Journal of Economic Behavior and Organization 194 (2022) 478–522

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Gender differences in preferences of adolescents: Evidence from a large-scale classroom experiment☆

Dániel Horn[a,b], Hubert János Kiss[a,b,*], Tünde Lénárd[c,a]

[a] KRTK KTI, Tóth Kálmán u. 4., Budapest, 1097, Hungary
[b] Department of Economics, Corvinus University of Budapest, Fővám tér 8, Budapest, 1093, Hungary
[c] SOFI, Stockholm University, Stockholm, SE-106 91, Sweden

ARTICLE INFO

Article history:
Received 14 January 2021
Revised 3 October 2021
Accepted 12 December 2021

JEL classification:
C80
C90
D91

Keywords:
Adolescents
Altruism
Competitiveness
Cooperation
Dictator game
Patience
Present bias
Public goods game
Risk preferences
Social preferences
Time inconsistency
Time preferences
Trust
Trustworthiness

ABSTRACT

In this study, we estimate unadjusted and adjusted gender gaps in time preference, risk attitudes, altruism, trust, trustworthiness, cooperation, and competitiveness using data on 1088 high school students from 53 classes. These data, collected by running incentivized experiments in Hungarian classrooms, are linked to an administrative data source on the students' standardized test scores, grades and family background. After taking into account class fixed effects, we find that females are significantly more altruistic, but are less present biased, less risk tolerant, less trusting, less trustworthy, and less competitive than males. At the same time we do not observe significant gender differences in patience, time inconsistency and cooperation at the 5% significance level. We also show that most of these initial gender differences do not change even if we control for age, family background, cognitive skills and school grades in a regression framework. We risk over-control when we include the time spent on each task as well as the other preference domains in our regressions, but the gender gap remains significant in social preferences (altruism, trust and trustworthiness), present bias and competition.

1. Introduction

Preferences matter. A large and growing literature shows that preferences strongly predict a wide array of real-life outcomes, including educational, labor market, financial and health choices (Dohmen et al., 2011; Falk et al., 2018; Golsteyn et al., 2014; Moffitt et al., 2011). Special attention has been given to gender differences in preferences as these may lead

☆ The project has been funded by the National Research and Development Office of Hungary (project no. 124396). The experiments were run in Hungarian, and the related legal documents are available in Hungarian here: https://www.mtakti.hu/kapcsolat/altalanos-tajekoztato-a-kiserletekrol/.

* Corresponding author at: KRTK KTI, Tóth Kálmán u. 4., Budapest, 1097, Hungary.
E-mail addresses: horn.daniel@krtk.hu (D. Horn), kiss.hubert@krtk.hu (H.J. Kiss), tunde.lenard@sofi.su.se (T. Lénárd).

https://doi.org/10.1016/j.jebo.2021.12.015

to inefficient social outcomes (Blau and Kahn, 2017; Buser et al., 2014; Ellison and Swanson, 2010; Paglin and Rufolo, 1990).

Preferences evolve throughout childhood and adolescence, and there are several studies shedding light on how gender shapes preferences, besides other important determinants like socioeconomic status. Understanding gender differences in preferences in childhood and adolescence is important as those preferences seem to be more malleable at younger ages (Ertac, 2020). Moreover, the gender intensification theory in psychology (Hill and Lynch, 1983) posits that adolescence reinforce societal expectations for gender-typed behavior (Rose and Rudolph, 2006), so investigating the factors that shape preferences in this age is an important endeavor.[1]

This paper investigates gender differences in time, risk, social and competitive preferences of high-school students using incentivized lab-in-the-field experiments conducted in 9 schools' 53 school classes, with overall 1088 students in Hungary.[2] The four most widely researched preference domains were measured in detail. We used the staircase (or unfolding brackets) method to measure time preferences (see Cornsweet, 1962; Falk et al., 2018). Following the beta-delta model proposed by Phelps and Pollak (1968) and Laibson (1997) we focus on the individual discount factor capturing the patience of the students (we often refer to it as *delta*), and on time consistency (we often call it *beta*) as well as on present-bias (when $beta<1$). We opted for the bomb risk elicitation task (Crosetto and Filippin, 2013) to assess *risk preferences*. We were interested in various aspects of social preferences, so we measured altruism (proxied by the dictator game), trust and trustworthiness (with the trust game), and cooperation (with a two-person public goods game). Moreover, we used the dictator game to assess altruism toward a classmate and a schoolmate, varying the degree of social distance between the dictator and the recipient. We call our corresponding measures *altruism* / *trust* / *trustworthiness* / *cooperation*, higher values of the measures indicating a higher level of the given preference. Competitive preferences were estimated using the established experimental procedure by Niederle and Vesterlund (2007) and we refer to this measure as *competition*. To ensure that the different tasks do not affect each other (e.g. receiving a low amount in the dictator game may influence how much a subject gives in the trust game), there was no feedback until the end of the experiment. A major strength of our study is that besides these preference measures and the related information gained from the tasks (time spent on each task), we obtained rich background information on the subjects from an administratively collected individual-level data source on the students' previous cognitive abilities (proxied by their standardized test scores in mathematics and literacy), school performance (grades) and family background.

There is consensus in the literature (see meta-analyses by Croson and Gneezy, 2009; Bertrand, 2011; Niederle, 2016) that there is a substantial gender difference in competitive preferences and no gender difference in time preferences.[3] However, there is an ongoing debate if there are gender differences in risk and social preferences. While Eckel and Grossman (2008), Croson and Gneezy (2009) and Bertrand (2011) claim that a robust gender difference exists in risk attitudes (women being more risk-averse), Niederle (2016) shows convincingly that this finding is dependent on the elicitation technique, a finding confirmed by Filippin and Crosetto (2016). Regarding social preferences, Bertrand (2011) argues that gender differences exist, Croson and Gneezy (2009) emphasize that women react in a more sensitive way to cues in the experimental context than men, and Niederle (2016) calls for further investigation to see if there are indeed gender differences. In a recent meta-analysis of the literature on the preferences of children and adolescents, Sutter et al. (2019) report findings mostly in line with the previous results that were blind to the age of the subjects. That is, findings on gender difference in time preferences point in all directions and are not conclusive, no clear gender differences have been reported in bargaining situations (captured by the ultimatum and the trust games) and in cooperation, but there are gender differences in risk preferences (females being more risk-averse), in individual decision-making (proxied by the dictator game where females are more altruistic), and in competitiveness (females being less competitive).

In Figs. 1–6 we directly compare our results to the studies that investigate preferences in the adolescence that use the same or a very similar elicitation tasks that we do.[4] These figures contain the rescaled coefficient of the female dummy that takes into account the features of the elicitation tasks and makes them comparable, by calculating the share of the endowment given / contributed / risked / transferred in each task. Standard errors were also rescaled accordingly. Albeit we report the most comprehensive specification of the studies, they are still not directly comparable to ours, as we use an extensive set of controls (in many cases we use much more controls than the studies that we consider, see Tables D.19–D.24 in Appendix D). In these figures we report the results from our full models, thus, it is no surprise that our coefficients are often closer to zero than that of the comparable studies, and hence we also believe they are closer to the real gender gap in each preference. We see that our findings are in line with most of these results. That is, females seem to be less risk tolerant, more altruistic and less competitive than their male counterparts, and there seems to be no gender difference in cooperation. Regarding time preferences, trust and trustworthiness the picture is still blurred.

[1] Andersen et al. (2013) and Alan and Ertac (2019) illustrate this point related to the emergence of gender difference in competitiveness with evidence from field experiments.

[2] 1108 students participated in our experimental sessions, but we know the gender of only 1088 students. As the paper focuses on gender differences, we use these 1088 observations.

[3] There are just a very few papers that report no gender differences in competitiveness (Price, 2012) or that document gender differences in time preferences (Dittrich and Leipold, 2014)

[4] For details on which studies we include, and short summaries of each see Appendix D.

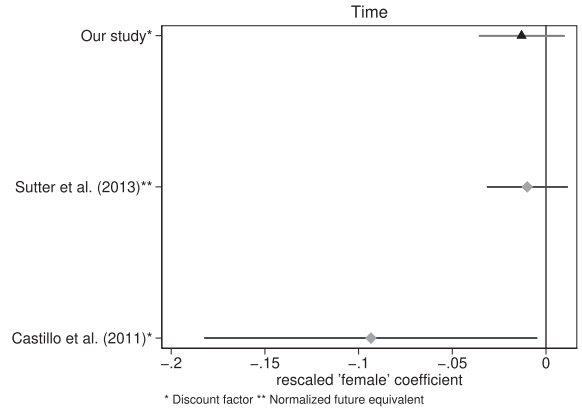

Fig. 1. Coefficient plots from the literature - Time preferences.

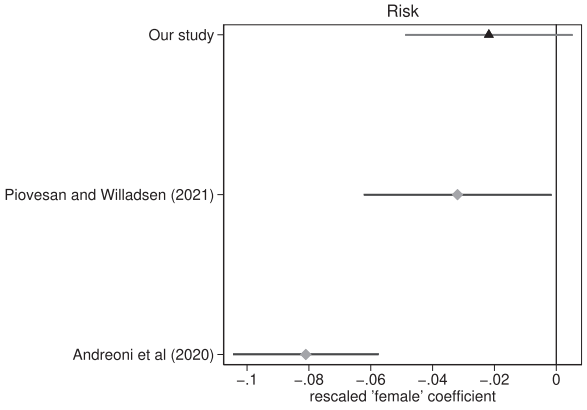

Fig. 2. Coefficient plots from the literature - Risk tolerance.

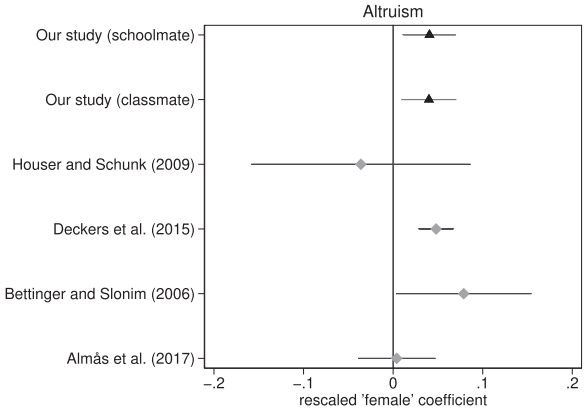

Fig. 3. Coefficient plots from the literature - Altruism.

This paper complements previous research on gender differences in preferences in two ways. First, we measure nine aspects of the four most widely used preferences *at once*, so we can measure gender differences more precisely, conditional on correlated preferences. The issue of correlated preferences has been addressed in some cases. For instance, risk preferences are often controlled for when measuring competitiveness, (see, for instance Buser et al., 2014), or when investigating time preferences, (see, for instance Andersen et al., 2008), but the same is typically not done when analyzing other preferences. Little is known about the rest of the potential correlations between the measured preferences and their effects on the gender gap. If there are significant correlations between the preferences - as we see in our data - not taking this into account might lead to an incorrect interpretation of gender differences in the various preference domains. While risk has been suspected of playing a role in many of the other preferences the same might be true for other preferences. For instance, we

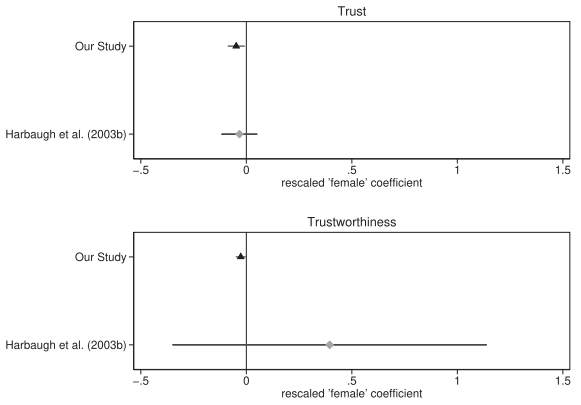

Fig. 4. Coefficient plots from the literature - Trust.

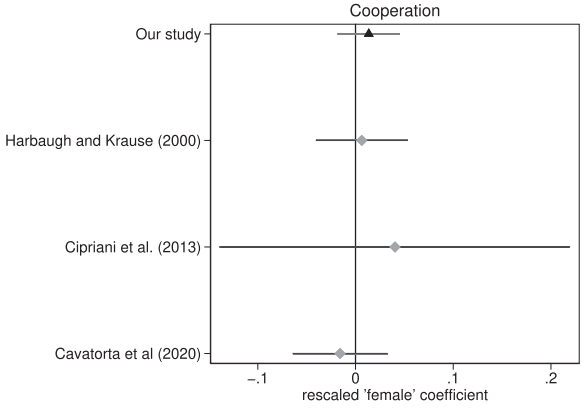

Fig. 5. Coefficient plots from the literature - Cooperation.

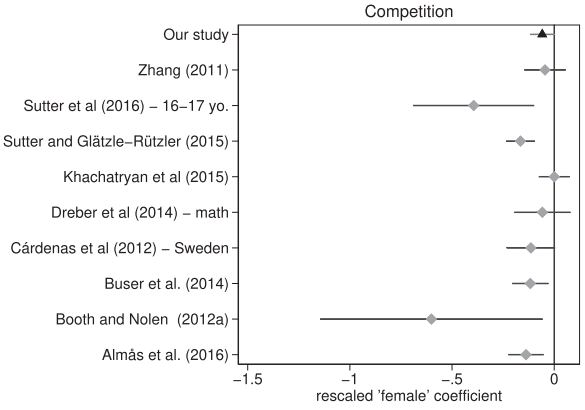

Fig. 6. Coefficient plots from the literature - Competition.

see that patience (captured by delta) correlates significantly with the amount sent in the trust game and the contribution in the public goods game. Therefore, estimating the gender gap in trust without taking into account the potential differences in patience could lead to under- or overestimated gender gaps. This study investigates if omitting other preferences is conducive to such issues or not.

Second and related to the previous point, given our rich data including information on family background, cognitive abilities and school performance (controls that have been found important determinants of preferences in the literature, see Sutter et al. (2019)) we can assess how the unadjusted gender differences change as we take into account more and more controls. In other words, we are able to see when the gender difference disappears (if it does at all), which may shed light on factors mediating the gender difference in the given domain. Note also that even though our sample is not representative,

by controlling for the factors mentioned above, we can account for potential individual confounders in a way that if we find gender difference in a given domain, then the difference is likely to be genuine.

Without any controls, we observe gender differences in all the preferences, females being less patient, less present-biased, more time inconsistent, more risk-averse, more altruistic (both with classmates and schoolmates), less trusting, less trustworthy, less cooperative and less competitive than males. However, our preferred baseline is when class fixed effects are taken into account as they control for many unobserved factors that affect the same group of students. Once class fixed effects are considered, the gender difference in time inconsistency and cooperation vanishes. After applying an extensive set of further controls, including age, family background, cognitive abilities and school grades, gender differences disappear in patience, but interestingly the rest of the previous findings do not change. Hence, there seems to be a solid gender difference in present bias, risk preferences, altruism, trust, trustworthiness and competitiveness in adolescence. If we take the analysis one step further and control also for the time spent on each task as well as the total time spent on the experiment, the gender gap in present bias weakens to marginal significance, but risk, competitiveness and the previously significant social preferences remain significant. By controlling for the rest of the preferences (risking over-control) the gender gap disappears in risk preferences and becomes only marginally significant in competitiveness, but significant gender difference (at 5%) remains in altruism, trust and trustworthiness. That is, gender differences in social preferences remain strong even if we control for family background, cognitive skills, school performance, time spent on the experiment and all other preferences. However, risk seems to be intertwined with other preferences, hence the insignificant gender gap.

One interesting finding from our data is that females are more altruistic (independently of the social distance from the recipient), while males are more trusting and trustworthy. We observe that in the dictator game females are more likely to split their endowment evenly than males, while in the trust game males are more likely to send all their endowment. These results are in line with previous findings in the literature according to which females are more egalitarian and concerned about payoff inequality (Fehr et al., 2013), while males are more efficiency-oriented and so more likely to make decisions that enhance the pie (Almås et al., 2010; Sutter et al., 2018).

The study is organized as follows. Section 2 contains information about the experimental tasks that we used, the procedures, the rest of the variables and a descriptive analysis of the data. In Section 3, we present the results, and Section 4 concludes.

2. Data

From March 2019 to March 2020, we visited 9 secondary schools in Hungary to assess the economic preferences of students. Overall, we measured time, risk, social and competitive preferences of 1088 students in 53 school classes (groups of students studying the major subjects together as of the start of their upper-secondary education). The experiments were anonymous, but we can link the preference measures to individual-level data from the National Assessment of Basic Competences (NABC) (for details see Sinka, 2010), providing useful information about the participants' previous standardized test scores, school grades and family background. With the detailed preference map of the students and the additional information on their background and school performance, we can study how gender differences in preferences observed in adolescence depend on other observable factors.

In this section, first we briefly describe the procedures related to the experiments and the experimental tasks that we used. Then, we present our variables related to family background and cognitive abilities. We finish the section with some descriptive statistics of the variables. For a more detailed description of the collection of the experimental data see Horn et al. (2020).

2.1. Procedures

At the beginning of the project, we contacted all educational providers in Hungary with at least one secondary school to request permission to run the experiment in their institutions. Our sample contains schools that were either suggested by the provider and schools that - once the provider gave permission - indicated voluntarily their willingness to participate. Half of the sample operates in Budapest and the other half in smaller rural towns of Hungary.

Our sample is not representative of the total school population of Hungary. The socioeconomic status of the participating students is higher than that of the corresponding population. In terms of school performance, students in our sample achieved a higher average mathematics test score on the NABC in 6th grade than the population of all 6th-grade students in 2017.[5]

After arranging the schedule with the schools, but before the experiment, we sent out a data protection statement to all parents and children, explaining that we would ask for the students' IDs used in the NABC so that we would be able

[5] In Horn et al. (2020) we provide more information on the differences between the sample used in this study and the overall student population in Hungary.

to connect our experimental data to anonymous NABC data on school performance and socioeconomic background at the individual level.[6] Participation was voluntary and anonymous.[7]

On experiment day, we unpacked our laptops in the school in a designated classroom, turning it into our laboratory for the day.[8] The experiment was conducted using the z-Tree (Fischbacher, 2007) software. We ran the experiments during school hours (courses in Hungary are 45 minutes long followed by a 15-minute break), so we had at most an hour to conduct the experiment with a given class and pay the participants. Participants knew each other as they were classmates in all sessions. After entering the classroom, participants were free to choose a seat. Once seated, the experimenter read aloud the instructions that students could also read from the sheet in front of them. Importantly, we explained in the instructions that participants would make decisions in 9 situations in 8 tasks,[9] many of them involving interaction with other participants, but we did not tell anything about the concrete experimental tasks. We emphasized that the experiment was not an exam, there were no correct answers, and that we were interested in how they would decide in a given situation. Participants were assured that all decisions remained confidential. In all sessions at least two experimenters were present who made sure that participants did not speak with each other or disturb each other in any way during the experiment. All doubts were answered aloud at the time when they were raised, so that each participant could hear it. We had run three pilot experiments, and we relied on the experience gained from them to eliminate the potential pitfalls. In some of the more complex questions (e.g. time preferences) we have included control questions to see whether students paid attention to the questions or understood them correctly. In other questions additional "practice" tables were provided before the real questions. We describe these processes and results in the research documentation (Horn et al., 2020).

There were no time limits in the different tasks (except for the real-effort task to measure competitiveness), the only constraint being that we had to end the experiment before the next class. We asked participants to occupy themselves silently after they have finished, because potentially there could be large differences in how much it would take for different participants to make all the decisions. Even though there was a large variance in the time that participants spent with the tasks, there were no incidents related to it.

