. His options are columns which are always indicated with two signs of the same kind. The first option of the other player is indicated with §§ and the second option is indicated with %%.

Based on the row that you have chosen and the column that your interaction partner has chosen the potential payment for both players is determined. This is done as follows: The table on the left hand side contains four possible payments dedicated to you and measured in game points. The table on the right hand side contains four possible payments dedicated for the opponent player – again measured in game points. The cell that is relevant for payment is dependent on your own decision and on the decision of your interaction partner.

Let’s assume you choose the row indicated with the sign § and your interaction partner chooses the column indicated with the signs §§. In this particular case you would get 65 points and your interaction partner would get 57 points. Let’s assume another case in which you choose the row indicated with the sign % and the other player chooses the column indicated with the signs §§. In this case you would get 23 points and your interaction partner would get 85 points.

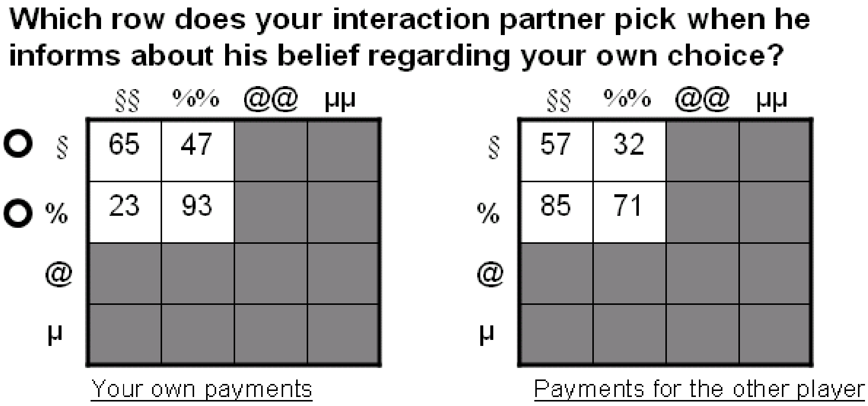
You can make your choice by marking the circle next to the sign you want to pick with a cross.

Example for decision 2:



In the middle of each sheet we ask you for your opinion about your interaction partner's choice.
For example if you belief that your interaction partner chooses the column indicated by the signs §§ you need to mark the circle above those signs with a cross. If you think that your interaction partner chooses the column marked with the signs %% you need to mark the circle above the signs %% with a cross. Please note that correct beliefs can result in considerable higher payments.

Example for decision 3:



In the lower part of each sheet we ask you for your opinion about your interaction partner's belief regarding your own choice.
For example if you think that your interaction partner expects you to choose the row indicated with the sign § you need to mark the circle next to the sign § with a cross. If you think that your interaction partner expects you to choose the row indicated with the sign % you need to mark the circle next to the sign % with a cross.
Please note that correct beliefs can result in considerable higher payments.

In the example illustrated above each player has two options from which he can choose. In general there are also games in which one of the two players has not only two, but three or four options from which he can choose.

Calculation of payments:

At the end of the whole experiment we will randomly select one specific sub-game which is relevant for your payment.

In order to realize that we provide a deck of cards showing the numbers 1 to 18 from which you (or the other player) will be asked to draw one card. The number shown on this card determines the sub-game which is relevant for your payment.

In addition one specific decision will be selected, which in combination with the selected sub-game will determine your payment.

In this context relevant decisions are:

- which row is your own choice (decision 1)
- which column chooses your interaction partner according to your opinion (decision 2)
- which row does your interaction partner pick when he informs about his belief regarding your own choice (decision 3)

The random selection of the decision that is relevant for your payment will be realized by drawing another card from a deck of three cards showing numbers 1-3.

Payment for decision 1:

In the case that the second card you (or the other player) have drawn shows the number 1, you and your interaction partner get paid for the first decision. Each point will be converted into Euro according to the following exchange rate

1 point = 0.10 Euros <u>(0.20 Euros)</u>

Let's assume you have chosen § and your interaction partner has chosen %%. In the example illustrated before this would mean that you earn 4.7 Euros <u>(9.4 Euros)</u> (0.1 * 47 points) <u>(0.2 * 47 points)</u> while your interaction partner earns 3.2 Euros <u>(6.4 Euros)</u> (0.1 * 32 points) <u>(0.2 * 32 points)</u>.

