Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Gender Differences in the Willingness to Compete Emerge Early in Life and Persist

Matthias Sutter, Daniela Glätzle-Rützler

To cite this article:
Matthias Sutter, Daniela Glätzle-Rützler (2015) Gender Differences in the Willingness to Compete Emerge Early in Life and Persist. Management Science 61(10):2339-23354. http://dx.doi.org/10.1287/mnsc.2014.1981

MANAGEMENT SCIENCE
Vol. 61, No. 10, October 2015, pp. 2339–2354
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.1981

Gender Differences in the Willingness to Compete Emerge Early in Life and Persist

Matthias Sutter
Department of Economics, University of Cologne, D-50923 Cologne, Germany; and Department of Public Finance, University of Innsbruck, A-6020 Innsbruck, Austria, matthias.sutter@uibk.ac.at

Daniela Glätzle-Rützler
Department of Public Finance, University of Innsbruck, A-6020 Innsbruck, Austria, daniela.ruetzler@uibk.ac.at

Gender differences in the willingness to compete have been identified as one important factor in explaining gender differences in labor markets and within organizations. We present three experiments with a total of 1,570 subjects, ages three to 18 years, to investigate the origins of this gender gap. In a between-subjects design we find that boys are more likely to compete than girls as early as kindergarten and that this gap prevails throughout adolescence. Re-examining the behavior of 316 subjects in a within-subjects design two years later, we show that these gender differences also largely persist over a longer time period and can thus be considered stable. Controlling for subjects' abilities in the different tasks, their risk attitudes, and expected performance, the gender gap in the willingness to compete is estimated in the range of 10–20 percentage points. We discuss the implications of our findings for policy interventions and organizational management.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2014.1981.

Keywords: competition; gender gap; experiment; children; teenagers; risk aversion
History: Received October 12, 2012; accepted March 10, 2014, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* October 31, 2014.

1. Introduction

A successful career in business, politics, or science does not only depend on an individual's ability and social skills, but also on a readiness to accept the challenge of competition for scarce jobs and rewards. However, recent research has provided ample evidence that women shy away from competition much more than men, a finding that is considered to be an important factor in explaining the persistent gender gap in wages and top-level positions in business, politics, or science (e.g., Gneezy et al. 2003, Gneezy et al. 2009, Gneezy and Rustichini 2004, Niederle and Vesterlund 2007 and 2010, Weichselbaumer and Winter-Ebmer 2007, Blau et al. 2010, Cason et al. 2010, Dohmen and Falk 2011, Datta Gupta et al. 2013).

Both in politics and within organizations, the gender gap in wages and top-level positions has been addressed by interventions in the spirit of affirmative action programs (Balafoutas and Sutter 2012, Miller and Segal 2012, Villeval 2012, Calsamiglia et al. 2013, Niederle et al. 2013). Such programs often provide incentives for women to compete to attract the best female talents for jobs. Although such programs make sense when addressing an existing imbalance between both genders in labor markets, they might fail to address the sources of the imbalance at their roots as long as it is not clear when one important component of gender differences in labor markets (i.e., the difference between men and women in their willingness to compete) emerges and whether such a difference may be persistent. For this reason, we examine two key issues in this paper.

First, we explore at which age gender differences in the willingness to compete emerge. Second, we check the persistence of a gender gap in the willingness to compete by presenting within-subjects evidence over an interval of two years. Addressing the latter issue is crucial for determining whether a competitive attitude is a rather stable personality trait that has a persistent influence on individual choices in competitive environments. The first issue is important to deal with to tailor the timing and type of policy interventions to prevent a gender gap altogether from emerging, or to close an existing gap already at an age when subjects can be targeted at presumably low costs. Achieving this aim would be beneficial for organizations in the medium and longer run. It would ensure that when hiring for upper-level positions, companies could expect the best men and women to apply, whereas existing evidence suggests that qualified female candidates may abstain from applications because of a reduced willingness to compete. For instance, a recent field experiment by Flory et al. (2015) shows that women do shy away from applying for a job if the job is

relatively more competitive. With respect to educational choices, recent papers by Zhang (2013) and Buser et al. (2014) show that subjects who are more likely to compete in experiments are more likely to choose a more demanding—and, in expected terms, a more profitable—education track than subjects who are less likely to compete. Buser et al. (2014), for instance, show that women choose the more demanding and rewarding education tracks less often, which correlates with their lower likelihood to compete in an experiment.

We present an experimental study with a total of 1,570 subjects, 853 of them aged three to eight years, and 717 aged nine to 18 years. Compared to previous studies with preadulthood samples (see below), this is the first paper to cover the entire preadulthood period from early childhood to late adolescence. In particular, this is the first paper to study gender differences in the willingness to compete before the age of seven and to incentivize the children in the task. This allows us to examine the possible emergence of a gender gap in choosing to compete as early as kindergarten age, a period in life where a human subject's cognitive and noncognitive skills are strongly shaped (Kail and Cavanaugh 2010) and when another important domain of economic decision making—other-regarding preferences—has been shown to develop (Fehr et al. 2008). Since other-regarding preferences and the willingness to compete are found to be correlated in adults—with inequality-averse subjects avoiding competition more often (Balafoutas et al. 2012)—and since women are, in general, more inequality-averse than men (Croson and Gneezy 2009), we hypothesize that gender differences in the willingness to compete arise around the same time as when inequality aversion develops, namely between ages three and eight.

Concerning the second issue examined in this paper—the persistence of attitudes toward competition—we have repeated one experiment with 316 teenagers two years after their first exposure to an experiment on competition. We are thus the first to study whether the choice to compete is intertemporally stable within subjects.[1] The 316 teenagers participating in our repetition were in 8th to 12th grade. After 9th grade, subjects may leave school and start working in Austria, and after 12th grade they either continue to college or university or start working. Hence, we have been able to track the preferences for competition at a time in life when teenagers may enter the labor market. If their preferences to compete would be highly volatile and thus unpredictable, policy interventions or affirmative action programs within companies would need to be viewed with large skepticism since preferences could easily change from one year to the next.

We find that gender differences in the willingness to compete emerge at kindergarten age when boys are more likely than girls to choose a competitive payment scheme rather than a noncompetitive one. We document such a gap (approximately 10–15 percentage points) from the age of three years onward for a presumably stereotypical male task (running 30 meters in the gym) and from the age of five years onward for a presumably stereotypical female manual task (picking items from a basket).[2] With regards to performance, there is no significant gender difference in running, and girls are significantly better-performing than boys in the manual task. Nevertheless, boys compete much more often than girls. The observed gap is robust to controlling for performance, for gender differences in risk attitudes, and children's expectations.

In the group of nine- to 18-year-olds, we then find a strong gender difference in the willingness to compete in a simple math task (adding two-digit numbers), where girls perform equally well as boys, but where boys choose a competitive payment scheme approximately twice as often as girls. Again controlling for risk attitudes, expected and actual performance, we estimate a gender gap of approximately 15–20 percentage points, and it is noteworthy (and worrisome) that the gap is even wider for the top-performing quartile of boys and girls. Moreover, the gender gap is still there when we let a subset of 316 teenagers repeat the math task two years later. We find that 77% of subjects stick with their initial choice, made two years earlier. For those who change their decision, girls are noticeably more likely to opt out of competition, whereas boys are approximately twice as likely as girls to opt in. Overall, these results indicate a high degree of persistent choices

[1] The issue of intertemporal stability of economic preferences has also recently attracted attention in different domains. Meier and Sprenger (2015) show that time preferences among approximately 1,400 adult subjects are reasonably stable over a period of two years. Volk et al. (2012) study the temporal stability of cooperation preferences. They find that in 66.4% of cases subjects do not change their type (conditional cooperator, free rider, or other) between two points in time. Dariel and Nikiforakis (2014), as well as Blanco et al. (2011), study the within-subjects stability of social preferences. Although Dariel and Nikiforakis (2014) find a strong correlation of prosocial behavior in the public goods and gift exchange game (i.e., subjects classified as cooperators reciprocate higher wages with higher levels of effort compared to noncooperators, who do not engage in gift exchange), Blanco et al. (2011) find less within-subjects stability of social preferences across four different experiments (ultimatum game, dictator game, sequential-move prisoner's dilemma, and public good game). Concerning the willingness to compete, Wozniak et al. (2014) let participants repeat an experiment on competition a few days later, finding a large degree of consistent choices. We are not aware of any study, however, that examines the persistence of competitive preferences beyond a period of only a few days.