Time and risk preferences were measured using individual tasks, so the payoffs did not depend on the choices of other participants. The measurement of social and competitive preferences involved strategic interaction, so payoffs were interdependent. To create random student pairs, we used z-Tree (Fischbacher, 2007). Matching pairs was carried out always at the end of the experiment, after each student made the decisions in each task.[10]

We incentivized the participants with meal vouchers that could be used in the school cafeterias as cash. We made clear to the students that from the 8 experimental tasks one would be randomly chosen by the computer for payment, and that the game for payment would be the same for all participants. We explained carefully that if a task involved several choices (as the time preference measures did), only one randomly picked choice would be payoff-relevant. We paid no show-up fee, as we went to the schools during school hours. Payoffs in the different tasks were designed so that the expected payoff was around 1000 HUF (around 3 EUR), approximately the price of a full meal at an average school cafeteria.

We informed participants about the details of the payment (e.g. random selection of tasks for payment, use of vouchers) at the beginning of each session. Payoffs not involving delay were handed out at the end of the session.

2.2. Experimental tasks

Time preferences Time preferences reveal how an individual trades off earlier and later benefits. Using the beta-delta model proposed by Phelps and Pollak (1968) and Laibson (1997) we can differentiate between patience (delta) and time consistency (beta). Patience indicates how an individual values the future relative to the present, while time consistency indicates if this relative valuation is the same at different points in time. Patient individuals value the future more relative to the present than their less patient counterparts. Time consistency implies the same trade-off between earlier and later benefits at different points in time when separated by the same time interval. In contrast to time consistent individuals, present-biased (future-biased) ones are more (less) impatient now than later.

To capture both aspects of time preferences, we measured decisions at two different time horizons. Participants had to choose between receiving a smaller amount today or a larger amount in 2 weeks (task 1) and they made the same decision also for the dates 4 weeks vs. 6 weeks (task 6). In both cases, participants made 5 interdependent choices using the staircase (or unfolding brackets) method (see Cornsweet, 1962; Falk et al., 2018). The benefit of this method is that it uses the available number of questions efficiently to find the approximate indifference point between the earlier and the later payoffs. In each case, the earlier amount was fixed (1000 HUF $\sim$ 3 EUR) while the later amount (X) was changed in an adaptive way, depending on the previous choices. For instance, a choice of 1000 HUF today instead of X=1540 HUF in 2 weeks indicated that the indifference point was higher than 1540 HUF (because +540 HUF was not enough for the student

[6] The NABC ID is a hash-code of the educational IDs of the students used only to identify students within the NABC surveys. It is not linked to any other data set. We notified the education providers that we would collect NABC IDs.

[7] There were two students who opted out from our experiment.

[8] In two schools we used the computers of the school to run the experiment.

[9] We divided the experiment to 8 tasks, as trust and trustworthiness were measured within one task.

[10] With an odd number of students in the room, the last pair of students was in fact a group of three participants and the payments of students in this group were affected by the decision of only one of the other students who was also randomly chosen by the program.

to wait two weeks), so in the next question X was increased. X ranged from 1030 to 2150 HUF. After five questions we have a fairly accurate information about the indifference point.[11] If the same participant in task 6 (4 weeks vs. 6 weeks) ends up with the same indifference point, then she is time consistent. A lower indifference point indicates present bias.

When one of the two time preference tasks was payoff-relevant, the computer chose randomly one of the 5 decisions and participants were paid according to their choice. That is, students who chose to receive a larger amount two, four or six weeks later were asked to put their vouchers in an envelope indicating the name of the student and the date when the payment was to be received, and we placed the envelopes at the school secretariat from where students could claim their payment in two, four or six weeks.[12]

Risk preferences Attitudes toward risk are informative about an individual's attitude toward uncertainty, so the corresponding tests generally involve some situation with uncertainty, mainly gambles (e.g. Eckel and Grossman, 2002; Gneezy and Potters, 1997; Holt and Laury, 2002). Based on our experience in a pilot experiment, gambles may seem strange to our student pool, so instead of gambles, we decided to use the bomb risk elicitation task by Crosetto and Filippin (2013), which is a more game-like measure of risk preferences, where higher values indicate higher risk tolerance.[13] Crosetto and Filippin (2016) examine four, widely used risk elicitation methods in experimental economics, including the bomb risk elicitation method, and report that it is a valid measure of risk preferences. We measured risk attitudes in task 4.

When this task was selected for payment, the computer generated a random number between 1 and 100 that determined the outcome of the risky situation and the earnings of the participants.

Social preferences There are many aspects of social preferences. In our experiment, we focused on four of them: *altruism, trust, trustworthiness and cooperation*.

Following standards of the profession, we measured altruism with the dictator game. There were two dictator games. In both tasks, participants were endowed with 2000 HUF. In the first one (task 2), the participants had to decide how to split their endowment with a classmate in the room, while in the second one (task 3), the other party was not somebody from the room, but a random schoolmate. Task 2 was incentivized, but task 3 was hypothetical as implementing the choice was not feasible. When this task was payoff-relevant, the computer paired the participants randomly and selected randomly a member of each pair to be the dictator and her / his choice was implemented.

We measured trust and trustworthiness using a modification of the trust game (also known as investment game) by Berg et al. (1995). The modification (also applied by Sutter and Kocher (2007)) consisted in that the receiver did not have an initial endowment. The game (task 7) consisted of two steps. In step 1, in the role of the sender, each participant decided how much of their endowment of 1000 HUF to send to a randomly selected receiver in the room, knowing that the amount would triple at the receiver, and in the second step, the receiver could send back any portion of that larger amount. The initially sent amount had to be a multiple of 100 and it is our measure of trust. In step 2, everybody assumed the role of the receiver and they had to choose how much they would return of the 3*X of all possible (and hypothetical) X amount received (X=0, 100, 200, ... 1000). That is, we have answers for all contingencies, and this stage provides information on the trustworthiness of the participants. More concretely, we calculate the share of the amount sent back for every possible amount received and tripled, and we use the average of theses shares as our measure of trustworthiness.[14] Everybody made a decision in both roles (as a sender and as a receiver). We modified the trust game to link it more to the dictator game where the recipient depends on the altruism of the dictator. Here, this motive is still present, but it is complemented with the possibility of reciprocity by the receiver. In the role of the receiver, the reciprocity motive may become stronger relative to the standard trust game as without the sender sending money, she would end up with nothing. Hence, the modification both intensified the senders' and the receivers' motives to be prosocial. When this game became payoff-relevant, students were paired, and one student in each pair was randomly selected as sender. We used the decision of the receiver that corresponded to the sent amount to determine the players' payoffs.

The third dimension of social preferences that we measured was cooperation. Using a two-person variant of the public goods game (task 5), we endowed everybody with 1000 HUF and matched each participant randomly with somebody else in the room. They had to decide how much of the endowment to contribute to a common account, without knowing the decision of the other participant. The amount not contributed to the common project added to their payoff. The marginal per capita return was 75%, so each of the two participants received 75% of the total contributions, independently of the individual contribution. Our proxy for cooperation is the contribution to the common project: the more a participant contributes,

[11] For example, if the participant in the last question chooses 1730 HUF in 2 weeks instead of 1000 HUF today, then (by the construction of the payoffs) we know that her indifference point is between 1730 HUF and the closest lower amount (1650 HUF). For simplicity, in this case, we assign the indifference point of 1650 to the participant, so she needs a 650 HUF compensation for waiting 2 weeks to receive the payment.

[12] We made sure to choose dates for the experiments so that these later payments could be received and the vouchers could be used without any problem, e.g. no later payment occurred during holidays. The Covid-19 outbreak and the sudden school closures have affected some of the later payments, so we agreed with the schools to distribute these later payments to the students when normal routine returns. Since the outbreak and the ensuing school closure was unexpected, the choices of the students should not have been influenced by these events.

[13] In this task, there is a store with 100 numbered boxes, one of which contains a bomb with uniform probability. Participants decide how many boxes to collect, following the numbering. If the bomb is in one of the boxes collected, then the participant earns no money, otherwise earnings increase with the number of boxes collected. The number of boxes collected is a proxy for risk tolerance.

[14] For instance, if X=300 and the receiver returns 450 HUF, then the share sent back is $\frac{450}{3*300} = 0.5$.

the more cooperative she is.[15] When this task was chosen for payment, the computer randomly paired the participants and based on their decisions the payoffs were calculated and paid.

Competitiveness We measured competitiveness in the last task (task 8), using the setup by Niederle and Vesterlund (2007), but instead of adding up numbers, participants faced a real-effort task where they had to count zeros in 5x5 matrices (as in Abeler et al. (2011)) for one minute. In the first stage (piece-rate) the number of correctly solved matrices determined the participants' earnings. In stage 2, the outcome of a tournament defined the payoffs, where only the best 25% of the participants earned money for the task, though in this case, earnings were 4 times as high per matrix solved as in stage 1. At the end of stage 1 and 2 we provided feedback about how many matrices the participants solved correctly, but no information was given about their relative performance. In stage 3, students could decide whether to get paid by the piece-rate or by the tournament scheme. The tournament choice is the indicator of a participant being competitive. After stage 3, participants were asked to rank themselves (being in the 1st / 2nd / 3rd / 4th quartile) based on their performance in stage 1 and 2. This belief elicitation was incentivized, those who guessed correctly received 300 HUF (but only if this task was picked for payment). At the end of the experiment, when this task was selected for payment, the computer picked one of the stages randomly and participants were paid according to their performance (and guesses) in that stage.

Order It was not obvious in which order to implement the 8 tasks. The following considerations governed our decision when establishing the order. Since participants might have unwittingly tried to be consistent in their choices in the two time preference tasks, we wanted to have them somewhat apart, introducing other tasks between them. In the two dictator games, the only difference was the reference group so we put these questions close to each other, since we did not think that participants would want to be consistent in giving the same amount to classmates and schoolmates. Our aim was that participants consider the different tasks as separate and independent decisions, so in the first 7 tasks, we did not give any feedback to them. Note that in the first 7 tasks there is no clear good choice. However, in the last one, that is, in the competitiveness task participants received feedback about their absolute performance (the number of matrices solved) and the (potential) earnings that those performances implied.[16] Knowing the absolute performance may affect the participants emotionally (e.g. having earned a lot of money in the piece-rate stage may cause elation), we put the competitiveness task at the end. All participants made decisions in the same order. An advantage of having a fixed order of tasks for all students is that students' decisions are directly comparable, while the drawback is that we do not know whether order effect influenced the choices (e.g., would students make the same decisions in the time preference tasks if those tasks were the last ones?). As a consequence, all our findings are conditional on the special order that the participants played the games.

2.3. Family background and cognitive abilities

Besides the preferences that we measured at schools, the other main variables of interest are related to demography (age, gender), the family background of the students, their school performance and their cognitive abilities. We obtained these student-level variables from the NABC database. Data on gender and age are missing only for a few cases, but socioeconomic status is missing in 16% of the cases and GPA in 24% of the cases, because these were self-reported in the NABC questionnaire. For family background, we transformed all categorical variables into dummy variables, where missing was a separate category. For the GPA, we imputed missing values with the sample mean and controlled for the imputed values with a separate missing dummy.

The family background variables that we consider are: the highest level of parents' education, father's employment status, whether the family receives regular child protection support and the number of books at home. We proxy cognitive ability with standardized mathematics and reading test scores measured in grade 6 (around age 12).

As a further set of controls for school performance, we use teacher-given class-marks from grade 6: GPA, as well as separate grades in mathematics, Hungarian language and literature.

2.4. Some descriptive statistics about the sample

In our sample of 1088 students, we have 611 females and 477 males in 53 classes in 9 schools. Students are 16.8 years old on average (min. 14.5 - max. 20.5). Table 1 shows the pairwise correlations of the preference measures as well as their significance level. Tables B.5 and B.6 in Appendix B show the average difference between males and females in all observed characteristics and in the preferences. Tables 2 and 3 below show the corresponding *t*-statistics.

Unsurprisingly, different preference measures within a preference domain are well correlated - i.e. altruism, trust, trustworthiness and cooperation within social preferences, as well as delta and beta within the time preference domain. Risk is correlated with most of the measured preferences, suggesting that it has a prime role among preferences. There are also

[15] To make the decision easier, on the decision screen, participants had two sliders, both of them going from 0 to 1000, the first representing their contribution and the second corresponding to their co-player's contribution. By moving the sliders, they could see the payoff consequences of different contribution combinations. Fig. 7 in Horn et al. (2020) contains a screenshot of the decision screen.

[16] We did not inform participants after stage 2 if they were in the best 25% of students. We only let them know the number of correctly solved matrices and the payoff if they happened to be in the best 25%, but we did not tell them if they were or were not.

Table 1
Pairwise correlations between preferences.

	Delta	Beta	Risk	Altruism	Trust	Trustworthiness	Cooperation
Beta	−0.394***	1					
Risk	0.156***	−0.115***	1				
Altruism	0.0278	0.0668*	0.130***	1			
Trust	0.145***	0.0520	0.220***	0.266***	1		
Trustworthiness	0.0213	0.0190	0.0157	0.294***	0.324***	1	
Cooperation	0.103***	0.00970	0.149***	0.138***	0.450***	0.252***	1
Competition	0.0184	0.0486	0.0813**	0.0280	0.0466	0.0297	0.0329

*/**/*** denotes significance at 10 / 5 / 1% level.

Table 2
Average difference of all variables between males and females - NABC data.

	T-test	
	Diff.	*t*-stat
NABC data		
Age (in months)	0.684	(0.83)
Family		
Parental ed.: low	0.00496	(0.62)
Parental ed.: medium	0.124***	(4.27)
Parental ed.: high	−0.108***	(−3.55)
Parental ed.: missing	−0.0216	(−1.17)
Father: employed	−0.0421	(−1.45)
Father: self-employed	0.0549*	(2.55)
Father: regural work	0.00405	(0.46)
Father: occasional work	−0.00741	(−1.01)
Father: childcare	−0.00368	(−0.50)
Father: retired	0.00890	(1.53)
Father: unemployed	0.00609	(1.35)
Father: disabled	0.00563	(1.08)
Father: missing	−0.0264	(−1.29)
Child support: no	−0.0109	(−0.42)
Child support: yes	0.0364	(1.81)
Child support: missing	−0.0255	(−1.34)
No. books: 0–50	0.00140	(0.09)
No. books: cca. 50	0.0337*	(2.01)
No. books: max. 150	0.0256	(1.11)
No. books: max 300	0.0246	(1.13)
No. books: 300–600	0.00239	(0.11)
No. books: 600–1000	−0.00712	(−0.32)
No. books: over 1000	−0.0599**	(−2.81)
No. books: missing	−0.0206	(−1.14)
Cognitive skills		
Math score, 6th grade	−0.521***	(−8.79)
Reading score, 6th grade	−0.120*	(−1.97)
Grades		
GPA, imputed	0.0361	(1.35)
GPA, missing	−0.00721	(−0.30)
Math grade, imputed	−0.105*	(−2.03)
Hungarian grade, imputed	0.0839	(1.92)
Literature grade, imputed	0.0768	(1.94)
Math grade, missing	−0.0305	(−1.42)
Hungarian grade, missing	−0.0251	(−1.17)
Literature grade, missing	−0.0298	(−1.37)

/***/**/* denote significance at 1 / 5 / 10%.

some less straightforward associations: delta is correlated with trust and cooperation.[17] Competition seems to be the most unique preference as it correlates only with risk.

According to Table 2, while in most cases there is no statistical difference between females and males, some variables are significantly different. Apparently, males in our sample have better family background, as their parents are relatively more educated and less likely to be self-employed. The number of books also indicates a higher socioeconomic status of males. Males have higher mathematics and reading test scores in grade 6 in our sample, indicating better cognitive abilities.

[17] This might be due to the fact that both the trust game and the public goods game have a slight time element in the sense that one has to wait until the other player decides, to know the outcome.

Table 3
Average difference of all variables between females and males (female-male) - Experimental tasks.

	T-test	
	Diff.	*t*-stat
Experiments		
Payoff	−23.01	(−0.45)
Delta	−0.0402***	(−4.27)
Beta	0.0532***	(3.85)
Risk	−6.466***	(−5.76)
Altruism (classmate)	5.932***	(5.34)
Altruism (schoolmate)	4.550***	(3.67)
Trust	−8.215***	(−5.36)
Trustworthiness	−3.058**	(−3.08)
Cooperation	−3.842*	(−2.28)
Competition	−0.101***	(−3.40)
Time		
Total time spent on tasks	74.19***	(6.04)
Time spent on Task 1 (time now vs 2 weeks)	6.438***	(5.50)
Time spent on Task 6 (time 4 vs 6 weeks)	5.311***	(3.61)
Time spent on Task 1 and 6 (both time tasks)	11.75***	(4.98)
Time spent on "Risk"	4.214**	(2.98)
Time spent on "Altruism (classmate)"	1.303	(1.10)
Time spent on "Altruism (schoolmate)"	1.908*	(2.41)
Time spent on "Trust"	8.173***	(3.96)
Time spent on "Trustworthiness"	33.25***	(7.27)
Time spent on "Cooperation"	9.328***	(4.01)
Time spent on "Competition"	9.371***	(4.14)

/***/**/* denote significance at 1 / 5 / 10%.

These differences are likely to be interrelated as better test scores may be due to better family background. If we look at within class differences in these variables, only very few of them remain significant (e.g. females' parents are more likely to be medium level educated, but not less or more educated), and few reverse their sign (e.g. females have higher GPA within class due to their higher Hungarian language and literature grades, but males have significantly higher math test scores).[18] This suggests that while our sample of classes are far from being representative, the within class gender differences resemble that of the total population better.[19]

Table 3 also reveals that females spend significantly more time on almost all experimental tasks and overall on the experiment (expressed in seconds). In the regression, we control for time spent on the specific tasks and the total time spent on the experiment.

3. Results

Our main variable of interest is the gender dummy (*female*) that indicates if females make different decisions in the given preference task. For each preference measure, the first specification is the raw difference between the genders: the female coefficient without any control variables. But as we have shown above, our sample is quite imbalanced if we do not control for the fact that our respondents are clustered within classes. Hence, the second specification adds the class fixed effects (*class FE*). Their inclusion allows us to take into account the following: i) all experiments were conducted within a classroom at a given time and place under similar circumstances; ii) participants play some of the games with their peers in the classroom; iii) students are likely to be selected into different classes (and hence our imbalance in the covariates). In fact, we believe that the results of this specification would probably be closer to a representative sample, had we have one. Henceforth, when including additional controls to our models, we will use the class fixed effect model as a reference. In the next specification, we control for age as it has been shown to be an important determinant of preferences during adolescence (see Sutter et al., 2019). Then, we control for family background by considering various aspects of the socioeconomic status (*family*), see Section 2.3. In the next specification, we include the mathematics and reading test scores from grade 6, assuming that they are good proxies of *cognitive skills*. Then, we also add *grades* to control for school performance. Besides the grade point average, we also take mathematics, Hungarian language and literature into account. Next we control for the time spent on the given task as well as the total time spent on the experiment. In the last specification we control for all

[18] See figures A.26-A.30 in Appendix A for more details.

[19] In the 2017 NABC 6th grade full database females have 0.04 standard deviation lower maths scores and 0.27 standard deviation higher reading scores than males. Females also have a 0.24 points (out of 5) higher GPA, 0.4 points higher Hungarian language and 0.33 higher literature grades than boys, while boys score 0.15 points higher in maths than females.

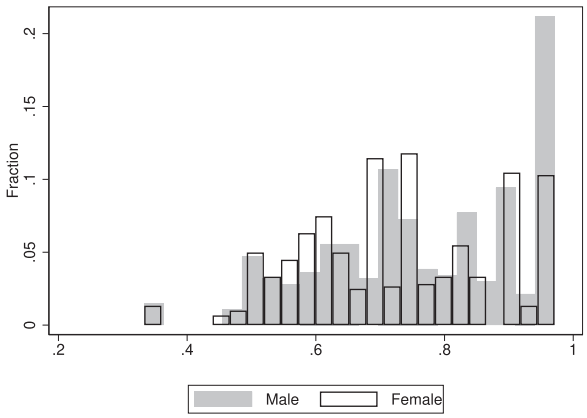

Fig. 7. Distribution of patience (delta) by gender.

other preferences, thereby testing if the association of gender with the preference of interest is confounded by the other preferences.[20]

By controlling for exogenous factors like age, family characteristics, cognitive skills and school performance, we do not only control for the biases in our sample but also try to speculate about the mechanisms through which gender associates with preference measures. Taking into account the time spent on the tasks, we expect to capture the effort that the participant made in the given task (or how seriously she or he took the task). Finally, by controlling for all other measured preferences, we test whether the given preferences have a direct (*ceteris paribus*) effect on the differences between genders.