Payment for decision 2:

In the case that the second card you (or the other player) have drawn shows the number 2, you and your interaction partner get paid for the second decision. In this case you will get 5 Euros <u>(10 Euros)</u> if your prediction of your interaction partner's choice was correct. If your prediction was incorrect you will get no payment in this case.

Payment for decision 3:

In the case that the second card you (or the other player) have drawn shows the number 3, you and your interaction partner get paid for the third decision. In this case you get 5 Euros <u>(10 Euros)</u> if your prediction of your interaction partner's expectation (regarding your own choice) was correct. If your prediction was incorrect you will get no payment in this case.

In addition to the payments described in the former each participant of the game gets a fixed amount of 2 Euros <u>(4 Euros)</u> for attending the game.

Which row do you want to choose? Example 1

		§§	%%	@@	µµ
⊗	§	77	39		
○	%	69	55		
	@				
	µ				

Your own payments

	§§	%%	@@	µµ
§	51	62		
%	35	44		
@				
µ				

Payments for the other player

Which column does your interaction partner choose according to your opinion?

	§§	%%	@@	µµ
§	77	39		
%	69	55		
@				
µ				

Your own payments

	§§	%%	@@	µµ
	○	⊗		
§	51	62		
%	35	44		
@				
µ				

Payments for the other player

Which row does your interaction partner pick when he informs about his belief regarding your own choice?

		§§	%%	@@	µµ
⊗	§	77	39		
○	%	69	55		
	@				
	µ				

Your own payments

	§§	%%	@@	µµ
§	51	62		
%	35	44		
@				
µ				

Payments for the other player

Participant ID: _________ **Class ID: ______**

Control-questions:

1) Assume you and your interaction partner get paid for decision 1.

   a) what is your payment in points / Euro in the case your interaction partner has chosen §§?

   __________ points ________ Euro

   b) what is your interaction partners payment in points Euro, in the case he has chosen %%?

   __________ points ________ Euro

   c) what is your payment in points / Euro in the case your interaction partner has chosen %%?

   __________ points ________ Euro

2) Assume you and your interaction partner get paid for decision 2.

   a) what is your payment in the case your interaction partner has chosen §§ when he took decision 1?

   __________ Euro

3) Assume you and your interaction partner get paid for decision 3.

   a) what is your payment in the case your interaction partner has chosen § when he took decision 2?

   __________ Euro

Which row do you want to choose? Example 2

		§§	%%	@@	µµ
○	§	67	33		
○	%	30	28		
○	@	47	68		
⊗	µ	71	32		

Your own payments

	§§	%%	@@	µµ
§	55	34		
%	69	34		
@	37	53		
µ	75	60		

Payments for the other player

Which column does your interaction partner choose according to your opinion?

	§§	%%	@@	µµ
§	67	33		
%	30	28		
@	47	68		
µ	71	32		

Your own payments

	§§	%%	@@	µµ
	⊗	○		
§	55	34		
%	69	34		
@	37	53		
µ	75	60		

Payments for the other player

Which row does your interaction partner pick when he informs about his belief regarding your own choice?

		§§	%%	@@	µµ
○	§	67	33		
○	%	30	28		
○	@	47	68		
⊗	µ	71	32		

Your own payments

	§§	%%	@@	µµ
§	55	34		
%	69	34		
@	37	53		
µ	75	60		

Payments for the other player

Participant ID: _________ **Class ID: ______**

Control-questions:

1) Assume you and your interaction partner get paid for decision 1.

   a) what is your payment in points / Euro in the case your interaction partner has chosen §§?

   __________ points ________ Euro

   b) what is your interaction partners payment in points / Euro, in the case he has chosen §§?

   __________ points ________ Euro

   c) what is your payment in points / Euro in the case your interaction partner has chosen %%?

   __________ points ________ Euro

2) Assume you and your interaction partner get paid for decision 2.

   a) what is your payment in the case your interaction partner has chosen §§ when he took decision 1?

   __________ Euro

3) Assume you and your interaction partner get paid for decision 3.

   a) what is your payment in the case your interaction partner has chosen @ when he took decision 2?