[2] For adults, Günther et al. (2010) report a gender gap in the willingness to compete in a math task (interpreted as a stereotypical male task) but not in a word task (presumably stereotypically female). Wozniak et al. (2014), however, find a significantly larger willingness of men to compete both in a math and word task.

over a relatively long period of time, supporting the interpretation that attitudes toward competition are formed early in life and then remain stable for a large majority of subjects.

In the following, we discuss in §2 how our paper relates to the most closely connected work on gender differences in the willingness to compete. Then we present two experiments run with three- to eight-year-old children in §§3 and 4. These experiments are the first ones to examine the competitive attitudes of kindergarteners. In §5 we continue with the third experiment, involving nine- to 18-year-olds, which allows us to study whether the gender differences found in the first two experiments persist in adolescence. In §6 we report the results of repeating the third experiment with the same children two years later, enabling us to study whether competitive attitudes are stable over time within subjects. In §7 we discuss the main findings and implications and conclude the paper.

2. Related Literature

The seminal paper on the competitiveness of children is by Gneezy and Rustichini (2004). They have shown for a sample of nine- to 10-year-old Israeli children that boys increase their performance in a 40-meter sprint more when competing against another child—instead of running alone—than girls do.[3] Concerning the decision whether to compete (when there is an option to avoid it), Andersen et al. (2013) have examined gender differences in seven- to 15-year-old children and teenagers of a patriarchal tribe and a matrilineal tribe in Northeast India. The task was to throw tennis balls into a bucket. They found that boys compete more often than girls starting in the early teenage years. In our experiment, the gender gap emerges earlier, in kindergarten. The very different background of the subject pools considered in both studies—a sample of children in Austria, one of the richest countries in the European Union, versus two tribes in what many would consider a developing country—might explain the different time period when the gap emerges. However, once it has emerged, both studies—Andersen et al. (2013) and ours—share the same result: the gap is also found in older age groups.

Booth and Nolen (2012) examine the willingness to compete in 15-year-old British teenagers when solving mazes, finding, in general, boys compete more often than girls, which matches our results. Booth and Nolen note one exception of no gender differences, for girls attending single-sex schools. Yet, they hint at the possibility that the latter result may have been affected by self-selection issues. Cardenas et al. (2012) have compared the willingness to compete in a word search and a math task of Colombian and Swedish nine- to 12-year-old children. Although they have not been able to find a gender difference in the willingness to compete of Colombian children, they have found that Swedish boys compete significantly more often than Swedish girls in this group of nine- to 12-year-olds in both tasks. Dreber et al. (2014) replicate the gender gap for Swedish 16- to 18-year-olds in the math task, but not in the word search task. It remains an open question why these results are partly different from Cardenas et al. (2012) for Swedish nine- to 12-year-old children.

In developing countries, neither Zhang (2013) nor Khachatryan (2012) have found significant gender differences in the willingness to compete. Khachatryan (2012) has studied gender differences in the willingness to enter a competition in a word search and a math task with Armenian eight- to 16-year-olds, finding no significant difference in any of the age groups observed. Similarly, Zhang (2013) does not find a significant gender difference in a math task for 15-year-old Chinese subjects from a rural and poor area in Ninglang County.

So, the evidence seems to suggest that in highly developed countries one can typically observe a gender gap in the willingness to compete, whereas in developing countries this is not the case or the gap emerges only later. Almås et al. (2014) provide important insights that may explain such a cross-country pattern. They show in a sample of 14- to 15-year-old Norwegians that there exists a large and significant gender gap in the willingness to compete in a math task among adolescents who have parents with high education levels. However, there are no significant gender differences among adolescents with parents who have low levels of education, indicating that it is, perhaps, the gap in parental education that matters. Generally speaking, these findings suggest that the overall education level, which presumably influences economic development, may explain why gender differences in the willingness to compete seem nonexistent among children in less-developed countries like Colombia, India, Armenia, or rural China, or at least why they emerge at a later age compared to children in highly developed countries like Sweden, Norway, or Austria. In fact, for our math task (for part of our sample) we have information on the educational attainment of the participants' parents. Our data confirm the findings of Almås et al. (2014) with respect to the influence of

[3] Dreber et al. (2011) have not been able to replicate this gender difference in performance in the response to competition in a sample of eight- to 10-year-old Swedish children. They argue that one reason for the lack of a gender difference (in improving performance as a response to competition) might be the very egalitarian treatment of men and women in Swedish society, lending support to the general finding in Gneezy et al. (2009) that cultural norms are important for competitive attitudes.

parents' educational background on gender difference in the willingness to compete.

3. Experiment I: Running (3- to 8-Year-Olds)[4]

3.1. Experimental Design

A total of 412 children, 321 of them in kindergarten, ages three to six, and 91 in second grade of primary school, ages seven and eight, participated in this experiment (see Table 1). The experiment was run during regular physical education classes in the gyms of the respective schools or kindergartens, and it was always conducted by female experimenters. The verbal script for explaining the experiment is reproduced in Online Appendix A1. (All online appendices are available at http://homepage.uibk.ac.at/~c40421/children-competition_Appendix.pdf.) Our design is a slight modification of Gneezy and Rustichini (2004). In the first stage, children had to run alone a distance of 30 meters. They were timed and then split in two groups along the median time (with feedback given only at the end of the experiment). Children in the faster group got two rewards (e.g., stickers, sweets, marbles, pencils, etc.), and the other children received one reward at the end of the experiment. Please note that this reward scheme does not meet the properties of a piece-rate, but is based on relative performance already. Although most of the literature on the willingness to compete uses a piece-rate scheme as a benchmark, the results we find for the running task with respect to the willingness to compete are well-aligned with the related literature, and hence the stage 1 payment scheme does not seem to have influenced the main findings of this experiment.

After the first sprint of 30 meters, children were informed that there was a second stage of the experiment in which they had to run once more, this time choosing between two possible incentive schemes as follows.

- Scheme 1: If a child chose to run alone, that child received two rewards if faster than the median running time in the first round of running (stage 1); otherwise, that child received one reward.

[4] All experiments reported in this paper were approved by the State Board of Education of Tyrol, the Internal Review Board of the University of Innsbruck, and the headmasters of all involved schools and kindergartens. Teachers also consented to run the experiments during regular school hours or kindergarten hours, thus practically avoiding any kind of self-selection issues. All parents were informed in a letter about the project and its aim to study the economic decision making of children, without revealing any specific details. Parents had to consent to their child's participation, and only five parents (of more than 1,500) opted their child out. Children were also instructed that participation was completely voluntary, but all of them agreed to participate. The whole series of experiments was run between 2008 and 2011 in Tyrol, i.e., in the western part of Austria.

Table 1 **Number of Participants by Age and Gender in Experiment I (Running Task)**

	Kindergarten		Elementary school
Age (in years)	3/4	5/6	7/8 (2nd grade)
Female	64	101	44
Male	66	90	47
All ($N = 412$)	130	191	91

- Scheme 2: If the child chose to compete against one other child, the number of rewards (according to Scheme 1) was doubled if the child won the race, but the child received zero in the case of losing the race. If Scheme 2 was picked, the child had to run against another child who had also opted for Scheme 2 and had been classified in the same group according to the median split after the first sprint. Although choosing Scheme 1 has a competitive component, we argue that the choice of Scheme 2 reveals a preference for an even more competitive payment scheme.

Choices were elicited individually and privately. Children were called one by one to an experimenter, and the second stage of the experiment was explained to them. Thereafter they made their choice, were asked about their expectation to win the race,[5] and were sent to a separate room where all children who had already made their decision were waiting for the second stage to be performed. Before and after performing or making choices, children were instructed not to talk to each other, and they were kept busy playing games to avoid observing others running. All rewards, also those from the first sprint, were distributed at the end of the experiment.[6]

Research has shown that there are no differences in performance of boys and girls in short sprints before puberty (see, e.g., Papaiakovou et al. 2009 and also the results in Gneezy and Rustichini 2004). Hence, this task does not imply an advantage of three- to eight-year-old boys over girls of the same age with respect to objective performance. Nevertheless, the task might be perceived

[5] Note that we also asked those children who opted for Scheme 1 about their (hypothetical) expectation whether they would have won against another child if they had opted for Scheme 2. This was done to keep the conditions identical for all children. We did not incentivize the children's expectations to avoid the scenario in which a child could hedge by stating an expectation to lose and then run slowly to deliberately lose the competition and earn some reward for a correct expectation.