To ease the exposition of the results, we use coefficient plots that visualize the estimation of the coefficient of the *female* dummy with the corresponding 95% confidence intervals. Appendix C contains the full regression tables. To present our findings in a structured way, we use the same descriptive statistics and the same specifications in all of the regressions for the analysis of the different preferences below.

3.1. Time preferences

The existing literature did not produce a clear-cut finding if there is a gender difference in the patience of adolescents. Some studies report females being more patient (Bettinger and Slonim, 2007; Castillo et al., 2011; 2019), while others document the opposite result (Golsteyn et al., 2014). Focusing on the studies that are directly comparable with our work (see Fig. 1 and Appendix D.1), we also see that Castillo et al. (2011) report females being significantly less patient than males, whereas Sutter et al. (2013) find no significant gender difference. Some individual factors are argued to have a direct effect on time preferences. Patience is shown to increase with age (Bettinger and Slonim, 2007), while low social status is likely to predict more impatient choices (Castillo et al., 2011). There is also some evidence that better cognitive abilities associate with more patience (Luehrmann et al., 2018). Regarding present bias, Tymula (2019) and Luehrmann et al. (2018) do not find gender differences.

Our measure of patience (delta) is the individual discount factor that we calculate based on task 6, the intertemporal choice between a lower amount of money in 4 weeks and a larger amount of money in 6 weeks. Assuming linear utility, the indifference amount of 1000 HUF to be received in 6 weeks (denoted by x_6) comes from the equation $1000 = delta * x_6$, were *delta* denotes patience. In our sample, delta ranges from 0.33 to 0.97. The mean for females is 0.73 and for males is 0.77.

Figure 7 shows the distribution of delta by gender and reveals that the difference between females and males is mainly due to the fact that there are more males at the upper end of the distribution. More precisely, more males exhibit the maximum level of patience than females.[21]

Figure 8 represents the coefficient plot of the regression analysis (see Table C.8 in the Appendix for the full regression). The first point in Fig. 8 shows the difference in the raw data, confirming that there is a significant gender difference in patience. However, once we add class fixed effects, the gender difference disappears and remains so in the rest of the specifications. Therefore, if the individual characteristics of the participants and features of the environment are controlled for, there seems to be no gender difference in patience.

[20] We only control for the "main" preferences from the four domains - delta, risk, altruism, trust, cooperation and competition - and we never control for preferences from the same domain. So we do not take delta into account when we look at beta, nor do we control for the other social preferences, when we look at altruism, trust or cooperation.

[21] The Wilcoxon rank-sum test (p-value$<$0.001) indicates that overall, males are more patient than females. The Kolmogorov-Smirnov test shows that the two distributions are not equal (p-value$<$0.001).

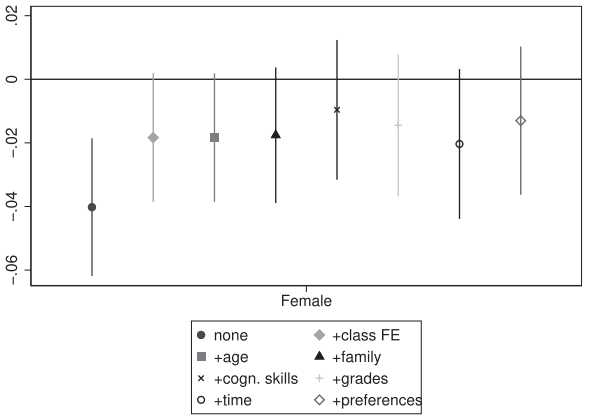

Fig. 8. Adjusted gender differences in patience (delta).

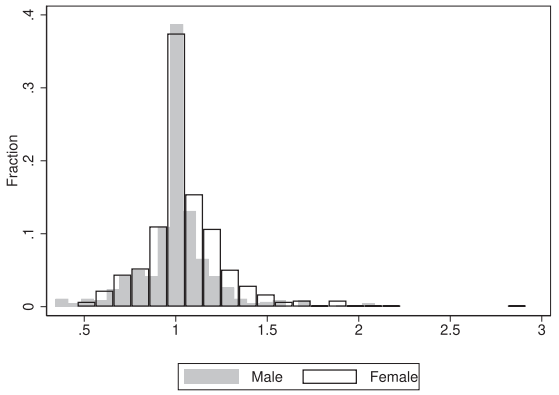

Fig. 9. Distribution of time inconsistency (beta) by gender.

We compute the time inconsistency parameter (beta) by applying the delta - beta model (Phelps and Pollak, 1968; Laibson, 1997) to the observations from experimental tasks 1 and 6. Using the previously computed delta parameter and denoting the indifference amount of today's 1000 HUF to be received in 2 weeks by x_2, beta is given by $1000 = beta * delta * x_2$. In our sample, *beta* ranges from 0.34 to 2.91. The mean for females is 1.05, while for males it is 1.[22] These values near 1 indicate that on average, females and males are quite time consistent. Even though the distributions of beta illustrated in Fig. 9 seem to be very similar for females and males, the Kolmogorov-Smirnov test rejects the equality of distributions (p-value$<$0.01). The Wilcoxon rank-sum test indicates that the beta of females is different from the beta of males (p-value$<$0.001). In order to be able to test time inconsistency in a linear regression framework we have transformed the beta parameter: the greater the distance from the value 1 the greater the time inconsistency. Hence, we subtracted 1 from beta and took its absolute value, before running the regressions below. Thereby, the female coefficient can straightforwardly be interpreted as gender differences in time inconsistency.

Figure 10 (Table C.9) shows that there is no significant gender difference in time inconsistency. The initial raw significant difference disappears after taking into account class fixed effects and the lack of gender gap remains even after we control for all observable characteristics and the rest of the preferences.

Time inconsistency comprises any deviation from time consistency: individuals being more impatient now than in the future (known as present bias) or the other way around (known as future bias). However, more attention has been given to present bias as it relates to procrastinating behavior and suboptimal life outcomes (Ariely and Wertenbroch, 2002; Moffitt et al., 2011; Daly et al., 2015; Wang and Sloan, 2018). To study present bias, we restrict our attention to beta$<$1, and generate a dummy variable where present bias=1 if beta$<$1 and present bias=0 if beta$\geq$1. 32.8% of the students, 29.8% of the females and 36.7% of the males are present biased. The test of proportions reveals that there is a significant difference in the proportion of present-biased students between females and males (two-tailed test, p-value=0.0179), indicating that males are more present-biased than females. Figure 11 shows that this difference is persistent and though it diminishes

[22] Luehrmann et al. (2018) report similar range of values for time inconsistency.

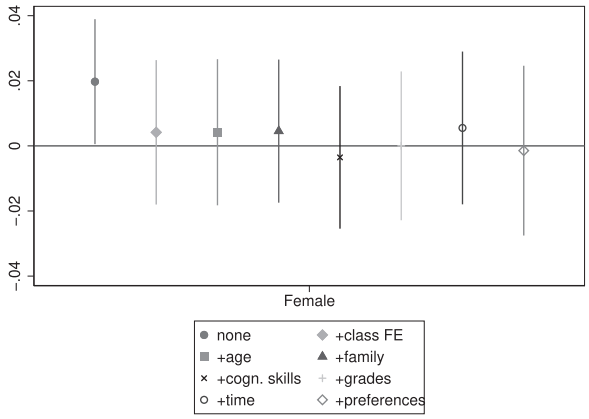

Fig. 10. Adjusted gender differences in time inconsistency ($|beta - 1|$).

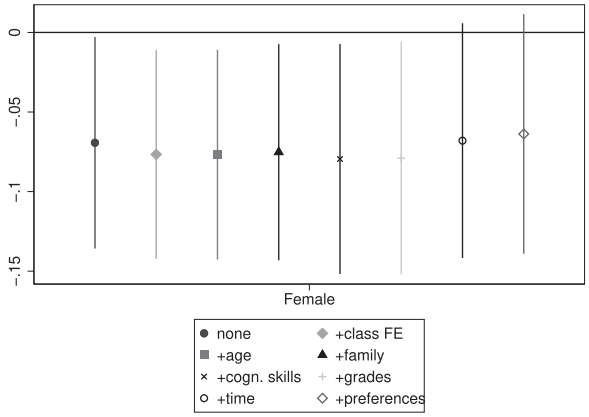

Fig. 11. Adjusted gender differences in present bias (beta<1).

somewhat when adding the observable characteristics, its size remains stable and remains statistically significant albeit only on the 10% level (see Table C.10 in the Appendix).[23]

When we compare the female coefficients across models, it is apparent, that the included exogenous covariates (age, SES, cognitive skills and grades) do not have a significant effect on its size, that is, no observable individual characteristic affects the gender gap in different dimensions of time preferences (see Table B.7 in the Appendix that shows the significance of the Chi-squared tests, a direct comparison of the female coefficients across models).

3.2. Risk preferences

Studies about risk preferences during adolescence mostly find that females are more risk-averse than males (Borghans et al., 2009; Booth and Nolen, 2012b; Eckel et al., 2012; Sutter et al., 2013). There seems to be an age trend, older children are less risk-taking (Harbaugh et al., 2002). Moreover, Khachatryan et al. (2015) find that the gender gap in risk-taking becomes larger in adolescence. Socioeconomic status also seems to matter, as low status associates with more risk-taking, though the evidence here comes mainly from the childhood (Deckers et al., 2015; 2017; Alan et al., 2017).

Regarding risk attitudes, there is mounting evidence that the type of risk elicitation task matters, as some tasks are more likely to reveal gender differences than others (see Filippin and Crosetto, 2016; Niederle, 2016). Crosetto and Filippin (2013, 2016) show that there is no general gender difference in risk-taking in the bomb risk elicitation task that we use, although the participants were older in both studies than the students in our sample. If we have a look at the studies that use the bomb risk elicitation task with children and adolescents (Andreoni et al., 2020; Piovesan and Willadsen, 2021), then we observe that females tend to be more risk averse even when this task is applied (see Fig. 2 and Appendix D.2).

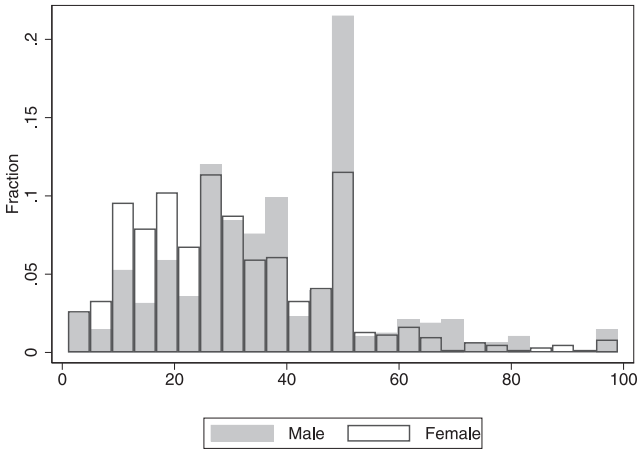

Fig. 12. Distribution of risk preferences by gender.

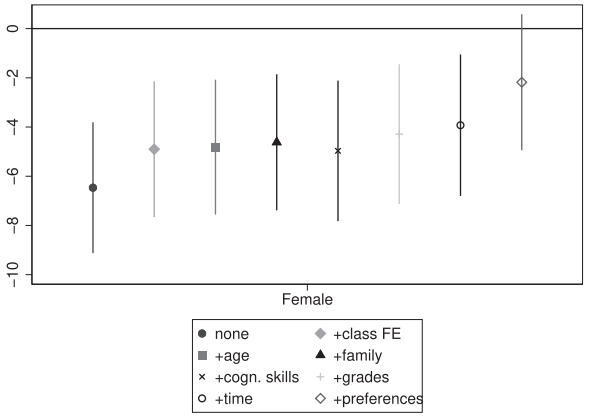

Fig. 13. Adjusted gender differences in risk preferences.

The distributions in Fig. 12 indicate that females tended to take out fewer boxes in the bomb risk elicitation task, that is, they are less risk-taking. There is also a marked difference in the choice of 50 boxes that seems to be a focal point. Males chose this number more often than females.[24]

While males were willing to take an average of 37.7 boxes out of the store, females chose to take out only 31.4.[25] Figure 13 indicates that this difference is statistically significant at 5% and persists even if we take into account class fixed effects, age, variables related to the family background and proxies for cognitive abilities, school performance and time spent on the tasks. However, when we control for the other preferences the difference becomes insignificant (though the sign of the coefficient does not change). When only the preference measures are considered, all preferences except trustworthiness associate with risk at a significance level of at least 5%, see Table 1. Moreover, many preferences (delta, altruism, trust and competition) associate with risk significantly, *ceteris paribus*, even if we control for all other preference measures, which suggests that risk is a preference present in many other domains (see Table C.11 in the Appendix). This result might also be interpreted as the result of risk preferences mirroring the gender effects of the other preferences, or conversely the gender difference in risk preferences drives some of the gender effect in the other preferences. While using these data we cannot tell which of these directions is more pronounced, it is important to underline that risk preference is inherent in many of the other preference domains and that the gender gap in risk also associates with gender differences within the other preferences.

[23] Note that in Fig. 11 the confidence intervals in the last specifications cross the zero line because they pertain to 95% confidence intervals.

[24] The Kolmogorov-Smirnov test reveals that the distributions are not equal ($p < 0.001$), and the Wilcoxon rank-sum test indicates that males are more risk-taking ($p < 0.001$).

[25] With non-children sample, usually both females and males are willing to take more risk in this task than our sample of students. For instance in Crosetto and Filippin (2013) females / males take out 43.4 / 44.2 boxes. However, Piovesan and Willadsen (2021) report numbers close to our observations: 33.6 / 37.2 for females / males.

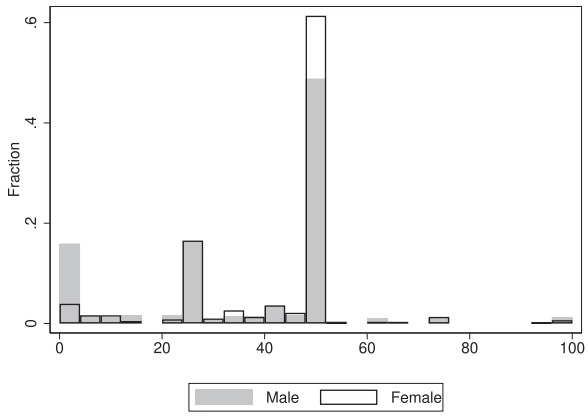

Fig. 14. Distribution of altruism (proxied by giving in the dictator game played with a classmate) by gender.

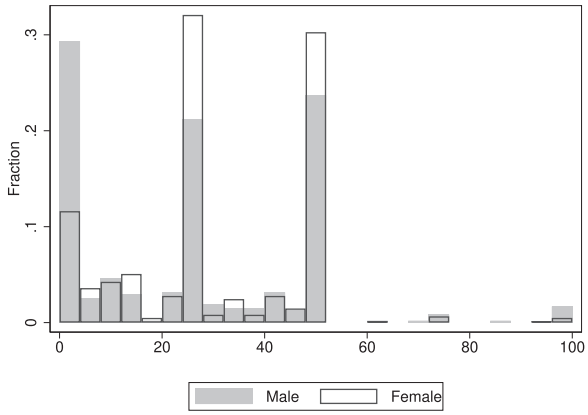

Fig. 15. Distribution of altruism (proxied by giving in the dictator game played with a schoolmate) by gender.

3.3. Social preferences

3.3.1. Altruism

During adolescence, females tend to be more altruistic (Harbaugh et al., 2003a; Bettinger and Slonim, 2006) and altruism increases with age (Harbaugh et al., 2003a; Bettinger and Slonim, 2006; Fehr et al., 2013).[26] The association of socioeconomic status with altruism is less clear. For younger children, low status correlates with giving less in the dictator game (Bauer et al., 2014; Deckers et al., 2017; Kosse et al., 2020). For adolescents, on the other hand, the only evidence (Almås et al., 2017) that we are aware of shows that low-status individuals are more egalitarian than individuals from a different background.

Following the literature, we proxy altruism with the amount given in the dictator game. While females in our sample gave 41.7% of their endowment to their classmates, males gave only 35.8%. Females being more altruistic than males is often found in the literature (see Fig. 3 and Appendix D.3 for a direct comparison with the literature that uses the same elicitation task).

Figure 14 indicates that females chose the egalitarian split more often than males (in line with findings by Fehr et al. (2013)), while the latter are more likely to give zero.[27]

We observe similar patterns when we consider how much the students give to a random schoolmate, but understandably the amount given decreases substantially. Figure 15 shows that females chose the egalitarian split more often and gave zero less frequently than males. Moreover, giving 25% of the endowment seems to be the focal point, and the share of females giving this percentage is higher than that of males. As to classmates, more males gave zero to a random schoolmate than females. Overall, both females and males gave less to a schoolmate than to a classmate (29.6% and 25.1% of their

[26] The effect of age is more complex as the change in altruism is intertwined with the application of meritocratic principles (Almås et al., 2010) and an increasing concern for efficiency (Maggian and Villeval, 2016; Sutter et al., 2018).

[27] The Wilcoxon rank-sum test and the Kolmogorov-Smirnov test indicate that the differences in the median and the distributions are significant (p-values <0.001 in both cases).

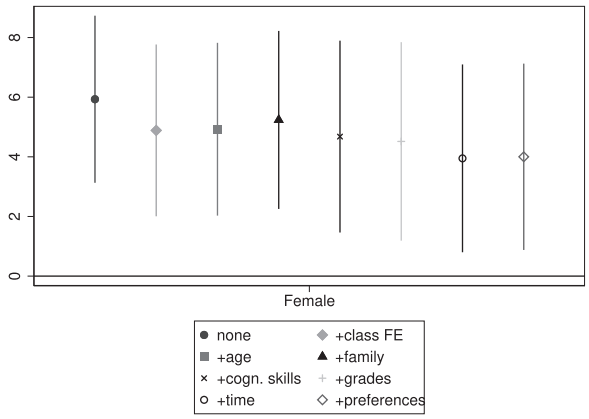

Fig. 16. Adjusted gender differences in altruism (proxied by giving in the dictator game played with a classmate).

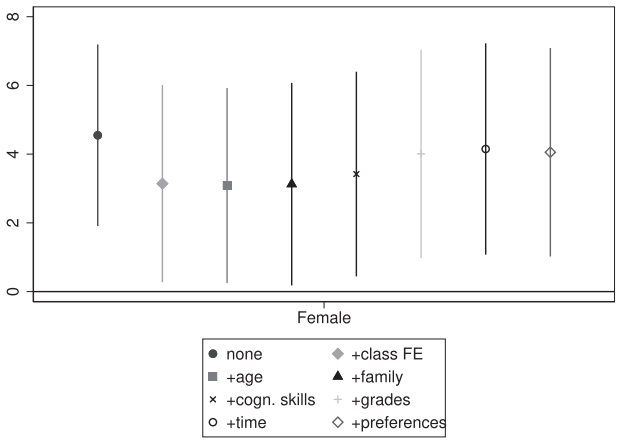

Fig. 17. Adjusted gender differences in altruism (proxied by giving in the dictator game played with a schoolmate).

endowment) in our sample, but the significant gender difference remains.[28] This suggests that the gender difference in altruism does not depend on the social distance between the dictator and the recipient.

Figures 16 and 17 (Tables C.12 and C.13) indicate that in line with the literature, females are significantly more altruistic than males, even if we add all the controls that we can observe. Hence, the significant difference is not due to differences in socioeconomic status, cognitive abilities, school grades or correlation with other preferences. There is no significant difference between the female coefficients across models (see Table B.7 in the Appendix).