   __________ Euro

B. Additional Tables

Table B.1: Percentages of decisions that comply with equilibrium, contingent on game-type

	Complexity (game # for rows // game # for columns)	**5th grade**	**7th grade**	**9th grade**	**11th grade**
Choice	*1 round of dominance to identify own equilibrium choice*				
	2x2 with dominant decision (#3, #13 // #1, #12)	91.11	85.58	82.69	84.52
	2x3 with dominant decision (#16 // #11)	93.33	90.38	80.77	83.33
	3x2 with dominant decision (#9 // #7)	82.22	86.54	88.46	73.81
	4x2 with dominant decision (#17 // #18)	91.11	82.69	82.69	83.33
	2 rounds of dominance				
	2x2, partner has dominant decision (#1, #12 // #3, #13)	41.11	45.19	45.19	34.52
	2x3, partner has dominant decision (#7 // #9)	40.00	32.69	40.38	33.33
	3x2, partner has dominant decision (#11 // #16)	31.11	34.62	36.54	35.71
	2x4, partner has dominant decision (#18 // #17)	44.44	32.69	46.15	42.86
	2x3 with 2 rounds of dominance (#2, #14 // #6, # 15)	38.89	35.58	36.54	28.57
	3 rounds of dominance				
	3x2 with 3 rounds of dominance (#6, #15 // #2, #14)	8.88	7.69	7.69	8.33
	No dominance				
	2x3, unique equilibrium, no dominance (#8, #10 // # 4, #5)	43.33	37.50	35.58	30.95
	3x2, unique equilibrium, no dominance (#4, #5 // #8, #10)	16.67	19.23	17.31	10.71
First order beliefs	*1 round of dominance to identify own equilibrium choice*				
	2x2 with dominant decision (#3, #13 // #1, #12)	87.78	86.54	89.42	89.29
	2x3 with dominant decision (#16 // #11)	86.67	98.08	84.62	90.48
	3x2 with dominant decision (#9 // #7)	82.22	88.46	88.46	80.95
	4x2 with dominant decision (#17 // #18)	91.11	78.85	80.77	83.33
	2 rounds of dominance				
	2x2, partner has dominant decision (#1, #12 // #3, #13)	15.56	14.42	19.23	16.67
	2x3, partner has dominant decision (#7 // #9)	26.67	11.54	26.92	16.67
	3x2, partner has dominant decision (#11 // #16)	0.00	7.69	5.77	11.90
	2x4, partner has dominant decision (#18 // #17)	20.00	5.77	15.38	21.43
	2x3 with 2 rounds of dominance (#2, #14 // #6, # 15)	14.44	15.38	15.38	11.90
	3 rounds of dominance				
	3x2 with 3 rounds of dominance (#6, #15 // #2, #14)	13.33	10.58	7.69	5.95
	No dominance				
	2x3, unique equilibrium, no dominance (#8, #10 // # 4, #5)	16.67	19.23	22.12	21.43
	3x2, unique equilibrium, no dominance (#4, #5 // #8, #10)	12.22	5.77	8.65	10.71
Second order beliefs	*1 round of dominance to identify own equilibrium choice*				
	2x2 with dominant decision (#3, #13 // #1, #12)	80.44	82.69	80.77	91.67
	2x3 with dominant decision (#16 // #11)	71.11	86.54	69.23	76.19
	3x2 with dominant decision (#9 // #7)	73.33	82.69	88.46	80.95
	4x2 with dominant decision (#17 // #18)	75.56	78.85	69.23	88.10
	2 rounds of dominance				
	2x2, partner has dominant decision (#1, #12 // #3, #13)	30.00*	16.35*	27.88*	16.67*
	2x3, partner has dominant decision (#7 // #9)	28.89	13.46	17.31	23.81
	3x2, partner has dominant decision (#11 // #16)	15.56	13.46	15.38	11.90
	2x4, partner has dominant decision (#18 // #17)	20.00	9.62	19.23	7.14
	2x3 with 2 rounds of dominance (#2, #14 // #6, # 15)	20.00	13.46	11.54	9.52
	3 rounds of dominance				
	3x2 with 3 rounds of dominance (#6, #15 // #2, #14)	22.22**	8.65**	8.65**	8.33**
	No dominance				
	2x3, unique equilibrium, no dominance (#8, #10 // # 4, #5)	22.22	19.23	25.96	29.76
	3x2, unique equilibrium, no dominance (#4, #5 // #8, #10)	20.00	10.58	13.46	8.33

** (*) significant difference at $p < 0.05$ ($p < 0.1$) across all age groups for equilibrium play in a particular set of games according to a two-sided Kruskal-Wallis test.