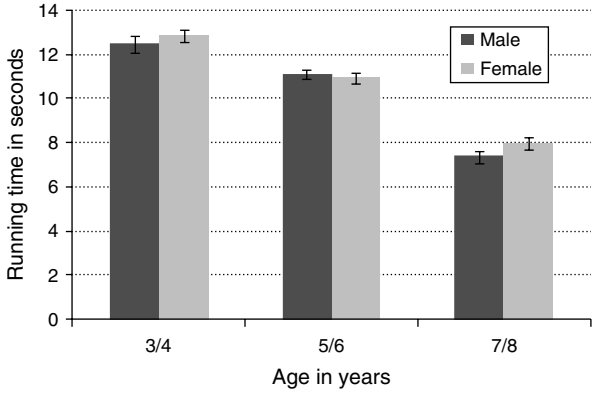
[6] When explaining the rules to children one-on-one, we always asked them to repeat the rules after the explanation (see the verbal script in Online Appendix A) to ensure understanding. We only report data for the 412 children who were able to repeat the rules and answer questions correctly. We had 40 children (34 of age three to four, and six of age five to six) who were not able to do so and who are excluded from the data analysis. Including these 40 children would not change any of the results reported here.

as stereotypically male, thus potentially favoring boys' likelihood to compete against others. To check for this possibility, we collected 96 teachers' expectations about whether boys or girls would be quicker in a 30-meter sprint. These teachers were not employed at any of the schools or kindergartens where we ran this experiment or the others, and none of them had heard about the project. We found, in fact, that teachers expected significantly better performance of boys by 1.41 seconds on average ($p < 0.001$ in single age groups, Wilcoxon signed-rank tests, two-sided, $N = 96$). Such perceptions might actually induce boys to compete more than girls in running, if teachers conveyed to young children—boys and girls alike—the impression that boys are better in this task, and it might discourage girls from competition.

3.2. Experimental Results

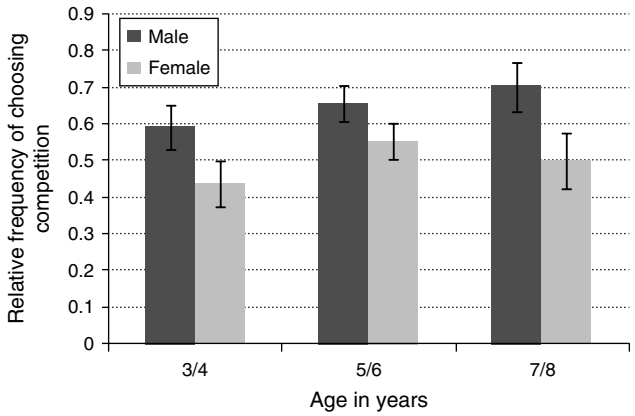
Figure 1(A) presents the average performance (measured in seconds) by boys and girls in the first sprint when children ran individually. The error bars in all figures indicate the mean ± standard error. In the group of seven- and eight-year-olds, girls are significantly slower, but if all data are pooled, there is no significant difference in running times between girls and boys (see Table B1 in Online Appendix B). Figure 1(B) demonstrates that, despite practically equal performance, boys choose to compete much more often than girls. In total, 64.5% of boys, but only 50.7% of girls, choose competitive payment Scheme 2 ($p = 0.005$, χ^2-test, two-sided, $N = 412$). The probit regressions in Table 2 estimate the gender gap in the willingness to compete at around 11–14 percentage points ($p < 0.05$).

Figure 1(A) Performance in Running in Experiment I ($N = 412$)



Note. Error bars indicate mean ± standard error.

Figure 1(B) Relative Frequency of Choosing the Competitive Payment Scheme in the Running Task in Experiment I ($N = 412$)



Note. Error bars indicate mean ± standard error.

Table 2 Probit Regressions on the Choice of Compensation Scheme in Experiment I

	Dependent variable: choice of competitive payment scheme	
Explanatory variables	1	2
Female (=1)	−0.139***	−0.114**
	(0.048)	(0.050)
Age	0.031*	0.044*
	(0.017)	(0.024)
Time in stage 1[a]		−0.010
		(0.011)
Expect to win in competition[b]		0.442***
		(0.059)
No. of observations	412	411[c]

Notes. The table presents the marginal effects of the coefficients. The results remain intact if we run logit regressions controlling for random effects on the school level. For instance, the marginal effect of "female" in model 2 would then be − 0.115 ($p < 0.05$).

[a]Number of seconds needed to complete the running task.

[b]This variable takes on the value 1 if a subject expects to be better than the other child in competition and 0 otherwise. (In case the child chose to run alone in the second stage, this question is hypothetical.)

[c]One child did not want to answer the expectation question.

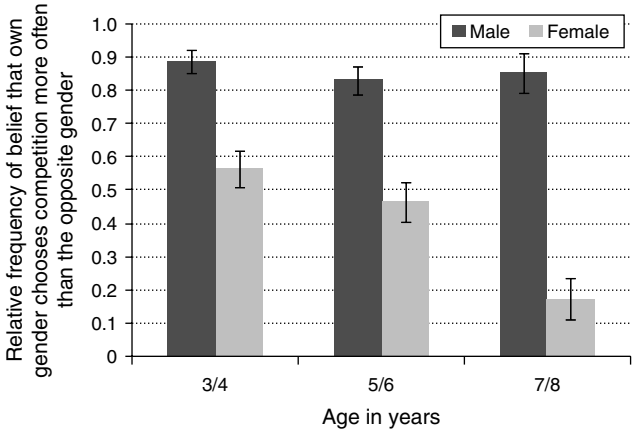
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively; robust standard errors in parentheses.

3.3. Do Children Expect a Gender Difference in the Willingness to Compete?

Approximately four weeks after the completion of this experiment, we asked 405 children about their expectations about the willingness of girls and boys to choose running in pairs (i.e., to compete) in the second sprint.[7] This part of our study was intended to check

[7] Note that 153 of these subjects had been participating in the running experiment, while 252 observations stem from new subjects. We were not able to return to all 405 subjects who had participated in the initial experiment (some kindergartens/schools did not have the temporal flexibility to allow us to visit them once more in the same academic year), for which reason we sampled 252 new children to get the same number of observations as in the experiment. In fact, the new subjects allow us to investigate whether the gender gap in expectations can also be found in the beliefs of those subjects who had not participated in the running experiment. In fact, this is what we find for these 252 children: for all age groups, there is a highly significant gender gap in beliefs (χ^2-test, $p < 0.02$ for each age group).

Figure 2 **Expectations of Children About Which Gender Chooses Competition More Often in the Running Task of Experiment I ($N = 405$)**



Note. Error bars indicate mean ± standard error.

whether children already at this very young age have a sense of both genders' willingness to compete. First we explained the task and the rules to each child and then asked the following question: "Consider there is a group of five girls and a group of five boys. Each of them has to decide whether to run in pairs for the second time or alone. What do you think: are more boys or more girls going to run in pairs?" Figure 2 shows that boys expect their own gender to choose competition significantly more often than girls expect from girls ($p < 0.001$ in each of the three age groups, χ^2-tests, two-sided).[8] These expectations are qualitatively well aligned with actual behavior, such that boys do compete more often, and they also fit with the teachers' expectations about boys performing better than girls in this task.

For this reason we were next looking for a task that potentially favored girls to see whether the gender gap in the willingness to compete might be reversed in such a task. Although a word task of unscrambling words is often used as stereotypically female (Günther et al. 2010, Wozniak et al. 2014), such a task was impossible to implement with kindergarten kids as young as three years old. After consultation with kindergarten teachers, we decided on the manual task described next.

4. Experiment II: A Manual Task (3- to 8-Year-Olds)

4.1. Experimental Design

This experiment was run with a total of 441 children, 367 of them in kindergarten, ages three to six, and 74 in second grade of primary school, ages seven and eight (see Table 3). The experimental task required subjects to pick all items with a cylindrical form (which were corks) from a basket with a plethora of different items. The verbal transcript of the explanation is included in §A2 of the online appendix.[9] The experiment was run in a separate room where only children who were working on the baskets were present. The children were given a one-minute time frame to pick all of the corks, and only the corks, from as many baskets as possible (see Figures A1 and A2 in Online Appendix A2). We put between three and eight corks into each basket so that children could not know how many corks to look for in a particular basket, since the number varied. The sequence of the number of corks in consecutive baskets was identical for all children. The experiment had two training phases of one minute each. After each training phase, the experimenter (always female) gave feedback on each basket and explained to the child whether everything was correct, and why if it was incorrect. After the two training phases, children were given the option to choose (in private and individually) between the following payment schemes, which were applied to a final one-minute time frame.

Table 3 **Number of Participants by Age and Gender in Experiment II (Manual Task: Picking Items from Baskets)**

	Kindergarten		Elementary school
Age (in years)	3/4	5/6	7/8 (2nd grade)
Female	70	104	40
Male	74	119	34
All ($N = 441$)	144	223	74

- Scheme 1: For each basket correctly worked on, the child got one reward, irrespective of the performance of other children.
- Scheme 2: For each basket correctly worked on, the child got two rewards if that child performed better than one other randomly paired child.[10] If that child performed worse, zero rewards per basket were given. In the case of a tie, the child had to repeat the task (for one more minute) until the tie was broken.