3.3.2. *Trust and trustworthiness*

Gender differences in the trust game have not been in the focus of the previous literature. The amount sent by the trustor and the amount returned by the trustee tend to increase with age (Harbaugh et al., 2003a; Sutter and Kocher, 2007).[29]

Similarly to Sutter and Kocher (2007), we let our participants play a modified version of the trust game as the receiver had no initial endowment. The modification of the game implies that the receiver ends up without money if the sender does not send her / him anything. Thus, altruistic motives behind the sending behavior of the sender are stronger than in the standard game. This small modification allows us to directly compare both stages of the trust game with the dictator game as both differ from the dictator game in one aspect only. The sending stage of the trust game differs from the dictator game in that the sender can expect some reciprocity, while the altruistic motives behind the decisions are similar (and certainly stronger than in the standard trust game). The return stage of the trust game differs from the dictator game only in that the trustee received the amount from the sender and not from the experimenter. Overall, the modification increases the altruistic motives compared to the standard trust game.

In this light, it seems important that we find that males sent more of their endowment than females in both stages of the trust game, while females sent more in the dictator game (41.7% vs. 35.8% of their endowments, as shown above). In the

[28] Again, both the Wilcoxon rank-sum test and the Kolmogorov-Smirnov test indicate that the differences are significant (p-values <0.001 in both cases).

[29] See Fig. 4 and Appendix D.4 for a direct comparison with the only comparable study that we are aware of.

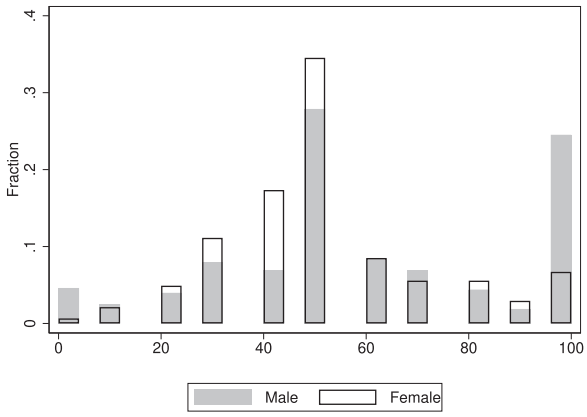

Fig. 18. Distribution of trust by gender.

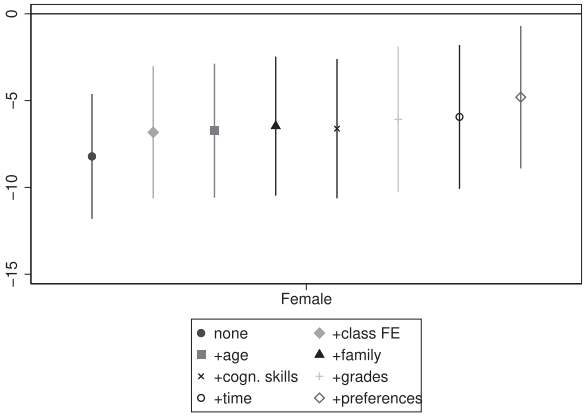

Fig. 19. Adjusted gender differences in trust.

first stage of the trust game, males sent 59.8% and females sent 51.6%. In the second stage, males - again - sent more than females (40.3% vs. 37.3%).[30]

Figure 18 indicates that the gender difference in the first stage of the trust game is due to the fact that almost 25% of the males sent their entire endowment to the receiver, while only less than 7% of females did so.[31] An explanation may be the difference in the weight that females and males assign to equality and efficiency (that is, making the overall pie bigger). Almås et al. (2010) and Maggian and Villeval (2016) point out that efficiency seeking becomes an important motive in adolescence (while egalitarianism matters less), and efficiency concerns are stronger in the case of males (Sutter et al., 2018).

Figure 19 (Table C.14) shows that as we add controls, the gender difference in trust shrinks, but it does not disappear. Males still send more of their endowment in the first stage of the trust game, even after all their observable characteristics - including their time, risk and competitive preferences - are controlled for.

Turning to trustworthiness (the second, return stage of the trust game), Fig. 20 indicates that males' decisions are more extreme: they are more likely to send nothing back, but they are also more likely to send half of the received (and tripled) amount back, or even above that.[32] Figure 21 (Table C.15) shows that the gender difference remains significant at 5% even if we add all the controls we have, including preferences.

Overall, we see that females tend to be more altruistic when they cannot expect anything in exchange and when the endowment is independent of the co-player. If any of these changes, males tend to send more. Figure 22 shows the distri-

[30] Remember that, in the experiment, when playing the role of the receiver, students made a decision on how much to send back to the sender for each possible amount that they could receive. That is, we asked for ten separate decisions. We asked that if s/he received 100/200/... /900/1000 HUF, how much s/he would send back from the tripled 300/600/... /2700/3000 HUF. We calculated the corresponding shares for each decision and computed the average. This is our proxy for trustworthiness.

[31] Both the Wilcoxon rank-sum test and the Kolmogorov-Smirnov tests indicate that males and females behave differently in the first stage of the trust game (p-value$<$0.0001 in both cases)

[32] The differences in the medians and the distributions are significant according to the Wilcoxon rank-sum test and the Kolmogorov-Smirnov test (p-values$<$0.0005 in both cases).

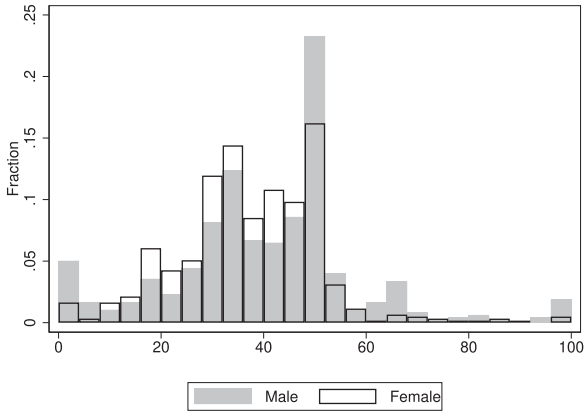

Fig. 20. Distribution of trustworthiness by gender.

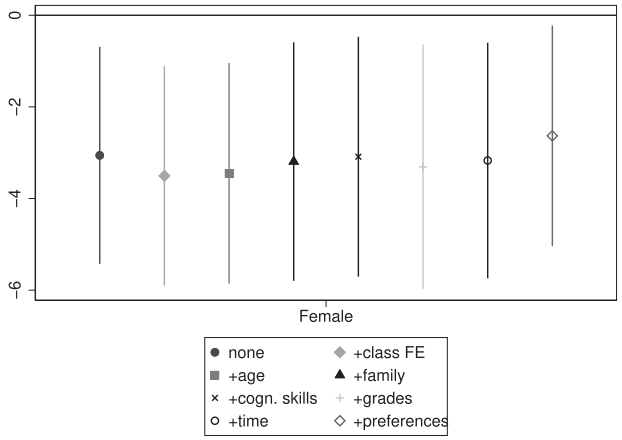

Fig. 21. Adjusted gender differences in trustworthiness.

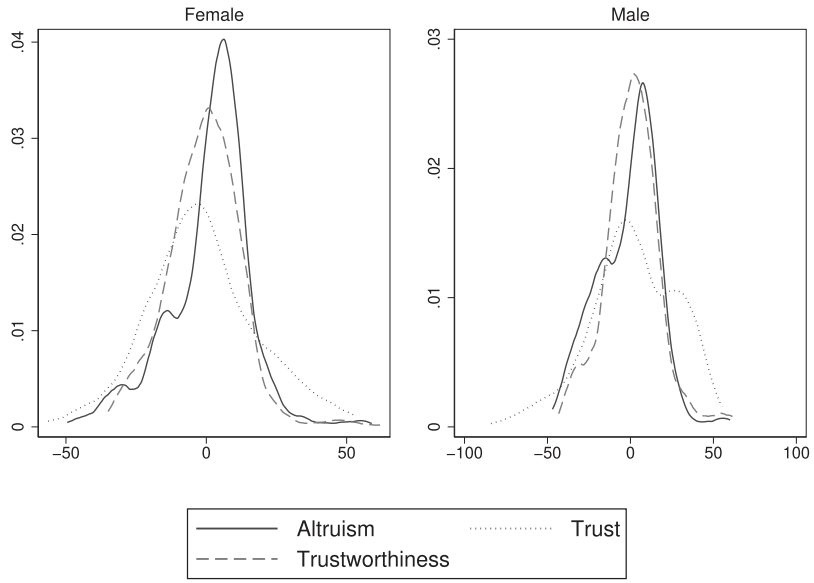

Fig. 22. Distribution of residuals in the dictator and the trust game by gender.

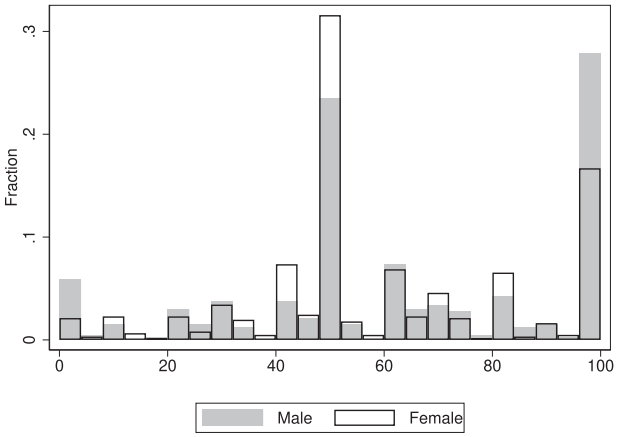

Fig. 23. Distribution of cooperation by gender.

bution of residuals from the three separate models of altruism, trust and trustworthiness, where we have controlled for all observable characteristics (including preferences) except the gender of the students. The more the distribution of trust and trustworthiness residuals differ from the dictator game, the more responsive students are to the changes between games. For females both the trust and the trustworthiness residuals are shifted to the left, indicating a general tendency to lower the amounts if parameters of the models change. Males, on the other hand, seem to react much less to the changes between the dictator game and the second stage of the trust game (albeit there is a slight bump at the lower end of the dictator game distribution), but much more to the changes between the dictator game and the trust game. That is, males react more to changes concerning reciprocity but less to changes in the source of the endowment, while females respond strongly to both.

3.3.3. Cooperation

The scant literature on cooperative behavior during adolescence does not report gender differences (see Fig. 5 and Appendix D.5 for a direct comparison with the literature that uses the same elicitation task).[33] There is some evidence that cooperativeness increases with age (Brocas et al., 2017), though the evidence is stronger in younger ages (Fan, 2000; Harbaugh and Krause, 2000; Angerer et al., 2016). The existing literature is silent on whether socioeconomic status or cognitive abilities associate with cooperative attitudes.

As explained in Section 2.2, we used a two-person variant of the public goods game in which the marginal per capita return on the offered amount was 75%. We proxy cooperativeness with the amount of contribution to the common project. While males contribute 62.6% of their endowment, females contribute 58.8%. Figure 23 reveals that this difference is mainly due to males contributing their whole endowment more frequently than females, while females choose to contribute half of the endowment more frequently than males.[34] Similarly to the behavior in trust games, the stronger efficiency-seeking motive present in males may explain why they contribute more to the public good. The OLS analysis reveals that once we take class fixed effects into account the gender difference vanishes and this finding does not change as we add more and more controls.

However, when comparing the female coefficients directly across models, it becomes apparent that the controls matter more than in the other preferences. The female coefficients in the class FE model (second specification) and in the third to last model (where all exogenous variables are controlled) differ significantly at the 5% level.[35] This difference becomes even more significant when we control for the time spent on tasks and other preferences. Thus, it seems that – unlike in any of the other preferences – exogenous controls do have a small but significant effect on the gender gap in cooperation (see Table B.7 in the Appendix).

3.4. Competitive preferences

There is ample evidence in the literature that females are less willing to enter competition than males during adolescence (Booth and Nolen, 2012a; Buser et al., 2014; Dreber et al., 2014; Sutter and Glätzle-Rützler, 2015; Almås et al., 2016; Sutter

[33] For younger children, Angerer et al. (2016) report that females cooperate more.

[34] The Wilcoxon rank-sum test indicates that the contribution levels differ across genders (p-value=0.0059), while the Kolmogorov-Smirnov test rejects the null hypothesis the distributions of contributions are equal (p-value=0.001).

[35] Males contribute around 1% more when controlling for class FE, while females contribute more by around 0.5% when all exogenous variables are controlled.

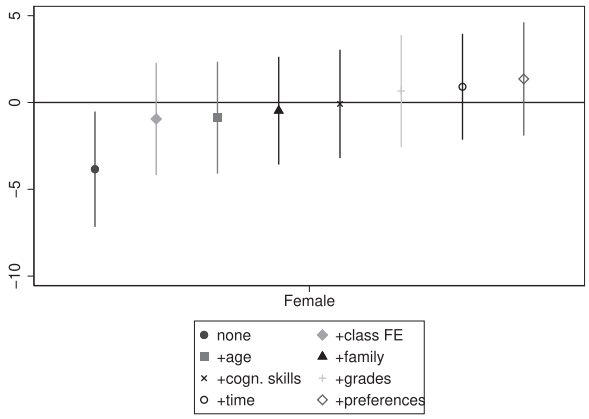

Fig. 24. Adjusted gender differences in cooperation.

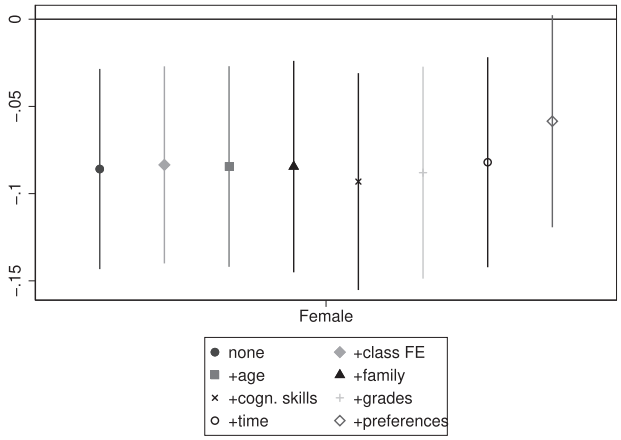

Fig. 25. Adjusted gender differences in attitudes toward competition.

et al., 2016).[36] Moreover, the environment may also shape competitive preferences, as Booth and Nolen (2012a) show that females in single-sex schools are more competitive. There is also some evidence that low-status adolescents (especially males) are less likely to compete (Almås et al., 2016, see Fig. 6 and Appendix D.6 for a direct comparison with the literature.

We use the experimental task developed by Niederle and Vesterlund (2007) to measure competitiveness, and we classify a student competitive if she chooses the tournament-based compensation in stage 3 of the competition task. Our data reveal that males are approximately 10 percentage points more likely to enter the tournament than females (66.2% vs 56.1%) if we do not take any of their observable characteristics into account.[37]

Figure 25 (Table C.17) indicates that the gender difference of 10 percentage points remains relatively stable and significant, even if we add more and more controls. Controlling for factors related to socioeconomic status, cognitive abilities, school performance, or the time spent on this task and overall on the experiment do not change the findings. Importantly, in the last step, we add the other preference measures to the regression, and the female dummy becomes only marginally significant (at 10%). As expected, more risk-taking students are more likely to enter the tournament, but even if we take this into account, the gender difference in competitiveness persists. There are no differences in the size of the female coefficient across models (see Table B.7 in the Appendix).

4. Discussion and conclusion

We carried out a large-scale experiment with Hungarian high-school students in their classrooms to measure a wide array of economic preferences that allows us to investigate gender differences in preferences during adolescence. Table 4

[36] In Fig. 6 and Appendix D.6 we compare our study directly with the literature that uses the same elicitation task. There are only some papers that do not report a gender difference, for instance, Khachatryan et al. (2015) or Zhang (2011) do not find a gender gap in Armenia and China, respectively. In fact, culture may affect gender difference in competitiveness as Andersen et al. (2013) report that females from patriarchal societies are less competitive.

[37] The test of proportions rejects the null hypothesis that the two proportions are equal (p-value=0.0007).

Table 4
Summary of the results.

	None	+Class FE	+Age	+SES	+Cognitive	+Grades	+Time	+Preferences
Patience (Delta)	M***	M*	M*	∅	∅	∅	M*	∅
Time inconsistency (Beta)	F**	∅	∅	∅	∅	∅	∅	∅
Present bias (Beta<1)	M**	M**	M**	M**	M**	M**	M*	M*
Risk tolerance	M***	M***	M***	M***	M***	M***	M***	∅
Altruism (classmate)	F***	F***	F***	F***	F***	F***	F**	F**
Altruism (schoolmate)	F***	F**	F**	F**	F**	F**	F***	F***
Trust	M***	M***	M***	M***	M***	M***	M***	M**
Trustworthiness	M**	M***	M***	M**	M**	M**	M**	M**
Cooperation	M**	∅	∅	∅	∅	∅	∅	∅
Competition	M***	M***	M***	M***	M***	M***	M***	M*

F / M represents females / males. */**/*** denotes significance at 10 / 5 / 1% level.

summarizes the main findings. The letters (F / M) indicate if the female dummy in the regressions is significant and which gender has a significantly higher measure in the given preference. Asterisks show the level of significance.

In order to take into account the factors related to the time and place of the experiments, selection into classes and the peer effects, we control for class fixed effects. Moreover, we consider many factors that have been proven to be important determinants of many preferences in adolescence according to the literature (Sutter et al., 2019). Thus, we control for age, family background (SES), cognitive skills and school performance (grades). By adding these controls, we make our non-representative sample reflect the total population of Hungarian adolescents better, and we also take the effect of potential individual confounders into account. Hence, if we observe gender differences even after considering these controls, it is strong evidence that those differences are real.

We observe that once we control for class fixed effects and individual exogenous factors (age, SES, cognitive skills and grades), there is no gender difference in patience (delta) and time inconsistency (beta), but males are more present-biased (beta<1) than females. There is a gender gap in risk attitudes, males being more risk-tolerant, that only ceases to be significant when we control for the other preferences. We see strong gender differences in two aspects of social preferences: while females are more altruistic than males (both with classmates and schoolmates), the opposite occurs regarding trust and trustworthiness. We detect no gender difference in cooperation. We also find that males are more competitive than females.

One might argue that as females tend to spend more time on (put more effort into) the tasks we see a biased gender gap. Also as preferences are correlated, focusing only on a single preference when investigating gender differences and not considering other preferences may be conducive to misleading conclusions. If preferences are not perpendicular they might be capturing the same non-cognitive traits. Thus, in our last two specifications we have controlled for the time spent on tasks (+Time column in Table 4), and for the other preferences (+Preferences in Table 4). Bear in mind that both time spent on tasks and other preferences might be endogenous factors in our gender gap estimation, as they might reflect gender differences that are affected by the preferences under scrutiny. (e.g. more patient individuals might spend more time on the tasks, or more patient individuals be more risk tolerant at the same time, while the direction of causality is not clear).

We find that taking the time spent on tasks into account has only a moderate effect on gender gaps in patience, present bias and altruism. Relative to the specification that includes grades, considering both time spent on tasks and other preferences affects present bias, risk, altruism, trust and competitive preferences between genders, the largest changes occurring in risk and competitive preferences (see Tables 4 and B.7).

By taking all exogenous factors as well as time spent on tasks and other preferences into account - hence risking over-control of the gender gap - we document (at least marginally) significant differences in present-bias, altruism (both towards classmates and schoolmates), trust, trustworthiness and competition.

By testing for different dimensions of time and social preferences separately we might have, by chance, found significant gender gaps. Admittedly, patience, time inconsistency and present bias are different aspects of time preferences; just as altruism, trust, trustworthiness and cooperation are aspects of social preferences. By running multiple but separate estimations we might find "false positives" - that is significant gender gaps - that are only due to chance. We correct for this, with a Familywise Error Rate Control correction, using our full specification with time spent on tasks and preferences as controls (see Table C.18 in Appendix C, where we show the estimated standardized coefficients, the original p-values as well as the Westfall and Young (Westfall et al., 1993) and the Bonferroni (Dunn, 1961) corrected p-values.). By increasing the standard errors of our most conservative models our results weaken, but still remain significant in all but one social preference domains, but the significant gender gap in present bias disappears.