At the end of the experiment, and before any feedback was given, children were asked whether they expected to have performed better or worse than the other child in the second training phase. Payments (e.g.,

[8] We are grateful for a reviewer's comment that choices in the experiment might be correlated with expectations (such that children opting for the competitive scheme might expect their own gender to compete more often). However, our data do not support this conjecture. We find no significant correlation between expectations and own choice.

[9] Like in Experiment I, we let children repeat the rules of this experiment and answer questions to control for understanding. We restrict the analysis to the 441 children with correct understanding. There were another 46 children (40 of ages three to four, and six of ages five to six) whom we exclude from the analysis because they were not able to repeat the rules and answer the questions appropriately. Including these 46 subjects would not change any result reported here.

[10] Note that the other child didn't necessarily have to choose Scheme 2.

pencils, stickers, sweets, etc., with an exchange rate of one item per reward point) were collected at the end of the experiment in a different room. Children who had opted for payment Scheme 2 were only informed whether they had performed better or worse than the other child, and how many rewards resulted from this. Children who opted for payment Scheme 1 did not get any information about other children. Children who completed the experiment were brought back to their classroom and instructed not to talk to each other. This was also enforced by the teachers in the classroom.

Like for Experiment I, we asked the same 96 teachers about the expected average performance of boys and girls in this manual task. In this case, the teachers expected girls to perform better than boys, by an average 0.65 baskets across all age groups ($p < 0.001$ in single age groups, Wilcoxon signed-rank tests, two-sided). Obviously, this task is perceived as favoring girls and hence is biased against finding that boys are more likely to compete than girls.

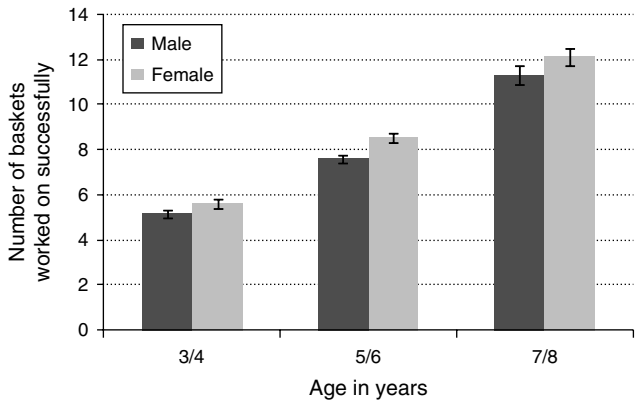
4.2. Experimental Results

Figure 3(A) presents the number of baskets correctly worked on by girls and boys in the second training phase before making their decision whether to compete. Across all age groups, girls perform significantly better by an estimated 0.7 baskets (see Table B2 in Online Appendix B). Nevertheless, Figure 3(B) demonstrates that girls shy away from competition more often than boys despite the fact that they perform better (and are expected to perform better by teachers). Across all cohorts, only 43.0% of girls, but 54.2% of boys, choose the competitive payment scheme ($p = 0.019$, χ^2-test, two-sided, $N = 441$). In the first age cohort of three- to four-year-olds, however, the difference in the willingness to compete is clearly insignificant (see Figure 3(B) and Table 4). This means that the gender gap in the willingness to compete emerges at approximately the age of five in this task and persists beyond this age, as is shown in the probit regressions in Table 4. There, the gender gap is estimated to be 14–16 percentage points. We control for a child's performance and expectation to win, and for risk aversion,[11] finding that more risk-averse subjects are significantly less likely to choose the competitive payment scheme.

4.3. Do Children Expect a Gender Difference in the Willingness to Compete?

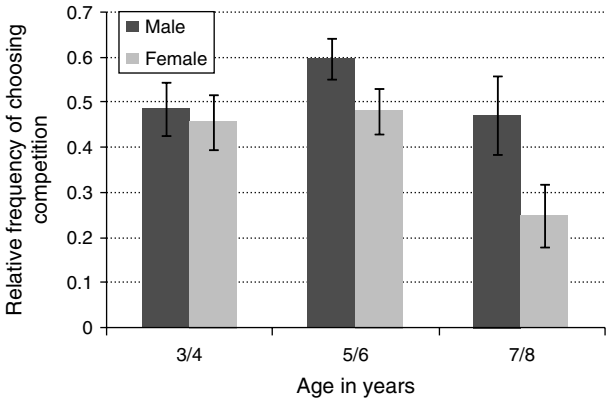
Approximately three weeks after the completion of this experiment, we asked 406 children about their expectations about the willingness of girls and boys to choose competition in this manual task.[12] Again, we first explained the task and the rules to each child and then asked the following question: "Consider there is a group of five girls and a group of five boys. Each of them has to decide whether to choose competition against another child. What do you think: are more

Figure 3(A) Performance in Picking Items from Baskets in Experiment II ($N = 441$)



Note. Error bars indicate mean ± standard error.

Figure 3(B) Relative Frequency of Choosing the Competitive Payment Scheme in Experiment II ($N = 441$)



Note. Error bars indicate mean ± standard error.

[11] Risk attitudes were measured (three to four days later) by letting children choose between receiving one present for sure or playing a risky game in which they could win two presents or zero with equal probability. We coded those children who played the risky game as risk taking, and the others as risk averse.

[12] In this case, 363 children had participated in the manual task experiment (since here we secured the opportunity to return to these children when preparing the agreement with kindergartens and schools for the experiment); 43 new children were also involved who had been absent (mostly due to illness) from the manual task experiment. We still let those 43 children take part in the follow-up survey. The number of new children is too small, however, to reliably infer whether the expectations data reported in Figure 3 would replicate in a pool of children who had not participated in the experiment. The expectations of those 363 children who participated in the experiment might have been shaped by talking with other children after the experiment about their choices. Their expectations might have also been influenced by a kind of false consensus effect of expecting others to behave similarly to oneself. The survey data for Experiment I in §3 suggest, however, that expectations do not necessarily differ among children who had participated in the experiment and those who had not.

Table 4 Probit Regressions on the Choice of Compensation Scheme in Experiment II

	Dependent variable: choice of competitive payment scheme					
	3/4 year olds		5/8 year olds		Pooled data	
Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Female (=1)	−0.026	−0.088	−0.144**	−0.160***	−0.110**	−0.143***
	(0.084)	(0.092)	(0.058)	(0.060)	(0.047)	(0.050)
Age	0.060	0.029	−0.057**	−0.099***	−0.021	−0.074***
	(0.097)	(0.117)	(0.025)	(0.034)	(0.017)	(0.028)
Number of baskets in training[a]		0.040		0.022		0.028**
		(0.029)		(0.015)		(0.013)
Expect to be better than the other child[b]		0.084		−0.012		0.008
		(0.114)		(0.065)		(0.056)
Risk aversion[c]		−0.242**		−0.196***		−0.211***
		(0.091)		(0.064)		(0.051)
No. of observations	144	135	297	295	441	430

Notes. The table presents the marginal effects of the coefficients. The results remain intact if we run logit regressions controlling for random effects on the school level, except that the age effect gets insignificant in model (6). The marginal effect of "female" in model (6), for instance, would be -0.138 ($p < 0.01$).

[a]Number of baskets correctly worked on in the second training phase.

[b]This variable takes on the value 1 if a subject expected to be better than the other child in the second training phase and 0 otherwise.

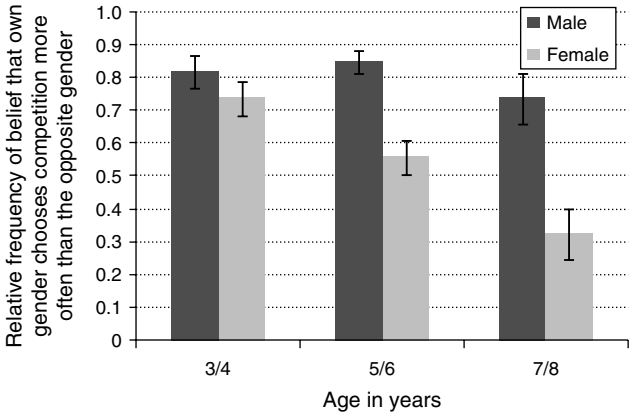
[c]Risk aversion was measured by letting subjects choose between receiving one present for sure and playing a risky game in which they can win two presents or zero with equal probability. Risk aversion is coded as 1 if the subject chooses the one definite present.

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively; robust standard errors in parentheses.

boys or more girls going to compete?" Figure 4 presents the results. For three- to four-year-olds we observe no gender difference. This means here that both girls and boys expect their own gender to be more likely to choose competition than the other gender and that both genders estimate roughly an equal fraction (of approximately 80%). This insignificance in expectations fits the results for three- to four-year-olds in actual behavior, where Figure 3(B) shows no gender difference either. However, for the two older age groups (five- and six-year-olds and seven- and eight-year-olds), expectations do differ between boys and girls. Girls expect their own gender to choose competition significantly less often than boys expect from boys ($p = 0.000$ in both cohorts, χ^2-tests, two-sided, $N = 190$ and $N = 74$, respectively). These results indicate that the gender gap in the willingness to compete does not only emerge in actual behavior around the age of five, but also that children's expectations about boys' and girls' competitiveness become different from this age on.[13]

The results presented thus far indicate that the gender gap in the willingness to compete emerges fairly early on in life, even in tasks where girls perform better and are expected to perform better. A necessary next step after having established this result is to check whether the gender gap is also found later in adolescence and whether it persists even within subjects across a longer time period.

Figure 4 Expectations of Children About Which Gender Chooses Competition More Often in the Manual Task of Experiment II ($N = 406$)



Note. Error bars indicate mean ± standard error.

5. Experiment III: A Math Task (9- to 18-Year-Olds)

5.1. Experimental Design

This experiment was run with 717 children and teenagers, ages 9 to 18 (see Table 5). The experiment was

[13] Please note that expectations and actual behavior in the experiment might be correlated, and this is partly the case. For three- to four-year-olds we do not find a significant correlation between expectations and own choice ($p > 0.1$, χ^2-test), but for the two older age groups the correlation is significant ($p < 0.01$, χ^2-tests).

Table 5 Distribution of Participants by Age and Gender in Experiment III (Math Task: Adding Two-Digit Numbers)

	Elementary school	Grammar school[a]			
Age in years	9/10	11/12	13/14	15/16	17/18
Grade	4th	6th	8th	10th	12th
Female	42	129	100	77	71
Male	51	79	66	51	51
All ($N = 717$)	93	208	166	128	122

Notes. We needed groups of four members in each age group. In those age groups where the number of participants is not a multiple of four, we had one group with fewer members, and we filled this group up by using another subject's decisions twice (without paying this subject twice, of course). We exclude eight observations with missing variables (these subjects were absent for the risk-aversion experiment) to have the same number of observations throughout the analysis of Experiment III.

[a]One of the four grammar schools was a girls-only school, and 88 subjects attended this school. We checked in the regression in Table 6 whether behavior in the girls-only school was different from the other schools but found no difference, allowing us to pool the data.

computerized (using z-Tree, Fischbacher 2007), and was run with notebooks in the students' classroom (during regular school hours, but only very rarely when students would have had their math course). The two authors (male and female) ran all experimental sessions. The experimental design is mainly taken from Niederle and Vesterlund (2007). The task was to add as many sets as possible of three two-digit numbers within two minutes. At the beginning of the experiment, subjects were randomly assigned to groups of four persons. The group assignment was fixed for all stages of the experiment. Each stage was only explained after the previous one had been finished. The instructions were explained verbally, following a fixed written script (see Online Appendix A3). Before the experiment started, all participants were given enough time to acquaint themselves with the software. At the end of each stage, participants were only informed about the number of calculations they had worked on and how many of them had been correct. The sequence of stages was as follows.

In stage 1 (piece-rate), each subject received €0.50 for each correct calculation. In stage 2 (competition), the four group members competed against each other. The group member who solved the most calculations correctly was paid €2.00 per calculation. The other three group members received nothing. Ties were broken randomly. Yet, there was no feedback given concerning the tournament outcome until the very end of the experiment. In stage 3 (choice), subjects chose privately between the following payment schemes before performing the same task as in stages 1 and 2.

- Scheme 1: Here the piece-rate scheme of stage 1 applied; hence, no competition was involved.
- Scheme 2: Here the competitive payment scheme of stage 2 applied. If Scheme 2 was chosen, a subject's performance in stage 3 was compared to the other group members' performances in stage 2. This method has several advantages. First, competition entry decisions do not depend on a subject's expectations about the other members' entry decisions. Second, stage 2 performances are competitive performances, and thus a subject competes against others when they were also exposed to a competitive payment scheme. Third, entering competition does not impose externalities on others. In principle, this means that stage 3 is an individual decision-making problem.[14]

There was a fourth stage that served as a control condition for stage 3 (which is the crucial stage). In stage 4, subjects did not have to perform the experimental task, but were only asked to submit their performance in stage 1 either to a piece-rate or competitive payment scheme.

At the end of the experiment, and before any feedback was given, subjects were asked about their expected rank within their group in stage 2 (with a rank of 1 indicating best performance). This allows us to control for the role of expectations about one's own performance in explaining the choice to compete. The actual performance in stage 1 can be used as a control for a subject's ability.

One stage was randomly selected for payment (to avoid income effects from cumulative earnings). Expectations were also incentivized and correctly guessed ranks yielded €0.50. Payoffs (which were on average €6.58) were distributed in sealed envelopes (with anonymized codes) several days after the experiment.

Motivated by the findings in Gneezy et al. (2003) that women might compete differently when competing only against other women, we implemented a between-subjects design with two treatments. In the *mixed* treatment, subjects were told that each group would consist of two boys and two girls. In the *single-sex* treatment, they were told that each group consisted of four members of their own gender. In fact, we do not find significant differences with respect to the choice of payment schemes in stage 3 between both treatments, for which reason we pool the data in the following. (Table B3 in Online Appendix B provides evidence that pooling is unproblematic.)

Concerning the expected performance of boys and girls, it is noteworthy that, in general, there are no

[14] This methodological approach is different from Experiment I where competition meant running side by side. We used this procedure here because children are used to running alongside each other (much more so than if we had told them that they would run alone and that their times would be compared with others'). In Experiment II, competition was against one other randomly paired child. Since in Experiment II we had not had a forced competition stage (but only one single stage that counted for payment), the approach of Experiment III was impossible to implement. Despite these slight methodological differences, the overall pattern of results is strikingly similar across all experiments.

gender differences in such easy math tasks as those used in this experiment (Hyde et al. 2008). This makes us unlikely to find a priori gender differences in the willingness to compete if the willingness to compete depends primarily on performance. As for Experiments I and II, we elicited the expected performance of boys and girls from the 96 teachers. For nine- to 14-year-olds, they expected practically identical performance of both genders (with 0.01 more calculations expected from boys on average; $p > 0.6$ in single age groups, Wilcoxon signed-rank tests, two-sided). Hence, the teachers judged this task as gender neutral for these age groups. Only for 15- to 18-year-olds did the teachers expect significantly better performance of boys (by approximately 0.6 calculations on average; $p < 0.05$ in single age groups, Wilcoxon signed-rank tests, two-sided).

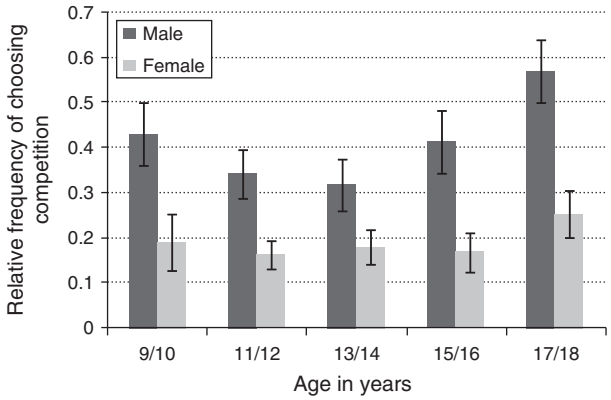
5.2. Experimental Results

Figure 5(A) presents the average performance of boys and girls in the noncompetitive stage 1, conditional on age. There are only very small, and always insignificant, differences between boys and girls (see also Table B4 in Online Appendix B). Nevertheless, there is a very strong difference in the relative frequency of boys and girls to choose the competitive payment scheme in stage 3, as Figure 5(B) illustrates. Across all age groups, 40.3% of boys, but only 18.6% of girls, are willing to compete ($p = 0.000$, χ^2-test, $N = 717$), and the difference is significant in each single age group ($p < 0.05$, χ^2-tests). The probit regressions in Table 6 show that the likelihood to choose the competitive payment scheme is significantly higher for boys. Model (2) in Table 6 controls for a subject's expectation to win in stage 2 (i.e., guessing a rank of 1), finding that those who expect to win are approximately 33 percentage points more likely to compete. Furthermore, better performance in stage 2 (number of correct calculations) is marginally positive. We also include a measure for risk aversion,[15] finding that more risk-averse subjects are less likely to compete. In Online Appendix B we show in Tables B5 and B6 that girls are significantly less optimistic in their expected ranks than boys (with an estimated difference of 0.38 ranks) and that girls are significantly more risk averse than boys. Despite these controls, the gender gap in the willingness to compete is estimated at approximately 17 percentage points, dropping only slightly from the estimated 22 percentage points without controls.[16]

Figure 5(A) Performance in the Math Task of Adding Two-Digit Numbers in Experiment III ($N = 717$)

Note. Error bars indicate mean ± standard error.