In short, by controlling for a large variety of exogenous factors, time spent on tasks, other preferences as well as for multiple hypothesis testing, we still find significant gender gap in altruism, trust, trustworthiness and competition, where females are more altruistic but are less trusting, less trustworthy and less competitive.

There are still many open questions. It is natural to ask what are the mechanisms behind the gender differences that we document, especially in the domain of social preferences. There is a growing literature that attempts to unearth these mechanisms. For example, Heilman and Chen (2005) show how gender role prescriptions affect altruism. Since we opted

for eliciting several preferences without varying the measurement technique, we cannot say anything about how different forms of elicitation or feedback would affect gender differences. However, our findings suggest three things. First, it is not simply differences in the socio-demographics, cognitive skills, or school performance that lie behind the documented gender differences, as we applied a wide array of such controls. Second, while there are significant gender differences in time spent on tasks, even by controlling for these we can explain away only a small fraction of the gender differences. Finally, correlations with other preferences cannot explain all the reported gender differences either, suggesting that these four sets of preferences have perpendicular aspects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Within class gender differences

Appendix A contains Figs. A.26, A.27,A.28, A.29, A.30.

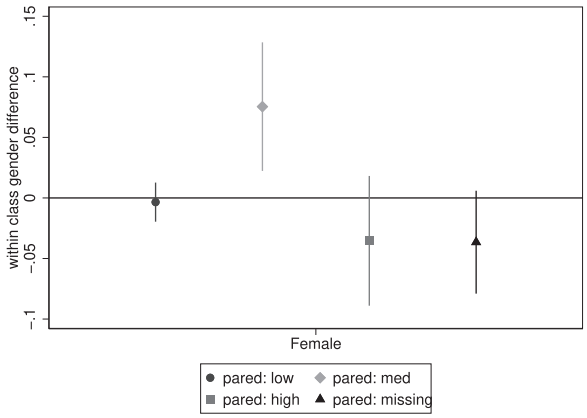

Fig. A.26. Within class gender differences, parental education.

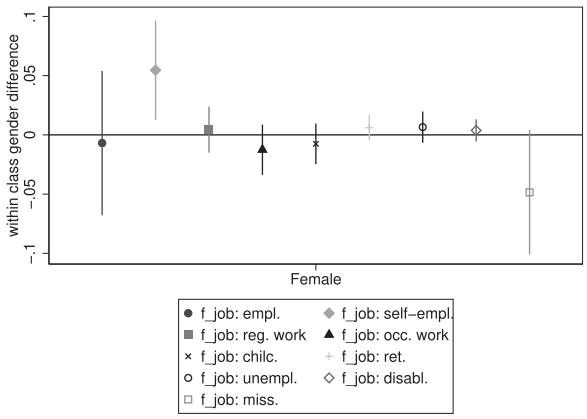

Fig. A.27. Within class gender differences, father's job.

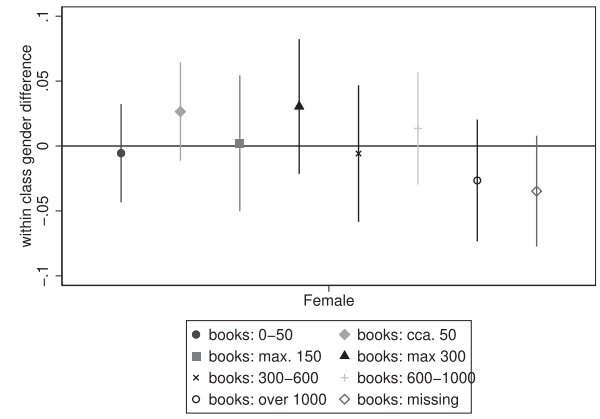

Fig. A.28. Within class gender differences, number of books.

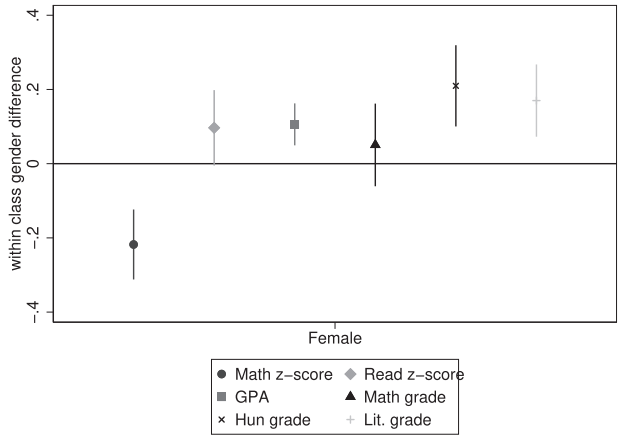

Fig. A.29. Within class gender differences, cognitive skills and school performance.

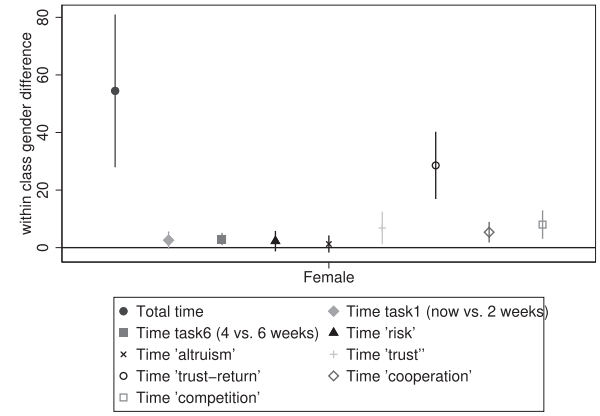

Fig. A.30. Within class gender differences, time spent on tasks.

Appendix B. Tables

Table B.5
Descriptive statistics by gender - NABC data.

	Female			***Male***		
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs
NABC data						
Age (in months)	202.34	13.61	611	201.65	13.34	477
Family						
Parental ed.: low	0.02	0.14	611	0.01	0.12	477
Parental ed.: medium	0.42	0.49	611	0.29	0.45	477
Parental ed.: high	0.47	0.50	611	0.58	0.49	477
Parental ed.: missing	0.09	0.29	611	0.11	0.32	477
Father: employed	0.64	0.48	611	0.68	0.47	477
Father: self-employed	0.17	0.38	611	0.12	0.32	477
Father: regural work	0.02	0.15	611	0.02	0.14	477
Father: occasional work	0.01	0.11	611	0.02	0.14	477
Father: childcare	0.01	0.11	611	0.02	0.13	477
Father: retired	0.01	0.11	611	0.00	0.06	477
Father: unemployed	0.01	0.09	611	0.00	0.05	477
Father: disabled	0.01	0.10	611	0.00	0.06	477
Father: missing	0.12	0.32	611	0.14	0.35	477
Child support: no	0.76	0.43	611	0.77	0.42	477
Child support: yes	0.14	0.35	611	0.10	0.30	477
Child support: missing	0.10	0.30	611	0.12	0.33	477
No. books: 0–50	0.06	0.24	611	0.06	0.24	477
No. books: cca. 50	0.10	0.30	611	0.06	0.24	477
No. books: max. 150	0.18	0.39	611	0.16	0.37	477
No. books: max 300	0.16	0.37	611	0.13	0.34	477
No. books: 300–600	0.14	0.35	611	0.14	0.35	477
No. books: 600–1000	0.15	0.36	611	0.16	0.37	477
No. books: over 1000	0.12	0.32	611	0.18	0.38	477
No. books: missing	0.09	0.28	611	0.11	0.31	477
Cognitive skills						
Math score, 6th grade	−0.23	0.91	609	0.29	1.03	472
Reading score, 6th grade	−0.05	0.99	609	0.07	1.01	472
Grades						
GPA, imputed	4.54	0.42	611	4.51	0.45	477
GPA, missing	0.19	0.39	611	0.20	0.40	477
Math grade, imputed	4.17	0.84	611	4.27	0.85	477
Hungarian grade, imputed	4.39	0.69	611	4.31	0.75	477
Literature grade, imputed	4.56	0.64	611	4.48	0.65	477
Math grade, missing	0.13	0.34	611	0.16	0.37	477
Hungarian grade, missing	0.13	0.34	611	0.16	0.37	477
Literature grade, missing	0.14	0.34	611	0.17	0.37	477

Table B.6
Descriptive statistics by gender - Experimental tasks.

	Female			***Male***		
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs
Experiments						
Payoff	1031.91	760.59	611	1054.93	925.16	477
Delta	0.73	0.15	602	0.77	0.16	467
Beta	1.05	0.23	590	1.00	0.21	457
Risk	30.80	18.18	606	37.27	18.45	474
Altruism (classmate)	41.72	15.78	611	35.79	20.88	477
Altruism (schoolmate)	29.60	18.44	611	25.05	22.42	477
Trust	51.55	21.23	611	59.77	29.25	477
Trustworthiness	37.26	14.11	611	40.31	18.60	477
Cooperation	58.76	25.38	611	62.60	30.12	477
Competition	0.56	0.50	611	0.66	0.47	477
Time						
Total time spent on tasks	877.96	198.71	611	803.76	203.99	477
Time spent on Task 1 (time now vs 2 weeks)	40.93	19.97	611	34.50	18.10	477
Time spent on Task 6 (time 4 vs 6 weeks)	41.60	23.19	611	36.29	25.17	477
Time spent on Task 1 and 6 (both time tasks)	82.54	38.38	611	70.79	38.87	477
Time spent on Risk	72.36	20.93	611	68.15	25.75	477
Time spent on Altruism (classmate)	39.12	17.72	611	37.82	21.45	477
Time spent on Altruism (schoolmate)	26.98	12.61	611	25.08	13.36	477
Time spent on Trust	63.55	36.46	611	55.38	29.94	477
Time spent on Trustworthiness	168.64	79.22	611	135.39	68.98	477
Time spent on Cooperation	86.62	38.07	611	77.30	38.06	477
Time spent on Competition	89.33	38.50	611	79.96	35.03	477

Table B.7
Comparing the coefficients of the female dummy in different specifications to the coefficients of the female dummy of the specification with the class fixed effects.

	Unadjusted	All Exogenous	All and time	All and time and preferences
Patience	.007	.443	.695	.364
Time inconsistency	.005	.496	.817	.48
Present bias	.776	.871	.553	.465
Risk tolerance	.059	.231	.05	0
Altruism (classmate)	.159	.419	.059	.157
Altruism (schoolmate)	.113	.155	.128	.23
Trust	.222	.352	.274	.024
Trustworthiness	.51	.656	.459	.149
Cooperation	.025	.026	.018	.006
Competition	.592	.835	.458	.07

Note: p-values of Chi-squared tests are in each cell. *All exogenous* refers to the specification that includes class FE, Age, Family, Cognitive skills and Grades.

Appendix C. Regression tables

Appendix C contains Tables C.9-C.18.

Table C.8
Adjusted gender differences in time discounting (delta).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−0.040***	−0.018*	−0.018*	−0.018	−0.010	−0.014	−0.020*	−0.013
Age (in months)			−0.000	0.000	−0.000	−0.000	−0.000	−0.000
Child support: missing				0.023	0.026	0.034	0.035	0.031
Child support: yes				0.006	0.008	0.008	0.009	0.008
Father: self-employed				−0.018	−0.019	−0.018	−0.018	−0.020
Father: regural work				0.012	0.017	0.021	0.016	0.016
Father: occasional work				−0.016	−0.029	−0.024	−0.022	−0.044
Father: childcare				−0.081*	−0.078*	−0.069	−0.070*	−0.066
Father: retired				−0.008	−0.016	−0.008	−0.007	−0.015
Father: unemployed				0.053	0.057**	0.041*	0.032*	0.034*
Father: disabled				−0.028	−0.031	−0.018	−0.011	−0.015
Father: missing				−0.003	−0.010	−0.007	−0.009	−0.008
Parental ed.: missing				−0.094	−0.097	−0.096	−0.097	−0.098
Parental ed.: medium				−0.042	−0.050	−0.046	−0.043	−0.050
Parental ed.: high				−0.031	−0.036	−0.035	−0.035	−0.040
No. books: cca. 50				0.035	0.040	0.038	0.034	0.028
No. books: max. 150				0.019	0.018	0.016	0.006	0.004
No. books: max 300				0.053**	0.051**	0.048**	0.041*	0.041*
No. books: 300–600				0.036	0.033	0.027	0.021	0.021
No. books: 600–1000				0.056**	0.052*	0.047*	0.043	0.042
No. books: over 1000				0.048*	0.043	0.035	0.030	0.023
No. books: missing				0.093*	0.105*	0.109**	0.101*	0.100*
Math score, 6th grade					0.037***	0.031***	0.029***	0.028***
Reading score, 6th grade					−0.007	−0.009	−0.008	−0.011
GPA, imputed						−0.006	−0.008	−0.004
GPA, missing						−0.012	−0.010	−0.016
Math grade, imputed						0.016*	0.015	0.014
Hungarian grade, imputed						0.009	0.009	0.009
Literature grade, imputed						0.006	0.007	0.007
Math grade, missing						−0.052	−0.050	−0.036
Hungarian grade, missing						0.053	0.053	0.054
Literature grade, missing						−0.008	−0.007	−0.023
Time spent on Task 6 (time 4 vs 6 weeks)							−0.001**	−0.001**
Total time spent on tasks							0.000***	0.000***
Competition								0.013
Altruism								0.000
Trust								0.000
Cooperation								0.000
Risk								0.001**
Constant	0.766***	0.754***	0.758***	0.740***	0.788***	0.706***	0.632***	0.572**
Observations	1069	1069	1069	1069	1063	1063	1063	1055
Adjusted R^2	0.016	0.089	0.088	0.091	0.114	0.116	0.131	0.139

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.9
Adjusted gender differences in time inconsistency (beta).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	0.020**	0.004	0.004	0.005	−0.004	0.000	0.006	−0.001
Age (in months)			0.000	0.000	0.000	0.000	0.000	0.001
Child support: missing				0.014	0.011	0.012	0.014	0.021
Child support: yes				0.004	0.004	0.005	0.002	0.003
Father: self-employed				0.018	0.019	0.016	0.016	0.018
Father: regural work				0.002	−0.004	−0.008	−0.003	−0.002
Father: occasional work				−0.005	0.003	−0.003	−0.009	−0.001
Father: childcare				0.111	0.110	0.104	0.105	0.097
Father: retired				−0.035	−0.027	−0.040	−0.045	−0.047
Father: unemployed				−0.020	−0.033	−0.019	−0.008	−0.002
Father: disabled				0.012	0.013	−0.001	−0.003	−0.000
Father: missing				−0.034	−0.030	−0.037*	−0.035*	−0.035
Parental ed.: missing				0.020	0.022	0.031	0.029	0.027
Parental ed.: medium				0.034	0.042	0.040	0.037	0.041
Parental ed.: high				0.036	0.040	0.040	0.039	0.042
No. books: cca. 50				0.014	0.011	0.011	0.014	0.020
No. books: max. 150				−0.022	−0.021	−0.019	−0.010	−0.010
No. books: max 300				−0.047*	−0.043	−0.043	−0.036	−0.035
No. books: 300–600				−0.055**	−0.052**	−0.049*	−0.044*	−0.042*
No. books: 600–1000				−0.035	−0.032	−0.029	−0.024	−0.025
No. books: over 1000				0.003	0.008	0.013	0.020	0.024
No. books: missing				−0.006	−0.014	−0.021	−0.014	−0.017
Math score, 6th grade					−0.028**	−0.021*	−0.020*	−0.019
Reading score, 6th grade					−0.000	0.000	−0.003	−0.001
GPA, imputed						−0.003	−0.001	−0.005
GPA, missing						0.039	0.037	0.040
Math grade, imputed						−0.018*	−0.017*	−0.016*
Hungarian grade, imputed						−0.012	−0.010	−0.012
Literature grade, imputed						0.012	0.010	0.012
Math grade, missing						0.042	0.045	0.035
Hungarian grade, missing						−0.005	−0.013	−0.006
Literature grade, missing						−0.070**	−0.066**	−0.064*
Time spent on 'beta"							0.000	0.000
Total time spent on tasks							−0.000***	−0.000***
Competition								0.003
Altruism								0.001
Trust								−0.001
Cooperation								−0.000
Risk								−0.000
Constant	0.127***	0.135***	0.130	0.060	0.039	0.105	0.176	0.191
Observations	1047	1047	1047	1047	1041	1041	1041	1033
Adjusted R^2	0.002	0.003	0.002	0.006	0.015	0.018	0.026	0.028

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.10
Adjusted gender differences in present bias (beta<1).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−0.069**	−0.077**	−0.077**	−0.075**	−0.080**	−0.079**	−0.068*	−0.064*
Age (in months)			−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
Child support: missing				−0.012	−0.011	−0.020	−0.012	−0.001
Child support: yes				−0.022	−0.022	−0.024	−0.029	−0.027
Father: self-employed				−0.040	−0.041	−0.042	−0.041	−0.041
Father: regural work				−0.192**	−0.192**	−0.195**	−0.189**	−0.182*
Father: occasional work				0.162	0.159	0.157	0.147	0.116
Father: childcare				−0.089	−0.092	−0.088	−0.087	−0.102
Father: retired				−0.081	−0.076	−0.071	−0.076	−0.089
Father: unemployed				0.332	0.343	0.338	0.360*	0.384*
Father: disabled				0.025	0.026	0.029	0.029	0.031
Father: missing				−0.034	−0.034	−0.023	−0.019	−0.019
Parental ed.: missing				−0.231	−0.226	−0.242	−0.256	−0.272
Parental ed.: medium				−0.066	−0.065	−0.075	−0.085	−0.092
Parental ed.: high				−0.045	−0.044	−0.056	−0.063	−0.071
No. books: cca. 50				−0.068	−0.068	−0.064	−0.058	−0.061
No. books: max. 150				−0.062	−0.064	−0.052	−0.036	−0.038
No. books: max 300				−0.029	−0.031	−0.020	−0.007	−0.008
No. books: 300–600				−0.047	−0.048	−0.033	−0.022	−0.025
No. books: 600–1000				−0.018	−0.018	−0.005	0.005	0.005
No. books: over 1000				−0.020	−0.021	−0.007	0.005	0.016
No. books: missing				0.141	0.133	0.139	0.151	0.152
Math score, 6th grade					−0.008	−0.011	−0.012	−0.011
Reading score, 6th grade					0.010	0.011	0.005	0.007
GPA, imputed						−0.040	−0.037	−0.043
GPA, missing						−0.006	−0.012	−0.021
Math grade, imputed						0.009	0.013	0.015
Hungarian grade, imputed						0.042	0.045	0.040
Literature grade, imputed						−0.033	−0.036	−0.035
Math grade, missing						0.053	0.058	0.067
Hungarian grade, missing						−0.025	−0.046	−0.050
Literature grade, missing						−0.012	0.005	0.005
Time spent on 'beta"							0.000	0.000
Total time spent on tasks							−0.000	−0.000
Competition								0.016
Altruism								0.000
Trust								−0.001
Cooperation								−0.001
Risk								0.001
Constant	0.368***	0.372***	0.409	0.513	0.521	0.643	0.786*	0.887*
Observations	1047	1047	1047	1047	1041	1041	1041	1033
Adjusted R^2	0.004	0.059	0.058	0.054	0.053	0.047	0.051	0.048

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.11
Adjusted gender differences in risk tolerance.