Figure 5(B) Relative Frequency of Choosing Competition in Experiment III ($N = 717$)



Note. Error bars indicate mean ± standard error.

Figure 6 addresses the gender gap in the willingness to compete from yet another perspective, by looking at the rate with which boys and girls in the four quartiles (of performance in stage 2) choose the competitive payment scheme in stage 3. We see that the gender gap is largest for the top-performing quartile (on the right side of Figure 6), with 40 percentage points difference (see also Table B8 in Online Appendix B). This is a particularly striking (and worrisome) result as it indicates that the best-performing girls do not close the gap (which could be the case because they are performing so well), but compared with the poorer-performing girls the gap gets even wider.

[15] The risk preferences of subjects were determined in a different experiment (approximately one year earlier) by letting them choose between a lottery and an increasing sure amount of money. The lottery yielded a prize $\pi > 0$ or zero with a 50:50 chance, and subjects had to indicate in a choice list whether they preferred the lottery or a safe amount that increased from $\pi/20$ to π in 20 steps of width $\pi/20$. From this we determined the interval including the certainty equivalent, took the midpoint of this interval, and then calculated our measure of risk aversion as $1 -$ (certainty equivalent/prize π). See Sutter et al. (2013) for more details on the risk experiment.

[16] It is worth noting that the performance of girls in stage 3 is *not* worse than boy's performance (see Table B7 in Online Appendix B). Rather, in some specifications, girls perform significantly better than boys in stage 3. In line with this result, we also note that there are no gender differences in the reaction to competition, measured by the relative increase of a subject's performance from stage 1 to 2.

Table 6 Probit Regressions on the Choice of Compensation Scheme in Experiment III

	Dependent variable: choice of competition in stage 3	
Explanatory variables	(1)	(2)
Female (= 1)	−0.219***	−0.165***
	(0.034)	(0.036)
Age	0.013**	0.008
	(0.006)	(0.007)
Number of correct calculations in stage 2		0.010*
		(0.005)
Expect to win in stage 2 (= 1)		0.331***
		(0.047)
Risk aversion		−0.239***
		(0.070)
No. of observations	717	717

Notes. The table presents the marginal effects of the coefficients. The results remain intact if we run logit regressions controlling for random effects on the session level. The marginal effect of "female" in model (2), for instance, would be − 0.166 ($p < 0.01$).

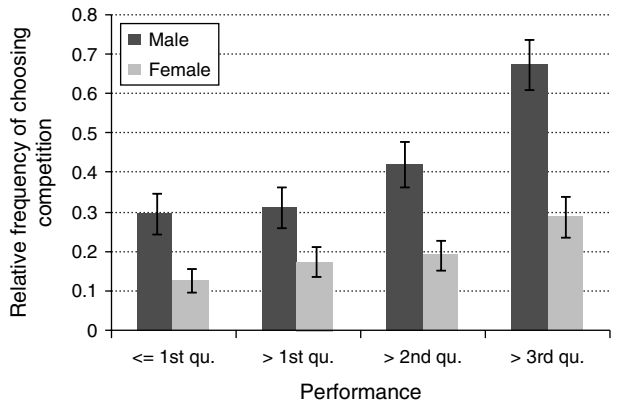
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively; robust standard errors in parentheses.

Since we have information on a child's number of siblings, we ran additional regressions where we include birth order, number of siblings, and a dummy for being the only child. We find that later-born children are more likely to enter the competition ($p = 0.035$) and that having more siblings is weakly significant ($p = 0.098$), whereas being an only child is not significant (see Table B9 in Online Appendix B). For Experiments I and II we also have information on these variables, but none turned out to be significant. For Experiment III we also have self-reported math grades for the previous year. Calculating for each child the relative math grade in comparison with the class average and adding this variable, we find that it is weakly significantly correlated with the willingness to enter the competition, meaning that children with better-than-average math grades are more likely to compete ($p = 0.087$, $N = 717$).

In Table B10 (see Online Appendix B) we confirm for a subsample of participants the earlier results of Almås et al. (2014) that the gender gap in the willingness to compete is influenced by the educational background of parents. Many parents filled in a voluntary questionnaire where we asked about the father's and mother's highest degree. In the group of students whose father has an above-median highest degree (i.e., a university degree), the gender gap is highly significant and 32 percentage points large (probit regression controlling for gender, age, performance in stage 2, confidence, and risk aversion, $p = 0.001$). For students with fathers with a below-median highest degree, the gender gap is insignificant (probit regression, $p = 0.291$). Taking the mothers' educational levels, we find a significant gender gap both for those students with mothers of above-median level (i.e., university entry qualification) ($p = 0.072$) and those with mothers of below-median level ($p = 0.024$).

The data for stage 4—in which subjects did not have to do the math task but simply submitted their stage 1 performance to piece-rate or a competitive scheme—reveal basically the same patterns as stage 3. In Table B11 in Online Appendix B, we provide an estimation result similar to the regression in Table 6, showing that girls are estimated to be 10 percentage points less likely to choose the competitive payment scheme. Given the similarity to the results discussed already in this section, we dispense with a more detailed description of stage 4 results.

Figure 6 Relative Frequency of Choosing Competition in Experiment III, Separated by Quartiles of Performance in Stage 2 ($N = 717$)



Note. Error bars indicate mean ± standard error.

6. Experiment IV: Repeating the Math Task Two Years Later

6.1. Experimental Design

Almost exactly two years after running Experiment III, we were able to return to a subset of 316 subjects to repeat the same experiment. These repeat participants were then in 8th, 10th, or 12th grade (see Table 7). The task was identical (adding sets of two-digit numbers), and the sequence of stages was also exactly as described previously in §5. We were able to match a subject's decisions in Experiment III and in Experiment IV

Table 7 Number of Participants by Age and Gender in Experiment IV (Repeating the Math Task)

	Grammar school		
Age (in years)	13/14	15/16	17/18
Grade	8th	10th	12th
Female	88	36	47
Male	73	27	45
All ($N = 316$)	161	63	92

because each child was uniquely identified with an eight-digit code that did not change across experiments.

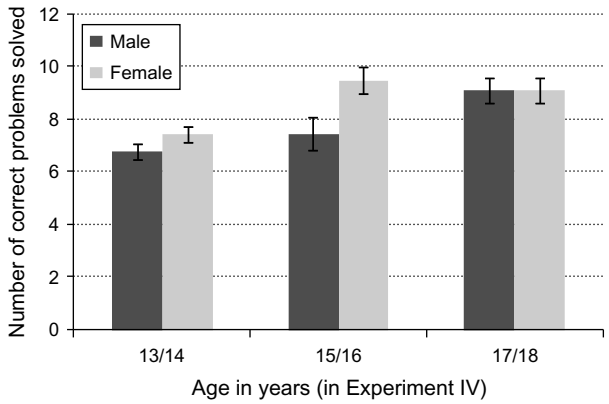
We could not return to the subjects who were in grades 4 or 12 in Experiment III, because after grade 4 children change schools in Austria, and after grade 12 they leave school (and continue to university or start working). Of the children who were in grades 6, 8, or 10 initially in Experiment III, we also lost most of those in grade 8, because in Austria after this grade approximately 60% of children typically change school once more, leaving us in the end with 316 subjects to repeat the experiment.