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−6.466***	−4.901***	−4.816***	−4.621***	−4.969***	−4.289***	−3.926***	−2.180
Age (in months)			0.119	0.131	0.130	0.144	0.141	0.119
Child support: missing				1.347	1.815	0.638	0.909	−1.740
Child support: yes				−1.664	−1.494	−1.678	−1.846	−1.530
Father: self-employed				1.808	1.619	1.761	1.743	1.579
Father: regural work				−2.813	−2.654	−2.989	−2.850	−3.571
Father: occasional work				12.211***	11.258***	11.448***	11.166***	12.609***
Father: childcare				0.758	0.741	0.425	0.495	3.266
Father: retired				8.632*	7.611	6.768	6.593	6.760
Father: unemployed				−6.952	−5.308	−3.841	−3.179	−4.175
Father: disabled				6.206	6.166	4.410	4.446	5.022
Father: missing				1.757	1.395	1.038	1.191	1.730
Parental ed.: missing				−0.053	0.073	−1.498	−2.050	−0.500
Parental ed.: medium				6.580**	6.671**	6.574**	6.259*	5.707*
Parental ed.: high				6.090*	6.095*	6.370*	6.153*	5.541
No. books: cca. 50				2.297	2.301	2.250	2.366	1.783
No. books: max. 150				1.899	1.339	1.044	1.395	2.088
No. books: max 300				0.462	−0.455	−0.226	0.042	0.029
No. books: 300–600				−0.153	−1.075	−0.588	−0.374	0.556
No. books: 600–1000				0.124	−0.697	−0.437	−0.206	0.224
No. books: over 1000				2.639	1.093	1.543	1.818	0.869
No. books: missing				7.480	6.354	5.799	6.059	7.510
Math score, 6th grade					0.559	0.835	0.811	−0.039
Reading score, 6th grade					2.701**	3.125***	2.979***	2.823**
GPA, imputed						−1.618	−1.558	0.368
GPA, missing						1.757	1.543	0.716
Math grade, imputed						0.011	0.105	−0.209
Hungarian grade, imputed						−1.908	−1.846	−2.023
Literature grade, imputed						−0.270	−0.362	−0.155
Math grade, missing						−7.356	−7.072	−4.062
Hungarian grade, missing						0.482	−0.311	−3.534
Literature grade, missing						9.807	10.412	10.034
Time spent on 'risk"							0.011	−0.008
Total time spent on tasks							−0.007*	−0.004
Competition								2.810**
Altruism								−0.146***
Trust								0.129***
Cooperation								0.011
Delta								10.767***
Constant	37.268***	36.390***	12.385	1.934	3.119	15.939	20.634	4.745
Observations	1080	1080	1080	1080	1073	1073	1073	1055
Adjusted R^2	0.029	0.076	0.077	0.078	0.091	0.097	0.098	0.145

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.12
Adjusted gender differences in altruism (classmate).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	5.932***	4.887***	4.926***	5.238***	4.679***	4.517***	3.949**	4.001**
Age (in months)			0.053	0.047	0.051	0.055	0.074	0.074
Child support: missing				−0.244	−0.187	−0.481	0.627	−0.198
Child support: yes				1.008	0.924	0.675	0.551	1.010
Father: self-employed				−1.150	−1.092	−1.249	−1.094	−0.736
Father: regural work				−3.360	−3.547	−3.304	−3.554	−3.948
Father: occasional work				6.638	6.661	6.766	7.915*	5.973
Father: childcare				1.181	1.123	1.056	0.678	0.900
Father: retired				5.080***	5.260***	5.090***	4.639**	4.695**
Father: unemployed				−2.885	−2.589	−3.602	−2.639	−3.010
Father: disabled				1.888	2.116	2.854	1.817	1.646
Father: missing				1.442	1.683	1.766	2.101	2.195
Parental ed.: missing				−1.788	−1.266	−0.011	−1.207	−1.238
Parental ed.: medium				−0.812	−0.576	−0.749	−0.845	−0.996
Parental ed.: high				0.370	0.571	0.299	−0.021	−0.376
No. books: cca. 50				0.276	0.314	0.518	0.524	0.900
No. books: max. 150				0.364	0.451	0.745	1.049	2.672
No. books: max 300				0.158	0.222	0.461	0.739	1.553
No. books: 300–600				2.487	2.605	2.758	3.166	3.841
No. books: 600–1000				1.420	1.590	1.845	1.752	3.182
No. books: over 1000				3.887	3.924	4.099	5.328	6.825**
No. books: missing				1.120	1.120	0.547	0.941	1.964
Math score, 6th grade					−1.435	−1.459	−1.344	−1.321
Reading score, 6th grade					1.117	0.828	0.658	0.984
GPA, imputed						1.393	1.133	1.394
GPA, missing						0.942	1.543	2.454
Math grade, imputed						0.210	0.109	−0.146
Hungarian grade, imputed						0.387	0.305	0.117
Literature grade, imputed						−0.015	0.261	−0.350
Math grade, missing						10.816*	10.811*	10.604*
Hungarian grade, missing						−22.208***	−21.889***	−22.343***
Literature grade, missing						9.659***	8.265***	9.292***
Time spent on 'altruism (classmate)'							−0.227***	−0.232***
Total time spent on tasks							0.018***	0.016***
Competition								−0.996
Risk								−0.080**
Delta								4.894
Constant	35.793***	36.380***	25.704	25.383	24.520	15.104	6.209	9.291
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.025	0.082	0.081	0.074	0.074	0.075	0.106	0.119

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.13
Adjusted gender differences in altruism (schoolmate).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	4.550***	3.142**	3.090**	3.127**	3.421**	4.008**	4.150***	4.056***
Age (in months)			−0.071	−0.076	−0.081	−0.063	−0.063	−0.038
Child support: missing				−1.776	−1.830	−1.384	−1.226	−2.826
Child support: yes				0.571	0.412	0.363	0.260	0.369
Father: self-employed				−0.914	−0.844	−0.892	−0.828	−0.516
Father: regural work				−2.525	−2.628	−2.791	−2.670	−2.842
Father: occasional work				3.420	3.549	3.676	3.684	3.496
Father: childcare				3.130	3.302	2.596	2.732	3.030
Father: retired				7.679**	8.135**	7.721**	7.690**	8.002**
Father: unemployed				5.402	4.694	5.975	6.348	5.248
Father: disabled				3.696	3.613	1.952	2.023	2.146
Father: missing				−0.577	−0.681	−1.126	−1.003	−0.705
Parental ed.: missing				6.882	6.998	8.292	8.114	8.522
Parental ed.: medium				5.508	5.397	5.564	5.416	5.507
Parental ed.: high				5.682	5.699	6.245	6.089	5.485
No. books: cca. 50				−1.071	−0.783	−0.648	−0.597	−0.317
No. books: max. 150				1.582	1.916	1.443	1.646	2.735
No. books: max 300				0.300	0.850	0.408	0.554	1.114
No. books: 300–600				1.417	1.986	1.562	1.596	1.889
No. books: 600–1000				−0.059	0.369	0.291	0.415	1.544
No. books: over 1000				3.433	4.266	4.158	4.248	5.526
No. books: missing				−0.398	0.700	0.713	0.681	1.533
Math score, 6th grade					0.885	1.459	1.417	1.003
Reading score, 6th grade					−1.803	−1.423	−1.468	−0.919
GPA, imputed						0.566	0.595	0.373
GPA, missing						−2.229	−2.313	−1.558
Math grade, imputed						−0.875	−0.803	−0.768
Hungarian grade, imputed						−3.252**	−3.230**	−3.232**
Literature grade, imputed						1.645	1.590	1.045
Math grade, missing						11.731	11.782	11.780
Hungarian grade, missing						−14.642*	−14.705*	−15.280*
Literature grade, missing						3.653	3.705	4.038
Time spent on 'altruism (schoolmate)'							0.035	0.026
Total time spent on tasks							−0.004	−0.004
Competition								−1.251
Risk								−0.068*
Delta								7.984*
Constant	25.048***	25.838***	40.207	34.806	35.277	39.177	41.188	36.783
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.011	0.054	0.053	0.044	0.045	0.047	0.046	0.052

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.14
Adjusted gender differences in trust.

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−8.215***	−6.825***	−6.732***	−6.471***	−6.616***	−6.080***	−5.940***	−4.802**
Age (in months)			0.126	0.163	0.164	0.183	0.176	0.169
Child support: missing				11.106*	11.617**	11.038*	12.517**	12.664**
Child support: yes				2.007	2.148	1.887	1.678	2.076
Father: self-employed				0.598	0.381	0.252	0.466	0.230
Father: regural work				−0.289	−0.023	−0.209	−0.544	−0.038
Father: occasional work				7.651*	6.429	6.387	6.213	3.244
Father: childcare				−11.834*	−11.806*	−11.858*	−12.490**	−11.800*
Father: retired				3.928	2.968	1.294	1.113	−1.111
Father: unemployed				4.099	5.824	6.179	6.758	7.082
Father: disabled				5.784	5.644	4.108	3.509	2.699
Father: missing				2.529	2.006	1.545	1.363	1.216
Parental ed.: missing				−0.562	−0.514	−0.822	−1.348	−2.374
Parental ed.: medium				3.839	3.762	3.548	3.103	1.062
Parental ed.: high				5.281	5.220	5.236	4.851	2.090
No. books: cca. 50				4.533	4.645	4.581	4.578	3.668
No. books: max. 150				0.303	−0.205	−0.158	−0.068	1.071
No. books: max 300				1.296	0.491	0.937	1.029	2.121
No. books: 300–600				0.535	−0.261	0.367	0.583	1.742
No. books: 600–1000				−2.299	−2.977	−2.442	−2.229	−0.072
No. books: over 1000				10.230***	8.934**	9.648**	9.539**	10.465**
No. books: missing				−8.885	−9.631	−11.045	−11.478	−12.548
Math score, 6th grade					1.273	1.491	1.564	0.990
Reading score, 6th grade					2.037	2.306	2.431	2.217
GPA, imputed						−5.175*	−5.206*	−5.200**
GPA, missing						6.357**	6.731**	7.197**
Math grade, imputed						0.880	0.992	0.980
Hungarian grade, imputed						−1.387	−1.319	−0.908
Literature grade, imputed						1.719	1.562	1.170
Math grade, missing						−2.809	−3.493	0.431
Hungarian grade, missing						−10.867*	−10.647*	−12.035**
Literature grade, missing						10.611	10.397	7.443
Time spent on 'trust"							−0.081**	−0.084**
Total time spent on tasks							0.008	0.009
Competition								0.780
Risk								0.192***
Delta								10.365*
Constant	59.769***	58.989***	33.555	18.953	19.711	32.632	32.563	19.218
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.025	0.054	0.054	0.071	0.075	0.077	0.080	0.110

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.15
Adjusted gender differences in trustworthiness (trust return).

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−3.058**	−3.505***	−3.449***	−3.193**	−3.088**	−3.309**	−3.170**	−2.632**
Age (in months)			0.074	0.067	0.063	0.070	0.074	0.087
Child support: missing				4.802	4.806	5.218	5.612	4.988
Child support: yes				2.189	2.234	2.406	2.317	2.018
Father: self-employed				−1.324	−1.348	−1.275	−1.138	−1.010
Father: regural work				1.513	1.553	1.757	1.560	1.919
Father: occasional work				5.961	5.906	6.052	5.542	2.112
Father: childcare				−2.788	−2.817	−2.640	−2.442	−2.284
Father: retired				−2.447	−2.575	−1.990	−2.001	−2.417
Father: unemployed				3.623	3.549	3.687	4.815	4.761
Father: disabled				6.629	6.586	6.842	6.619	6.259
Father: missing				−1.584	−1.638	−1.977	−1.756	−1.731
Parental ed.: missing				−3.150	−3.237	−3.206	−3.672	−4.294
Parental ed.: medium				0.881	0.836	1.418	0.904	−0.169
Parental ed.: high				0.844	0.769	1.391	0.855	−0.553
No. books: cca. 50				−0.176	−0.227	−0.303	0.137	0.164
No. books: max. 150				0.016	−0.079	−0.415	0.070	1.107
No. books: max 300				−0.037	−0.206	−0.453	0.204	1.049
No. books: 300–600				1.079	0.886	0.343	0.901	1.706
No. books: 600–1000				−1.716	−1.932	−2.319	−1.864	−0.280
No. books: over 1000				5.353**	5.072*	4.634*	5.274*	6.296**
No. books: missing				1.395	1.157	0.941	1.244	1.518
Math score, 6th grade					0.339	0.584	0.492	0.444
Reading score, 6th grade					0.117	−0.126	−0.436	−0.228
GPA, imputed						0.240	0.121	0.101
GPA, missing						0.048	−0.246	−0.023
Math grade, imputed						−1.016	−0.808	−1.020
Hungarian grade, imputed						−0.353	−0.310	−0.516
Literature grade, imputed						2.203*	2.171*	1.746
Math grade, missing						−1.590	−1.629	−0.979
Hungarian grade, missing						3.143	2.931	2.711
Literature grade, missing						−1.197	−0.846	−0.724
Time spent on 'trust-return"							0.023**	0.025**
Total time spent on tasks							−0.015***	−0.016***
Competition								−0.972
Risk								0.012
Delta								4.105
Constant	40.313***	40.564***	25.511	25.123	26.038	19.186	26.605	25.869
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.008	0.042	0.042	0.048	0.046	0.044	0.051	0.051

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.16
Adjusted gender differences in cooperation.

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−3.842**	−0.947	−0.876	−0.473	−0.083	0.662	0.904	1.356
Age (in months)			0.096	0.094	0.096	0.110	0.067	0.075
Child support: missing				5.313	5.971	6.132	5.847	5.616
Child support: yes				2.606	2.787	2.826	3.624	3.966
Father: self-employed				0.416	0.132	0.049	−0.236	−0.039
Father: regural work				6.335	6.960	6.439	5.193	5.869
Father: occasional work				21.340***	19.805***	19.507***	18.027***	17.014***
Father: childcare				−7.196	−7.143	−7.378	−6.690	−5.946
Father: retired				−0.499	−1.885	−1.759	−5.488	−6.679
Father: unemployed				5.067	7.379	9.463	10.719	11.181
Father: disabled				6.578	6.209	3.989	2.488	2.554
Father: missing				−1.687	−2.535	−2.339	−3.325	−3.172
Parental ed.: missing				−4.577	−4.880	−6.367	−5.996	−6.861
Parental ed.: medium				2.906	2.484	1.942	0.626	−0.831
Parental ed.: high				3.724	3.505	3.368	2.057	0.024
No. books: cca. 50				0.593	0.814	1.269	2.084	1.942
No. books: max. 150				−2.759	−3.397	−2.873	−2.880	−1.716
No. books: max 300				0.610	−0.409	0.269	−0.073	0.800
No. books: 300–600				−7.037	−8.054	−6.926	−6.847	−6.098
No. books: 600–1000				−5.642	−6.478	−5.256	−5.826	−3.920
No. books: over 1000				4.884	3.349	4.773	4.626	4.725
No. books: missing				8.713	8.026	8.854	8.630	8.248
Math score, 6th grade					3.015*	3.702**	3.737**	3.653**
Reading score, 6th grade					1.729	2.202	3.062**	2.924**
GPA, imputed						−5.050*	−6.153*	−6.487**
GPA, missing						−1.322	−2.106	−2.133
Math grade, imputed						−1.370	−0.514	−0.774
Hungarian grade, imputed						0.544	0.920	1.347
Literature grade, imputed						−0.194	−1.060	−1.449
Math grade, missing						3.354	5.194	7.360
Hungarian grade, missing						−2.334	−2.530	−2.996
Literature grade, missing						1.280	0.605	−1.283
Time spent on 'cooperation"							−0.234***	−0.231***
Total time spent on tasks							0.019**	0.020**
Competition								0.830
Risk								0.109***
Delta								7.835
Constant	62.601***	60.975***	41.592	38.668	39.148	62.511	79.368**	70.126*
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.004	0.034	0.034	0.052	0.062	0.061	0.110	0.119

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.17
Adjusted gender differences in competition.

	None	Class FE	Age	Family	Cogn. skills	Grades	Time	Preferences
Female	−0.086***	−0.084***	−0.084***	−0.085***	−0.093***	−0.088***	−0.082***	−0.058*
performance in 1st round	0.007	0.037***	0.037***	0.037***	0.036***	0.037***	0.037***	0.041***
performance 2nd-1st round	0.054***	0.067***	0.067***	0.067***	0.066***	0.066***	0.067***	0.069***
Age (in months)			−0.001	−0.001	−0.001	−0.001	−0.002	−0.001
Child support: missing				−0.022	−0.018	−0.064	−0.073	−0.078
Child support: yes				−0.034	−0.037	−0.041	−0.048	−0.045
Father: self-employed				0.032	0.033	0.032	0.035	0.030
Father: regural work				0.073	0.070	0.063	0.070	0.059
Father: occasional work				0.186	0.182	0.173	0.159	0.212**
Father: childcare				0.037	0.035	0.029	0.018	0.022
Father: retired				0.261**	0.263**	0.240**	0.244*	0.225*
Father: unemployed				−0.289	−0.289	−0.273	−0.276*	−0.294*
Father: disabled				−0.101	−0.097	−0.106	−0.095	−0.108
Father: missing				−0.094	−0.092	−0.091	−0.091	−0.099
Parental ed.: missing				0.209	0.224	0.162	0.206	0.222
Parental ed.: medium				−0.006	−0.003	−0.022	0.001	−0.002
Parental ed.: high				−0.024	−0.021	−0.039	−0.016	−0.021
No. books: cca. 50				−0.039	−0.035	−0.035	−0.028	−0.028
No. books: max. 150				−0.076	−0.075	−0.057	−0.052	−0.038
No. books: max 300				−0.086	−0.088	−0.059	−0.052	−0.053
No. books: 300–600				−0.140	−0.142	−0.102	−0.105	−0.082
No. books: 600–1000				−0.093	−0.093	−0.068	−0.058	−0.037
No. books: over 1000				−0.063	−0.068	−0.035	−0.028	−0.021
No. books: missing				−0.185	−0.183	−0.200	−0.201	−0.240*
Math score, 6th grade					−0.011	−0.010	−0.014	−0.027
Reading score, 6th grade					0.024	0.023	0.015	0.008
GPA, imputed						−0.041	−0.038	−0.029
GPA, missing						0.134*	0.130*	0.152**
Math grade, imputed						−0.005	−0.001	−0.004
Hungarian grade, imputed						0.054	0.050	0.053
Literature grade, imputed						−0.072**	−0.069**	−0.073**
Math grade, missing						−0.317*	−0.301**	−0.264*
Hungarian grade, missing						0.128	0.131	0.124
Literature grade, missing						0.207*	0.187*	0.153
Time spent on 'competition"							−0.002***	−0.002***
Total time spent on tasks							0.000	0.000
Altruism								−0.001
Trust								0.001
Cooperation								0.000
Delta								0.142
Risk								0.002**
Constant	0.536***	0.319***	0.573	0.635	0.682	0.968	0.972	0.736
Observations	1088	1088	1088	1088	1081	1081	1081	1055
Adjusted R^2	0.081	0.126	0.125	0.124	0.124	0.134	0.142	0.150

Note: 1st and 2nd round performance and interruptions during the competition game are controlled for. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.18
Corrections for multiple hypothesis testing (p-values).

	Full model coeff.	p-value (orig.)	Westfall-Young	Bonferroni
Patience (Delta)	−0.013	0.267	0.259	0.310
Time inconsistency (Beta)	0.027	0.155	0.248	0.310
Present bias (Beta<1)	−0.064	0.095	0.208	0.286
Risk tolerance	−2.180	0.120	.	.
Altruism (classmate)	4.001	0.013	0.056	0.052
Altruism (schoolmate)	4.056	0.010	0.045	0.049
Trust	−4.802	0.023	0.071	0.068
Trustworthiness	−2.632	0.033	0.071	0.068
Cooperation	1.356	0.408	0.414	0.408
Competition	−0.058	0.059	.	.

Appendix D. Comparison of our results with the literature

In this section we carry out a more detailed comparison of our findings with the literature. To do so, for each preference we consider the studies that use an elicitation method that yields a gender dummy comparable to our gender dummies.[38] There are many studies that use a young subject pool (e.g. children in kindergarten or pre-school). We do not consider studies with such young participants as they are arguably very different from our subject pool of adolescents. Therefore, we discard studies that have only subjects younger than 10 years. We report rescaled coefficients with confidence intervals for sake of comparison, so we restrict our attention to those studies that report the standard errors (or statistics that allow to infer the standard error) of the point estimates. For each preference, we describe the inclusion criteria and provide a summary table that contains all the relevant information. We also provide a graph of the rescaled coefficients that allows for a direct comparison between these studies and ours (see Figures 1-6 in the main text).