We can check whether those children who were initially in grades 6, 8, or 10 and repeated the experiment two years later differ in some notable respects from those who changed or quit school. Most importantly, for each of grades 6, 8, and 10 in Experiment III, and separated for each gender, we do not find a different propensity in the willingness to compete in stage 3 of Experiment III between children who repeated the experiment two years later and those who didn't (χ^2-tests, $p > 0.23$ in any case). There is one significant difference (of six comparisons) between subjects who stay in the subject pool of Experiment IV and those who don't, and this refers to the number of correct solutions in stage 1 of Experiment III: girls who were in grade 10 in Experiment III and continued in Experiment IV perform better than girls in grade 10 who drop out later ($p < 0.05$, Mann–Whitney U-test; all other comparisons fail significance). Moreover, for risk aversion, we do not find significant differences between the two groups for any of the three age groups (Mann–Whitney U-test, $p > 0.2$ for all age groups). Concerning their self-reported math grade (for the year prior to running Experiment III), for the first and second age groups (11- to 12-year-olds and 13- to 14-year-olds in Experiment III) we do not find significant differences between both groups (11- to 12-year-olds in Experiment III, $p = 0.210$; 13- to 14-year-olds, $p = 0.257$). For the third age group (15- to 16-year-olds in Experiment III), we find that those subjects who change or quit school (those not participating in Experiment IV) have significantly worse math grades ($p = 0.026$). Hence, the remaining subjects for Experiment IV have presumably somewhat better cognitive skills in this age group.

6.2. Experimental Results

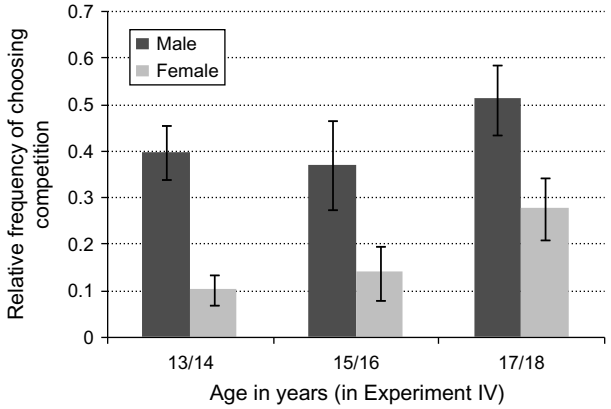
Like in Experiment III, the performance of girls in stage 1 of Experiment IV is not significantly worse than the performance of boys, as can be seen in Figure 7(A). In fact, 13- to 14-year-old and 15- to 16-year-old girls perform significantly better than boys. However, in Experiment IV there is also a strong gender difference in the frequency of choosing the competitive payment scheme in stage 3. Across all age groups, girls enter competition in 15.8% of cases, but boys do so in 42.8%

Figure 7(A) Performance in the Math Task of Adding Two-Digit Numbers in Experiment IV ($N = 316$)



Note. Error bars indicate mean ± standard error.

Figure 7(B) Relative Frequency of Choosing Competition in Experiment IV ($N = 316$)



Note. Error bars indicate mean ± standard error.

of cases ($p < 0.001$, χ^2-test, $N = 316$) (the difference is significant in each single age group, $p < 0.05$, χ^2-tests). Figure 7(B) presents the entry frequencies for each age group separately. So far, the evidence confirms that, in the aggregate, the data patterns look very similar across an interval of two years. In the following, we analyze the data on an individual level.

First, we look at the relative frequency with which subjects made the same decision in stage 3, both in Experiment III and IV. These are 243 of 316 subjects (77%), which is remarkably large, and a significantly larger fraction than if choices had been random ($p < 0.001$, binomial test, $N = 316$; see also Table 8). Girls make the same choice in both experiments in 80% of cases, and boys in 74%, with no significant gender difference in this respect ($p > 0.2$, χ^2-test, $N = 316$). Although the large majority of teenagers obviously has a stable preference, it is interesting to look also at those 23% of subjects who choose once the competitive and once the noncompetitive scheme. Here we find strong gender differences again. Of those 234 subjects who had

Table 8 Relative Frequency with Which Subjects Make Same Choice in Stage 3 of Experiment III and Experiment IV (Two Years Later)

Age in Experiment III Age in Experiment IV	Overall	11/12 13/14	13/14 15/16	15/16 17/18
Female (%)	80	83	81	72
Male (%)	74	73	74	76
Number of subjects	316	161	63	92

not chosen the competitive scheme in Experiment III, only 11% of girls then chose the competitive scheme in Experiment IV, compared to 26% of boys ($p < 0.01$, χ^2-test). Hence, relatively more boys than girls opt into the competitive scheme two years later. The mirror image of this finding is that relatively more girls than boys opt out of competition two years later. Of 82 subjects who had chosen the competitive scheme in Experiment III, 27% of boys prefer the noncompetitive scheme in Experiment IV, compared to 63% of girls ($p = 0.001$, χ^2-test).

In summary, the data pattern shows that competitive preferences persist in a large majority of subjects over two years. If subjects change their preferences, girls are relatively more likely to opt out of competition and less likely to opt in than boys. This overall pattern can also be seen in the probit regression reported in Table 9. It is, in principle, the same regression that has been used in Table 6 for Experiment III, but it uses as the dependent variable the choice of the competitive payment scheme in Experiment IV and adds as independent variables a subject's choice in Experiment III and the fraction of classmates of one's own gender who chose to compete in Experiment III. The latter variable is intended to capture potential peer effects. We can see that the decision to enter competition in Experiment III two years ago increases the likelihood of choosing the competitive payment scheme in Experiment IV by approximately 30–40 percentage points. Consistent with the previous analysis, the female dummy is significantly negative, showing that girls are significantly less likely to choose competition even when controlling for their previous choices, meaning that girls are much more likely to opt out of competition in Experiment IV, even if they competed in Experiment III. We also see that the performance in stage 2 (positive), expectations to win (positive), and risk aversion (negative) have the same effects on the willingness to compete as discussed in Experiment III.[17] We find no evidence of peer effects.

[17] When receiving their payments from Experiment III, students also got feedback about their performances in all stages, from which they potentially might have inferred their relative performance and thus made their decision to compete in Experiment IV contingent on this inference. When adding the true relative performance (within one's class) to the regression in Table 9, this variable is not significant, indicating that feedback most likely had a very minor impact on choices two years later.

Table 9 Probit Regressions on the Choice of Compensation Scheme in Experiment IV

	Dependent variable: choice of competition in stage 3	
Explanatory variables	(1)	(2)
Choice of competitive scheme in Experiment III	0.377***	0.313***
	(0.062)	(0.069)
Female	−0.220***	−0.161***
	(0.052)	(0.059)
Age	0.027*	0.018
	(0.015)	(0.017)
Number of correct calculations in stage 2		0.018*
		(0.011)
Expect to win in stage 2 (= 1)		0.231***
		(0.082)
Risk aversion[a]		−0.560***
		(0.152)
No. of observations	316	287

Notes. The table presents the marginal effects of the coefficients. If we add to model (2) as independent variable the fraction of classmates of one's own gender who competed in stage 3 of Experiment III (two years earlier), this variable is insignificant and does not affect the order of magnitude and the level of significance of the remaining independent variables. Hence, peer effects within one's own gender do not seem to play an important role. The results remain intact if we run logit regressions controlling for random effects on the session level, except that the variable "number of correct calculations in stage 2" gets insignificant in model (2). The marginal effect of "female" in model (2), for instance, would be − 0.164 ($p < 0.01$).

[a]The risk preferences of subjects were determined approximately two months prior to Experiment IV with the same procedure as described in Footnote 15 in §5.

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively; robust standard errors in parentheses.

7. Discussion and Conclusion

In this paper we have studied the willingness to compete of 1,570 children and teenagers, ages three to 18. We let boys and girls choose between a competitive and noncompetitive payment scheme in three different tasks. In two of them (running and adding two-digit numbers), we observed that boys and girls performed equally well; in the third one (a manual task of picking items from a basket), girls performed significantly better than boys. Despite the fact that girls were obviously not performing worse than boys in any task, we found that boys chose the competitive payment scheme significantly more often in each single task. Controlling for performance, risk attitudes, and expectations to win, the gender gap in the frequency of choosing the competitive payment scheme was estimated to lie in the range of 10–20 percentage points, with the gap fluctuating somewhat across childhood and adolescence, showing no clear directional development with age.

These results are important as they document that already at an early age (i.e., in kindergarten), there are significant differences in the willingness of boys and girls to expose themselves to competition. Since an

ability to cope with competition is vital for a successful career in business, the early formation of a gender gap in the willingness to compete may have important long-term consequences for unequal outcomes of men and women in labor markets (Blau et al. 2010).