D1. Time

The main inclusion criterion in case of time preferences is that the study in question should use an elicitation method that allows to infer discount rates or discount factors.

Studies that do not qualify because they are not comparable with our study: Golsteyn et al. (2014), Kosse and Pfeiffer (2012).

Studies that use an elicitation method that does not allow comparison: Luhrmann et al. (2018).

In Table D.19 we summarize the main features of the studies that are comparable to ours.

D2. Risk

The main inclusion criterion in case of risk attitudes is that the study in question should use the same elicitation method that we use. It is the bomb risk elicitation task (BRET) (Crosetto and Filippin, 2013).

Most of the studies do not use the BRET, but other elicitation tasks (e.g. Harbaugh et al. (2002) or Booth and Nolen (2012b)). From the few studies that use the BRET, some do not have children as subjects (Nielsen, 2019).

Piovesan and Willadsen (2021) applied a version of BRET, in which participants received a decision sheet with a road comprising 100 steps (from 1 to 100). They had to choose how many steps they wanted to take by crossing out the boxes on the sheet. They received one point for each step they take, but behind one of the steps there is a “bomb”, and stepping on it means they lose their points and receive zero points. The more steps a participant takes, the more risk taking the participant is.

In Andreoni et al. (2020), the authors use a method similar to BRET. There were 10 pencils in a jar in front of the participants. One of the pencils had a red mark on the bottom. Participants could choose to take as many pencils as they wanted from the jar and the pencils became theirs, as long as none of the pencils that they took had a red mark on the bottom. If a pencil had the red mark, then they lost all the pencils.

In Table D.20 we summarize the main features of the studies.

D3. Altruism

For sake of comparability, we consider only studies that use the standard dictator game. There are many studies that use a simplified version of the dictator game in which participants have to choose between two (or more) allocations. Other studies combine the standard dictator game with another task (e.g., a production phase) that introduces motives into the dictator’s decision beyond altruism. We disregard such papers. Furthermore, we consider studies that provide clear statistics and / or use an OLS regression so that results can be compared. Some studies report gender coefficients in a standard or ordered probit specification that makes comparison impossible with our findings.

The studies in Table D.21 satisfy all the inclusion criteria and hence are comparable to our findings. However, relevant differences remain. For instance, in Bettinger and Slonim (2006), Houser and Schunk (2009) and Deckers et al. (2015) participants are considerably younger than in our sample. In Bettinger and Slonim (2006) the sample consists of low-SES students, while we have a rather mid-to-high SES sample. The dictator game in Almås et al. (2017) is a bit special (the natural division there is giving half) that makes comparison tricky. Furthermore, the receiver’s identity also changes from study to study. Generally, the dictator splits the endowment with somebody that is less known to her / him than a classmate.

Here, we list the papers that were excluded and the criterion that they did not meet.

No data reported on gender differences: Eckel et al. (2011), Blake et al. (2015) John and Thomsen (2015).

Not a standard dictator game used to measure altruism: Fehr et al. (2008), Almås et al. (2010), Martinsson et al. (2011), Fehr et al. (2013), Bauer et al. (2014), Maggian and Villeval (2016), Brocas et al. (2017), Sutter et al. (2018).

Not an OLS specification: Harbaugh et al. (2003b), Benenson et al. (2007), Angerer et al. (2015), Chen et al. (2016).

In Table D.21 we summarize the main features of the studies that are comparable to ours.

[38] Our starting point to find the relevant studies was Sutter et al. (2019) that we complemented with the relevant papers published since 2019 that cite Sutter et al. (2019).

Table D.19
Time preferences - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Measure	Regression	Other controls
Castillo et al. (2011)	13–15, avg. 13.8	878 adolescents from 4 public middle schools in Georgia (US)	Yes, with Wal-Mart gift card.	USD 49 paid in a month vs USD X in 7 months, 20 decisions, X increasing in each decisions, ranging from USD 50.83 to USD 98.02	Raw gender difference in discount rates (for consistent choices) = 12.5 pps (85.4% for males, and 72.9% for females, t-stat=−3.09, p-value=0.002). Restricting attention to white students, the corresponding numbers are 78.6% for males, 71.2% for females, t-stat=−1.37, p-value=0.173.	
					OLS. Dep. var.: discount rate (0–100). First specification. Male dummy=10.68, SE=4.28.	Black, Hispanic/multi-racial, school fixed effects.
					OLS. Dep. var.: discount rate (0–100). First specification. Male dummy=9.34, SE=4.53.	Previous, plus gifted, special education, math and reading scores, free and reduced lunch (proxy for income), neighborhood characteristics.
Sutter et al. (2013)	10–18	661 students in 28 classes. Grades 5,7,9 and 11.	Yes, with money.	4 time horizons: today vs. 3 weeks, 3 weeks vs. 6 weeks, today vs. 1 year, 3 weeks vs. 1 year and 3 weeks. Two levels of stake: early amount EUR 4.05 vs. EUR 10.1, later payments increased gradually from EUR 4.05 (EUR 10.1) to EUR 5.95 (EUR 13.9)	OLS. Dep. var.: normalized future equivalent (future equivalent divided by the size of the early payment). Three weeks delay (today vs. 3 weeks, 3 weeks vs. 6 weeks). Female dummy= −0.01, SE=0.011; Female × upfront delay=0.003, SE=0.006; Female × high stakes= −0.007, SE=0.007.	Upfront-delay dummy, high stakes dummy, upfront-delay ×high stakes, age, risk aversion, ambiguity aversion, number of siblings, pocket money, German grade, Math grade.
					OLS. Dep. var.: normalized future equivalent (future equivalent divided by the size of the early payment). One year delay (today vs. 1 year, 3 weeks vs. 1 year and 3 weeks). Female dummy= −0.006, SE=0.013; Female × upfront delay=0.003, SE=0.004; Female × high stakes= −0.01, SE=0.006.	Same as above
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Yes, with school canteen vouchers.	Staircase method. 2 time horizons: today vs. 2 weeks, 4 weeks vs. 5 weeks. Early amount (HUF 1000 ∼ EUR 3), later amount ranging from HUF 1030 to HUF 2150.	OLS. First specification. Dep. var.: Discount factor (δ). Female dummy=−0.018, SE=0.01	Class fixed-effect.
					OLS. Last specification. Dep. var.: Discount factor (δ). Female dummy=−0.013, SE=0.0116	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.

Table D.20
Risk attitudes - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Regression	Other controls
Andreoni et al. (2020)	avg. 13.56	1295 adolescents from low-income and high-minority suburbs near Chicago (US)	Yes, with money.	OLS with school-level controls. Dep. var.: the number of pencils taken (out of 10). Female dummy=−0.81, SE=0.12.	Cognitive index, executive functioning index, personality measures (self-control, grit), demographic and SES variables (not specified).
Piovesan and Willadsen (2021)	7–16, avg. 11.15	340 Danish students in 19 classes.	Yes, with gifts.	Girls / boys make 33.6 / 37.2 steps (equivalent of boxes), Mann-Whitney test; p value=0.019.	
				OLS. Specification 1: Dep. var.: Steps taken (out of 100). Female dummy= −3.313, SE= 1.55.	Age in years, number of older / younger sisters / brothers.
				OLS. Last specification: Dep. var.: Steps taken (out of 100). Female dummy= −3.194, SE= 1.549.	Same as above, plus the personality traits measured by the HEXACO measure.
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Yes, with school canteen vouchers.	OLS. First specification. Dep. var.: Boxes taken out (out of 100). Female dummy=−4.9, SE=1.37	Class fixed-effects.
				OLS. Last specification. Dep. var.: Boxes taken out (out of 100). Female dummy=−2.180, SE=1.38	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.

D4. Trust

We focus on studies that apply the standard trust (also known as the investment) game (Berg et al., 1995). As before, we consider only studies that report gender differences and provide proper statistics (standard errors or t-values).

Felfe et al. (2021) uses trust game with adolescents, but gender differences are not reported. Sutter and Kocher (2007) report descriptives on gender differences, but do not provide appropriate statistics (only state that using Mann-Whitney tests the null hypothesis of no gender differences cannot be rejected).

In Table D.22 we summarize the main features of the studies comparable to ours.

D5. Cooperation

We consider studies that use the public goods game (PGG) to measure cooperation, have children and adolescents as subjects and report gender differences. We also exclude studies in which the subjects are too young compared to our sample.

Here we list the papers that were excluded and the criterion that they did not meet.

Not using the public goods game: Fan (2000),Houser et al. (2012), Cárdenas et al. (2014), Lergetporer et al. (2014), Angerer et al. (2016).

No gender dummy reported: Peters et al. (2004), John and Thomsen (2015)

Subjects too young: Hermes et al. (2020), because subjects are 6 years old.

In Table D.23 we summarize the main features of the studies comparable to ours.

D6. Competition

Several studies focus on how being exposed to competition affects performance, often reporting gender differences. However, here we concentrate on the studies that investigate who enters competition, so we exclude studies that examine other aspects of competition. Furthermore, we focus on studies that apply the protocol by Niederle and Vesterlund (2007). As before, the study should report gender differences (in form of a gender dummy), with the corresponding standard error also.

Here we list the papers that were excluded and the criterion that they did not meet.

A different aspect of competition is studied: Gneezy and Rustichini (2004)

Table D.21
Altruism - Summary of the main findings in the literature.

Study	Age	N	Receiver	Incentivized	Regression	Other controls	Remarks
Bettinger and Slonim (2006)	6–14	574 from Ohio (US)	Another child at an event.	Yes, with gift certificates.	OLS. Dep. var.: amount given to peers out of USD 10. Male dummy= −0.788, SE= 0.385.	Winner / loser of scholarship for poor children; Exchange rate in the dictator game; Family income; Number of children living at home; Private lottery; African-American, Age, Amounts parents gave	The means of giving in three dictator games with different cost of giving are used as a measure of altruism. The subjects come from low-income families (qualify for federal reduced/free lunch programs).
Houser and Schunk (2009)	8–10	151 elementary school students from 2 rural towns in Germany.	Unknown student from another school.	Yes, with candies.	OLS. Dep. var.: amount given to peers out of 20 M&Ms. Male dummy= 0.72, SE= 1.25.	Math grade, Social behavior grade, Treatment 2, Treatment2*male, Location.	Subjects divided 20 M&Ms. Three treatments. Treatment 1 is closest to ours.
Deckers et al. (2015)	7–10	732 children recruited using official registry in Bonn and Cologne (Germany).	Another unknown child of similar age from the same city / A child living in a city nearby / A child living in Africa	Yes, with toys.	OLS. Dep. var.: average share of endowment sent to the receiver (in 3 dictator games). Male dummy= −0.048, SE= 0.01.	Log income, Parental education, Child age.	Three different dictator games: a simple binary choice game (either 100% or 50% for the dictator) and two continuous dictator games in which subjects had to distribute 6 stars between themselves and the receiver.
Almås et al. (2017)	14–15	524 middle school students from 9 public schools in Bergen (Norway).	Classmate	Yes, with money.	OLS. Dep. var.: share of money sent. Female dummy= 0.004, SE= 0.022.	Low-SES dummy (bottom quintile of both income and parent education).	The dictator splits overall fixed earnings. The natural division is half-half. Not so clear comparison with standard dictator game.
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Classmate	Yes, with school canteen vouchers.	OLS. First specification. Dep. var.: share of endowment given to classmate (in %). Female dummy= 4.89, SE= 1.44	Class fixed-effects.	
					OLS. Last specification. Dep. var.: share of endowment given to classmate (in %). Female dummy= 4.001, SE= 1.557.	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.	
			Schoolmate	Not incentivized.	OLS. First specification. Dep. var.: share of endowment given to schoolmate (in %). Female dummy=3.142, SE= 1.43	Class fixed-effects.	
					OLS. Last specification. Dep. var.: share of endowment given to schoolmate (in %). Female dummy=4.056, SE=1.512.	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.	

Table D.22
Trust - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Co-player	Measure	Regression	Other controls
Harbaugh et al. (2003a)	8–17	153 children from Oregon (US)	Yes, with toys for third-graders and with money for older ones.	Students from 4 different grades and an adult. When playing against another student, the sender and the receiver were from different classes of the same grade.	Sender can send 0–4 tokens. Multiplier: 3.	OLS. Dep. var. for sender: amount passed to receiver (0–4). Sender: Male dummy= 0.13, adjusted t= 0.75.	Receiver's grade, survey on trust, sender's age, birth order, risk taking, relative height.
						OLS. Dep. var. for receiver: amount returned to sender (0–4). Receiver: Male dummy = −4.74, adjusted t= −1.04.	Sender's grade, survey on trust, receiver's age, number of siblings, birth order, tokens passed, relative height.
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Yes, with school canteen vouchers.	Classmate	Sender can send any amount between 0 and 1000 Ft. Multiplier: 3.	OLS. First specification. Dep. var. for sender: share of the endowment passed to receiver (in %). Sender: Female dummy = −6.825, SE=1.897	Class fixed-effects.
						OLS. Last specification. Dep. var. for sender: share of the endowment passed to receiver (in %). Sender: Female dummy = −4.802, SE=2.047.	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences. Class fixed-effects.
						OLS. First specification. Dep. var. for receiver: share of received (and multiplied) amount returned to sender (in %). Receiver: Female dummy= −3.505, SE= 1.191	
						OLS. Last specification. Dep. var. for receiver: share of received (and multiplied) amount returned to sender (in %). Receiver: Female dummy= −2.632, SE= 1.199.	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.

Table D.23
Cooperation - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Co-players	PGG parameters	Regression	Other controls
Harbaugh and Krause (2000)	6–12	208 children from New Mexico (US)	Yes, with gifts.	Participants at after-school and summer recreational programs. Possibly schoolmates.	Groups of 6. MPCR: $\frac{1}{3}$ and $\frac{2}{3}$.	OLS. Dep. var.: contribution in round 1 (out of 5 chips). First specification. Male dummy =−0.13, SE=0.11.	high-MPCR dummy, group attachment / familiarity with others measure, age, number of siblings, single parent household dummy.
						OLS. Dep. var.: contribution in round 1 (out of 5 chips). Last specification. Male dummy=−0.032, SE=0.12.	Same as above, except single parent household dummy, plus allowance, TV watching and church attendance.
Cipriani et al. (2013)	5–12	38 students from a public elementary school in Washington DC (US).	Yes, with toys.	Classmates.	Groups of 6. MPCR: $\frac{1}{2}$.	Panel regression. Dep. var.: contribution in round t (out of 5 tokens). First specification. Male dummy= −0.202, SE= 0.457.	Others' contribution in $t-1$.
						Panel regression. Dep. var.: contribution in round t (out of 5 tokens). Most comprehensive specification. Male dummy= −0.258, SE= 0.460.	Others' contribution in $t-1$, children's first round contribution, parents' first round contribution.
Cavatorta et al. (2020)	avg. 17–18 years	1172 adolescents from Palestine	Yes, with money.	Most probably classmates as experiments were run during regular school hours.	Groups of 4. MPCR: $\frac{1}{3}$	OLS. Dep. var.: contribution in one-shot PGG (out of 5 tokens). Female dummy=−0.079, SE (clustered at the school class) = 0.124.	Dummy on if the student has to cross a military checkpoint to go to school, age, class size, dummy on when the data was collected.
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Yes, with school canteen vouchers.	Classmates.	Groups of 2. MPCR: 0.75.	OLS. First specification. Dep. var.: share of the endowment contributed. Female dummy= −0.947, SE= 1.611.	Class fixed-effect.
						OLS. Last specification. Dep. var.: share of the endowment contributed. Female dummy= 1.356, SE= 1.626.	Age, child support, father employment, parental education, number of books, math and reading score, GPA, response time, all other preferences.

Not using the Niederle-Vesterlund protocol: Bartling et al. (2012),Andersen et al. (2013), Khadjavi and Nicklisch (2018), Alan and Ertac (2019).

No regressions with appropriate standard erros reported: Samak (2013).

In Table D.24 we summarize the main features of the studies comparable to ours. Note that in all cases, the dependent variable in the regression is whether the participant chooses to compete in round 3. When probit regressions were run, the marginal effects are reported, Table D.25 and D.26.

Table D.24
Competition -Summary of the main findings in the literature.

Study	Age	N	Incentivized	Co-player	Task	Regression	Other controls
Zhang (2011)	15–18	544 middle and high school students from rural China.	Yes, with money.	Class- or schoolmate. (Not clear from the description.)	Adding up two-digit numbers	Probit. For middle school: male dummy = −0.01, robust SE (clustered by session) = 0.143	Ethnic dummies (Mosuo and Yi), male x ethnic interactions, the logarithm of winning the tournament, risk tolerance, overconfidence, age, age^2, number of sisters, number of brothers, a dummy if the household is engaged in agriculture, education of the househead, and school fixed effects.
						Probit. For high school: male dummy= 0.045, robust SE (clustered by session) = 0.052	Same as above.
Booth and Nolen (2012a)	avg. 15	260 English students from grades 10 or 11 from 8 public schools.	Yes, with money.	(Mostly) with students from other schools.	Solving mazes.	Probit. Female dummy (without other controls) = −0.267, SE= 0.061. Female dummy (last specification)= −0.601, SE= 0.278.	Tournament score in round 2, tournament-piece rate score, single sex school dummy, female×single-sex, all-girls group dummy, all-boys group dummy, risk attitude, female×risk attitude, mother went to university dummy, father went to university dummy, number of brothers, number of sisters, student is 14 dummy.
Buser et al. (2014)	avg. 15	362 secondary school students in grade 9 in and around Amsetrdam (Netherlands).	Yes, with money.	Classmates.	Adding up two-digit numbers.	OLS. Female dummy (first specification) = −0.233, SE= 0.047. Female dummy (last specification)= −0.117, SE= 0.045.	First specification: school fixed effects, the tournament performance, the difference between the tournament and the piece-rate performance and the probability of winning the round 2 tournament. Last specification: previous plus guessed rank in the round 2 tournament (a proxy for confidence), decision to enter a lottery (a proxy for risk attitude), answer to a question on risk-taking (another proxy for risk attitude), math grade, GPA, math relative (a ranking of the student in the class according to the math grade), math quartile (students' self-reported ranking in the class), math difficulty (how difficult the student finds math).
Dreber et al. (2014)	15–19, avg 17	216 high-school students from 5 schools in the Stockhom area (sweden).	Yes, with money.	Class- or schoolmate. (Not clear from the description.)	Two tasks: 1. adding up two-digit numbers, and 2. word search.	OLS. Math task. Female dummy (first specification, without controls) = −0.191, SE (bootstrapped, clustered on class)= 0.062. Female dummy (last specification)= −0.058, SE= 0.07.	Last specification: performance in the tournament (round 2) and beliefs about the performance in the tournament (a proxy for confidence) and risk attitude.
						OLS. Verbal task. Female dummy (first specification, without controls) = −0.056, SE (bootstrapped, clustered on class)= 0.089. Female dummy (last specification)= 0.066, SE= 0.089.	Same as above.

Table D.25
Competition - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Co-player	Task	Regression	Other controls
Khachatryan et al. (2015)	7–16	216 Armenian students. We focus on the experiments that use the Niederle and Vesterlund (2007) protocol.	Yes, students aged 7–12 with pens, students aged 12–16 with money.	Schoolmates (possibly classmates).	Two tasks: math task and verbal task, in random order.	Probit. Math task. Female dummy (first specification, without controls) = −0.0135, SE (clustered on class)= 0.0309. Female dummy (last specification)= 0.000308, SE= 0.0387.	Last specification: performance in round 2 (tournament), the difference in performance between round 1 and round 1, age, beliefs on relative performance, risk attitude, a measure on how gendered the participant perceived the math task, a measure on how much the participant enjoys competing, a measure of importance that the participant attaches to winning a competition overall / against a male / against a female.
						Probit. Verbal task. Female dummy (first specification, without controls) = −0.0140, SE (clustered on class)= 0.0359. Female dummy (last specification)= −0.0326, SE= 0.0414.	Same as above.
Sutter and Glätzle-Rützler (2015)	9–18	717 elementary and grammar school students from Austria. (We ignore a young sample of children aged 3–8.)	Yes, with money.	Most probably classmates as experiments were run during regular school hours.	Adding up two-digit numbers	Probit. Female dummy (first specification) = −0.2195, SE)= 0.034. Female dummy (last specification)= −0.165, SE= 0.036.	First specification: age. Last specification: previous plus performance in stage 2 (tournament), beliefs on winning in stage 2, and risk aversion.
						With a subsample of 316 students, the experiment was repeated two years later. Probit. Female dummy (first specification) = −0.2205, SE)= 0.062. Female dummy (last specification)= −0.161, SE= 0.069.	First specification: age and dummy on the choice in round 3 two years earlier. Last specification: previous plus performance in stage 2 (tournament), beliefs on winning in stage 2, and risk aversion.
Almås et al. (2016)	14–15	523 students from 11 public middle schools in the Bergen municipality (Norway).	Yes, with money.	Schoolmates (possibly classmates).	Adding up two-digit numbers	OLS. Female dummy (first specification, without further controls) = −0.194, SE)= 0.044. Female dummy (last specification)= −0.138, SE= 0.044.	Controls in the last specification: performance in round 1, overconfidence (the difference between actual performance and guessed performance), risk attitude, patience, being egalitarian, being selfish, extraversion, agreeableness, conscientiousness, neuroticism, openness, low SES dummy.
Sutter et al. (2016)	10–17	588 from Austrian secondary schools. Three grades: grade 5 (aged 10–11), grade 8 (aged 13–14), grade 11 (aged 16–17).	Yes, with money.	Most probably classmates as experiments were run during regular school hours.	Adding up two-digit numbers	Probit. 10–11 years old. Female dummy (unique specification) = −0.136, SE)= 0.091.	A dummy for preferential treatment, a dummy for minimum quota treatment (the baseline is the standard Niederle-Vesterlund (2007) setup), performance in round 2 (tournament), belief in winning in round 2.
						Probit. 13–14 years old. Female dummy (unique specification) = −0.138, SE)= 0.081.	Same as above.
						Probit. 16–17 years old. Female dummy (unique specification) = −0.394, SE)= 0.151.	Same as above.

Table D.26
Competition - Summary of the main findings in the literature.

Study	Age	N	Incentivized	Co-player	Task	Regression	Other controls
Cárdenas et al. (2012)	9–12	1200 students from Bogotá (Columbia) and Stockholm (Sweden)	Yes, with pens and erasers.	Classmates.	Math and verbal task.	OLS. Findings for Columbia. Female dummy (first specification) = 0.0274, SE= 0.0553. Female×math (in the same specification) coefficient=−0.0533, SE=0.0783. In the last specification: Female dummy= 0.102, SE= 0.0714; Female×math (in the same specification) coefficient=−0.207, SE=0.101.	The first specification also contains a math dummy. The additional controls in the last specification are performance (no information which round), expected performance, age, risk attitudes, a measure of importance that the participant attaches to winning a competition against a male / a female, a measure on how gendered the participant perceived the math / verbal task.
						OLS. Findings for Sweden. Female dummy (first specification) = −0.114, SE= 0.0540. Female×math (in the same specification) coefficient=−0.140, SE=0.0764. In the last specification: Female dummy= −0.114, SE= 0.0606; Female×math (in the same specification) coefficient=−0.0368, SE=0.0832.	Same as above.
Our study	16–21, avg. 17.5	1088 high-school students from Hungary.	Yes, with school canteen vouchers.	Classmate	Counting 0s and 1s in a matrix.	OLS. First specification. Female dummy= −0.084. SE=0.028	Class fixed-effect, performance in stage 1, difference in performance in stage 1 and 2.
						OLS. Last specification. Female dummy= −0.058. SE=0.0303	Age, child support, father employment, parental education, number of books, math and reading score, GPA, performance in stage 1, difference in performance in stage 1 and 2, response time, all other preferences.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.12.015.

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101 (2), 470–492.
Alan, S., Baydar, N., Boneva, T., Crossley, T.F., Ertac, S., 2017. Transmission of risk preferences from mothers to daughters. J. Econ. Behav. Organ. 134, 60–77.
Alan, S., Ertac, S., 2019. Mitigating the gender gap in the willingness to compete: evidence from a randomized field experiment. J Eur Econ Assoc 17 (4), 1147–1185.
Almås, I., Cappelen, A.W., Salvanes, K.G., Sørensen, E.Ø., Tungodden, B., 2016. Willingness to compete: family matters. Manage. Sci. 62 (8), 2149–2162.
Almås, I., Cappelen, A.W., Salvanes, K.G., Sørensen, E.Ø., Tungodden, B., 2017. Fairness and family background. Polit. Philos. Econ. 16 (2), 117–131.
Almås, I., Cappelen, A.W., Sørensen, E.Ø., Tungodden, B., 2010. Fairness and the development of inequality acceptance. Science 328 (5982), 1176–1178.
Andersen, S., Ertac, S., Gneezy, U., List, J.A., Maximiano, S., 2013. Gender, competitiveness, and socialization at a young age: evidence from a matrilineal and a patriarchal society. Rev. Econ. Stat. 95 (4), 1438–1443.
Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2008. Eliciting risk and time preferences. Econometrica 76 (3), 583–618.
Andreoni, J., Di Girolamo, A., List, J.A., Mackevicius, C., Samek, A., 2020. Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. J. Econ. Behav. Organ. 179, 729–742.
Angerer, S., Glätzle-Rützler, D., Lergetporer, P., Sutter, M., 2015. Donations, risk attitudes and time preferences: astudy on altruism in primary school children. J. Econ. Behav. Organ. 115, 67–74.
Angerer, S., Glätzle-Rützler, D., Lergetporer, P., Sutter, M., 2016. Cooperation and discrimination within and across language borders: evidence from children in a bilingual city. Eur. Econ. Rev. 90, 254–264.
Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: self-control by precommitment. Psychol. Sci. 13 (3), 219–224.
Bartling, B., Fehr, E., Schunk, D., 2012. Health effects on children's willingness to compete. Exp. Econ. 15 (1), 58–70.
Bauer, M., Chytilová, J., Pertold-Gebicka, B., 2014. Parental background and other-regarding preferences in children. Exp. Econ. 17 (1), 24–46.
Benenson, J.F., Pascoe, J., Radmore, N., 2007. Children's altruistic behavior in the dictator game. Evol. Hum. Behav. 28 (3), 168–175.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10 (1), 122–142.
Bertrand, M., 2011. New perspectives on gender. In: Handbook of Labor Economics, Vol. 4. Elsevier, pp. 1543–1590.
Bettinger, E., Slonim, R., 2006. Using experimental economics to measure the effects of a natural educational experiment on altruism. J. Public Econ. 90 (8–9), 1625–1648.
Bettinger, E., Slonim, R., 2007. Patience among children. J. Public Econ. 91 (1–2), 343–363.

Blake, P.R., Piovesan, M., Montinari, N., Warneken, F., Gino, F., 2015. Prosocial norms in the classroom: the role of self-regulation in following norms of giving. J. Econ. Behav. Organ. 115, 18–29.
Blau, F.D., Kahn, L.M., 2017. The gender wage gap: extent, trends, and explanations. J. Econ. Lit. 55 (3), 789–865.
Booth, A., Nolen, P., 2012. Choosing to compete: how different are girls and boys? J. Econ. Behav. Organ. 81 (2), 542–555.
Booth, A.L., Nolen, P., 2012. Gender differences in risk behaviour: does nurture matter? Econ. J. 122 (558), F56–F78.
Borghans, L., Heckman, J.J., Golsteyn, B.H., Meijers, H., 2009. Gender differences in risk aversion and ambiguity aversion. J. Eur. Econ. Assoc. 7 (2–3), 649–658.
Brocas, I., Carrillo, J.D., Kodaverdian, N., 2017. Altruism and Strategic Giving in Children and Adolescents.
Buser, T., Niederle, M., Oosterbeek, H., 2014. Gender, competitiveness, and career choices. Q. J. Econ. 129 (3), 1409–1447.
Cárdenas, J.-C., Dreber, A., von Essen, E., Ranehill, E., 2014. Gender and cooperation in children: experiments in colombia and sweden. PLoS ONE 9 (3), e90923.
Cárdenas, J.-C., Dreber, A., Von Essen, E., Ranehill, E., 2012. Gender differences in competitiveness and risk taking: comparing children in colombia and sweden. J. Econ. Behav. Organ. 83 (1), 11–23.
Castillo, M., Ferraro, P.J., Jordan, J.L., Petrie, R., 2011. The today and tomorrow of kids: time preferences and educational outcomes of children. J. Public Econ. 95 (11–12), 1377–1385.
Castillo, M., Jordan, J.L., Petrie, R., 2019. Discount rates of children and high school graduation. Econ. J. 129 (619), 1153–1181.
Cavatorta, E., Zizzo, D.J., Daoud, Y., et al., 2020. Does Exposure to Violence Affect Reciprocity?: Experimental Evidence from the West Bank. Technical Report. University of Queensland, School of Economics.
Chen, J., Houser, D., Montinari, N., Piovesan, M., 2016. Beware of popular kids bearing gifts: a framed field experiment. J. Econ. Behav. Organ. 132, 104–120.
Cipriani, M., Giuliano, P., Jeanne, O., 2013. Like mother like son? Experimental evidence on the transmission of values from parents to children. J. Econ. Behav. Organ. 90, 100–111.
Cornsweet, T.N., 1962. The staircase-method in psychophysics. Am. J. Psychol. 75 (3), 485–491.
Crosetto, P., Filippin, A., 2013. The “bomb” risk elicitation task. J. Risk Uncertain. 47 (1), 31–65.
Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Exp. Econ. 19 (3), 613–641.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474. doi:10.1257/jel.47.2.448.
Daly, M., Delaney, L., Egan, M., Baumeister, R.F., 2015. Childhood self-control and unemployment throughout the life span: evidence from two British cohort studies. Psychol. Sci. 26 (6), 709–723.
Deckers, T., Falk, A., Kosse, F., Pinger, P.R., Schildberg-Hörisch, H., 2017. Socio-economic Status and Inequalities in Children’s IQ and Economic Preferences.
Deckers, T., Falk, A., Kosse, F., Schildberg-Hörisch, H., 2015. How Does Socio-economic Status Shape a Child’s Personality?.
Dittrich, M., Leipold, K., 2014. Gender differences in time preferences. Econ. Lett. 122 (3), 413–415.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550.
Dreber, A., von Essen, E., Ranehill, E., 2014. Gender and competition in adolescence: task matters. Exp. Econ. 17 (1), 154–172.
Dunn, O.J., 1961. Multiple comparisons among means. J. Am. Stat. Assoc. 56 (293), 52–64.
Eckel, C., Grossman, P.J., Johnson, C.A., de Oliveira, A.C., Rojas, C., Wilson, R., 2011. Social norms of sharing in high school: teen giving in the dictator game. J. Econ. Behav. Organ. 80 (3), 603–612.
Eckel, C.C., Grossman, P.J., 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. Evol. Hum. Behav. 23 (4), 281–295.
Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: experimental evidence. Handb. Exp. Econ. Results 1, 1061–1073.
Eckel, C.C., Grossman, P.J., Johnson, C.A., de Oliveira, A.C., Rojas, C., Wilson, R.K., 2012. School environment and risk preferences: experimental evidence. J. Risk Uncertain. 45 (3), 265–292.
Ellison, G., Swanson, A., 2010. The gender gap in secondary school mathematics at high achievement levels: evidence from the american mathematics competitions. J. Econ. Perspect. 24 (2), 109–128.
Ertac, S., 2020. The formation and malleability of preferences and noncognitive skills. In: Handbook of Labor, Human Resources and Population Economics, pp. 1–27.
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global evidence on economic preferences. Q. J. Econ. 133 (4), 1645–1692.
Fan, C.-P., 2000. Teaching children cooperation-an application of experimental game theory. J. Econ. Behav. Organ. 41 (3), 191–209.
Fehr, E., Bernhard, H., Rockenbach, B., 2008. Egalitarianism in young children. Nature 454 (7208), 1079–1083.
Fehr, E., Glätzle-Rützler, D., Sutter, M., 2013. The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. Eur. Econ. Rev. 64, 369–383.
Felfe, C., Kocher, M.G., Rainer, H., Saurer, J., Siedler, T., 2021. More opportunity, more cooperation? The behavioral effects of birthright citizenship on immigrant youth. J. Public Econ. 200, 104448.
Filippin, A., Crosetto, P., 2016. A reconsideration of gender differences in risk attitudes. Manage. Sci. 62 (11), 3138–3160.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112 (2), 631–645.
Gneezy, U., Rustichini, A., 2004. Gender and competition at a young age. Am. Econ. Rev. 94 (2), 377–381.
Golsteyn, B.H., Grönqvist, H., Lindahl, L., 2014. Adolescent time preferences predict lifetime outcomes. Econ. J. 124 (580), F739–F761.
Harbaugh, W.T., Krause, K., 2000. Children’s altruism in public good and dictator experiments. Econ. Inq. 38 (1), 95–109.
Harbaugh, W.T., Krause, K., Liday, S.G., Vesterlund, L., 2003. Trust in children. In: Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research, vol. 302, p. 322.
Harbaugh, W.T., Krause, K., Liday, S.J., 2003. Bargaining by Children. University of Oregon Economics Working Paper 2002-4.
Harbaugh, W.T., Krause, K., Vesterlund, L., 2002. Risk attitudes of children and adults: choices over small and large probability gains and losses. Exp. Econ. 5 (1), 53–84.
Heilman, M.E., Chen, J.J., 2005. Same behavior, different consequences: reactions to men’s and women’s altruistic citizenship behavior. J. Appl. Psychol 90 (3), 431.
Hermes, H., Hett, F., Mechtel, M., Schmidt, F., Schunk, D., Wagner, V., 2020. Do children cooperate conditionally? Adapting the strategy method for first–graders. J. Econ. Behav. Organ. 179, 638–652.
Hill, J.P., Lynch, M.E., 1983. The intensification of gender-related role expectations during early adolescence. In: Girls at Puberty. Springer, pp. 201–228.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655.
Horn, D., Kiss, H. J., Lénárd, T., 2020. Economic preferences in the classroom - research documentation. 10.2139/ssrn.3615292
Houser, D., Montinari, N., Piovesan, M., 2012. Private and Public Decisions in Social Dilemmas: Evidence from Children’s Behavior.
Houser, D., Schunk, D., 2009. Social environments with competitive pressure: gender effects in the decisions of German schoolchildren. J. Econ. Psychol. 30 (4), 634–641.
John, K., Thomsen, S.L., 2015. School-track environment or endowment: what determines different other-regarding behavior across peer groups? Games Econ. Behav. 94, 122–141.
Khachatryan, K., Dreber, A., Von Essen, E., Ranehill, E., 2015. Gender and preferences at a young age: evidence from armenia. J. Econ. Behav. Organ. 118, 318–332.
Khadjavi, M., Nicklisch, A., 2018. Parents’ ambitions and children’s competitiveness. J. Econ. Psychol. 67, 87–102.
Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., Falk, A., 2020. The formation of prosociality: causal evidence on the role of social environment. J. Polit. Economy 128 (2), 434–467.
Kosse, F., Pfeiffer, F., 2012. Impatience among preschool children and their mothers. Econ. Lett. 115 (3), 493–495.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112 (2), 443–478.
Lergetporer, P., Angerer, S., Glätzle-Rützler, D., Sutter, M., 2014. Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. Proc. Natl. Acad. Sci. 111 (19), 6916–6921.
Luhrmann, M., Serra-Garcia, M., Winter, J., 2018. The impact of financial education on adolescents' intertemporal choices. Am. Econ. J. 10 (3), 309–332.
Luehrmann, M., Serra-Garcia, M., Winter, J., 2018. The impact of financial education on adolescents' intertemporal choices. Am. Econ. J. 10 (3), 309–332.
Maggian, V., Villeval, M.C., 2016. Social preferences and lying aversion in children. Exp. Econ. 19 (3), 663–685.
Martinsson, P., Nordblom, K., Rützler, D., Sutter, M., 2011. Social preferences during childhood and the role of gender and ag-an experiment in austria and sweden. Econ. Lett. 110 (3), 248–251.
Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., et al., 2011. A gradient of childhood self-control predicts health, wealth, and public safety. Proc. Natl. Acad. Sci. 108 (7), 2693–2698.
Niederle, M., 2016. Gender. In: Handbook of Experimental Economics. Princeton University Press, pp. 481–553.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Nielsen, K., 2019. Dynamic risk preferences under realized and paper outcomes. J. Econ. Behav. Organ. 161, 68–78.
Paglin, M., Rufolo, A.M., 1990. Heterogeneous human capital, occupational choice, and male-female earnings differences. J. Labor Econ. 8 (1, Part 1), 123–144.
Peters, H.E., Ünür, A.S., Clark, J., Schulze, W.D., 2004. Free-riding and the provision of public goods in the family: a laboratory experiment. Int. Econ. Rev. 45 (1), 283–299.
Phelps, E.S., Pollak, R.A., 1968. On second-best national saving and game-equilibrium growth. Rev. Econ. Stud. 35 (2), 185–199.
Piovesan, M., Willadsen, H., 2021. Risk preferences and personality traits in children and adolescents. J. Econ. Behav. Organ. 186, 523–532.
Price, C.R., 2012. Gender, competition, and managerial decisions. Manage. Sci. 58 (1), 114–122.
Rose, A.J., Rudolph, K.D., 2006. A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. Psychol. Bull 132 (1), 98.
Samak, A.C., 2013. Is there a gender gap in preschoolers' competitiveness? An experiment in the us. J. Econ. Behav. Organ. 92, 22–31.
Sinka, E., 2010. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Technical Report.
Sutter, M., Feri, F., Glätzle-Rützler, D., Kocher, M.G., Martinsson, P., Nordblom, K., 2018. Social preferences in childhood and adolescence. a large-scale experiment to estimate primary and secondary motivations. J. Econ. Behav. Organ. 146, 16–30.
Sutter, M., Glätzle-Rützler, D., 2015. Gender differences in the willingness to compete emerge early in life and persist. Manage. Sci. 61 (10), 2339–2354.
Sutter, M., Glätzle-Rützler, D., Balafoutas, L., Czermak, S., 2016. Cancelling out early age gender differences in competition: an analysis of policy interventions. Exp. Econ. 19 (2), 412–432.
Sutter, M., Kocher, M.G., 2007. Trust and trustworthiness across different age groups. Games Econ. Behav. 59 (2), 364–382.
Sutter, M., Kocher, M.G., Glätzle-Rützler, D., Trautmann, S.T., 2013. Impatience and uncertainty: experimental decisions predict adolescents' field behavior. Am. Econ. Rev. 103 (1), 510–531.
Sutter, M., Zoller, C., Glätzle-Rützler, D., 2019. Economic behavior of children and adolescents - a first survey of experimental economics results. Eur. Econ. Rev. 111, 98–121. doi:10.1016/j.euroecorev.2018.09.004.
Tymula, A., 2019. An experimental study of adolescent behavior under peer observation: adolescents are more impatient and inconsistent, not more risk–taking, when observed by peers. J. Econ. Behav. Organ. 166, 735–750.
Wang, Y., Sloan, F.A., 2018. Present bias and health. J. Risk Uncertain. 57 (2), 177–198.
Westfall, P.H., Young, S.S., Wright, S.P., 1993. On adjusting p-values for multiplicity. Biometrics 49 (3), 941–945.
Zhang, J., 2011. Do Girls in China Compete Just as Much as Boys? Evidence from an Experiment That Predicts Educational Choice. Technical Report. Citeseer.