We consider three aspects of our study as particularly novel. First, we have been the first to examine competitive attitudes in kindergarten age. All previous studies had children of at least seven years of age (Booth and Nolen 2012, Cardenas et al. 2012, Andersen et al. 2013, Almås et al. 2014). Our findings show a gender gap as early as kindergarten. Only for three- to four-year-old girls and boys in the manual task did we not find a significant difference. This latter result is a first hint that early-age socialization may have an important impact on the emergence of the gender gap in competitiveness.[18] If the willingness to compete—and the differences between boys and girls therein—were solely determined by genetic factors, then we should not have been able to find a task wherein the youngest participants showed no gender difference in the likelihood to choose competition. Of course, the interplay between environmental and biological factors is far from settled (and it would have been beyond the scope of this paper to resolve it), but future research might devote resources to addressing it in more detail.[19]

Second, we have covered the whole pre-adulthood period from ages three to 18. Previous studies have had considerably less variation in age. Yet, the overall picture emerging from the literature and our study is pretty consistent. Once the gender gap in the willingness to compete has emerged, it does not seem to vanish later in life, although the size of the gap varies across age. The study of Andersen et al. (2013) found that in poorly developed regions in India the gender gap emerges in the early teenage years and persists thereafter in older cohorts. Although the cultural and economic background of both studies is very different—one of the richest countries in the European Union versus two tribes in what many would consider a developing country—and although the different background may explain why the age when the gender gap emerges is different, both Andersen et al. (2013) and our study share the insight that the gap does not close or vanish in older cohorts. Our findings are also consistent with results of Booth and Nolen (2012) for 15- to 16-year-old British children, those of Cardenas et al. (2012) for nine- to 12-year-old Swedish children, and those of Almås et al. (2014) for Norwegian adolescents. These studies have found (with different tasks) that girls are less likely to compete than boys. So far, the results from developing countries seem less supportive of a gender gap in the willingness to compete. Cardenas et al. (2012) in Colombia, Zhang (2013) in a poor region of China, and Khachatryan (2012) in Armenia do not find gender differences in the willingness to compete. It is beyond the scope of this paper to determine why this is the case. Recent work of Almås et al. (2014) suggests that educational levels might play a role, as we have also found in our Experiment III, but more work is certainly needed here.

Third, to the best of our knowledge, this is the first paper to examine whether competitive attitudes are persistent *within* subjects over a longer time period. By repeating the experiment with the math task two years later, we have found that 77% of subjects stick with their initial choices, meaning that a large majority of subjects do not change their preferences toward competition. Of those who do, girls are much more likely to opt out of competition than boys, and boys are significantly more likely to opt in. Unfortunately, we have not been able to repeat the first two experiments one or two years later with the same children to study whether such choices are already persistent in the first 10 years of life, which would provide further evidence about the optimal timing of policy interventions. Nevertheless, given the very high degree of persistent choices of teenagers, an important question for future research seems to be how competitive attitudes could be influenced and which interventions might work to induce (in particular very well-performing) women to choose competition as often as men do.

It has been shown that different forms of affirmative action programs (like introducing quotas for female winners or giving women a head start in competition) induce adult women to compete more and that such programs do not seem to have noticeable efficiency costs (e.g., of passing by better-qualified men) because they encourage, in particular, highly-qualified women to enter the competition (Balafoutas and Sutter 2012, Niederle et al. 2013). Although such programs are obviously successful in bridging an existing gap, it seems more practical—and potentially also more (cost-)

[18] In fact, from developmental psychology it is known that parental socialization of toys appropriate to play with in the first three years of life generates common gender stereotypes of toys both for boys and girls, indicating that parental socialization forms children's understanding of gender during the preschool and pre-kindergarten years (see Eisenberg et al. 1985, Freeman 2007). It is conceivable that the different toys and games for boys and girls are also associated with different degrees of competitiveness, supporting more competitive behavior of boys and less competitive behavior of girls (Weinberger and Stein 2008).

[19] A recent strand of literature has examined whether hormonal factors contribute to the gender gap in the willingness to compete. Unfortunately, the seminal studies by Buser (2012) and Wozniak et al. (2014) report largely contradictory results concerning the influence of hormones on women's likelihood to compete. Although some of the divergent results may have been driven by design differences, it seems that more research is needed to establish a common understanding of how hormones contribute to the gender gap in the willingness to compete.

effective—to close the gap right from the start, i.e., from early in life if it has been shown for a society that the gap already exists in young children, rather than reaching only adults with different forms of policy intervention. Our results show that interventions early in life might be important to bridge the gap. Which interventions work (best) is still poorly understood and this needs to be investigated more rigorously in the future. In doing so, two issues will need to be examined very carefully. First, does any intervention have long-term effects in influencing the gender gap in the willingness to compete? Second, and importantly, it is necessary to compare the potential costs of interventions with their potential benefits, an aspect that the literature on gender differences in the willingness to compete has largely neglected so far but that ought to be examined in future work.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.1981.

Acknowledgments

The authors thank the associate editor and three reviewers, as well as Loukas Balafoutas, Björn Bartling, Gary Charness, Ernst Fehr, Uri Gneezy, Magnus Johannesson, Martin Kocher, John List, Muriel Niederle, Pedro Rey-Biel, Bettina Rockenbach, Daniel Schunk, Karl Sigmund, Lise Vesterlund, Marie-Claire Villeval, and seminar participants at the University of Arizona in Tucson; University of California at Santa Barbara; the Universities of Erfurt, Gothenburg, Konstanz, and Munich; the Stockholm School of Economics; INSEAD Business School in Fontainebleau; Max Planck Institute of Economics in Jena; and Economic Science Association meetings in Innsbruck and Chicago for helpful comments. The authors are particularly grateful to Thomas Plankensteiner from the State Board of Education in Tyrol (*Landesschulrat für Tirol*); Michaela Hutz, Barbara Raithmayr, and Martina Zabernig from the Board of Tyrolean Kindergartens; and the headmasters Peter Arnold, Karin Dürnberger, Hildegard Flöck, Brigitte Fröhlich, Maria Frotschnig, Max Gnigler, Sigmund Heel, Gottfried Heiss, Christine Kager, Ulrike Künstle, Hermann Lergetporer, Isabella Mölk, Claudia Pertl, Bernhard Schretter, and Peter Paul Steinringer for making this study possible. Financial support was provided by the Austrian Science Foundation [Project P22772-G11] and the Austrian Central Bank [Jubiläumsfonds Projekte 12588 and 14680].

References

Almås I, Cappelen AW, Salvanes KG, Sørensen EØ, Tungodden B (2014) Willingness to compete: Family matters. Working paper, Norwegian School of Economics (NHH), Bergen, Norway.

Andersen S, Ertac S, Gneezy U, List JA, Maximiano S (2013) Gender, competitiveness and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Rev. Econom. Statist.* 95:1438–1443.

Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the lab. *Science* 335:579–582.

Balafoutas L, Kerschbamer R, Sutter M (2012) Distributional preferences and competitive behavior. *J. Econom. Behav. Organ.* 83:125–135.

Blanco M, Engelmann D, Normann H-T (2011) A within-subject analysis of other-regarding preferences. *Games Econom. Behav.* 72:321–338.

Blau FD, Ferber MA, Winkler AE (2010) *The Economics of Women, Men, and Work*, 6th ed. (Prentice Hall, Englewood Cliffs, NJ).

Booth AL, Nolen PJ (2012) Choosing to compete: How different are girls and boys? *J. Econom. Behav. Organ.* 81:542–555.

Buser T (2012) The impact of the menstrual cycle and hormonal contraceptives on competitiveness. *J. Econom. Behav. Organ.* 83:1–10.

Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness and career choices. *Quart. J. Econom.* 129:1409–1447.

Calsamiglia C, Franke J, Rey-Biel P (2013) The incentive effects of affirmative action in a real-effort tournament. *J. Public Econom.* 98:15–31.

Cardenas J-C, Dreber A, von Essen E, Ranehill E (2012) Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *J. Econom. Behav. Organ.* 83:11–23.

Cason TN, Masters WA, Sheremeta RM (2010) Entry into winner-take-all and proportional-prize contests: An experimental study. *J. Public Econom.* 94:604–611.

Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Literature* 47:448–474.

Dariel A, Nikiforakis N (2014) Cooperators and reciprocators: A within-subject analysis of pro-social behavior. *Econom. Lett.* 122:163–166.

Datta Gupta N, Poulsen A, Villeval M-C (2013) Gender matching and competitiveness: Experimental evidence. *Econom. Inquiry* 51:816–835.

Dohmen T, Falk A (2011) Performance pay and multi-dimensional sorting—Productivity, preferences and gender. *Amer. Econom. Rev.* 111:556–590.

Dreber A, von Essen E, Ranehill E (2011) Outrunning the gender gap—Boys and girls compete equally. *Experiment. Econom.* 14: 567–582.

Dreber A, von Essen E, Ranehill E (2014) Gender and competition in adolescence: Task matters. *Experiment. Econom.* 17:154–172.

Eisenberg N, Wolchik SA, Hernandez R, Pasternack JF (1985) Parental socialization of young children's play: A short-term longitudinal study. *Child Development* 56:1506–1513.

Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454:1079–1084.

Fischbacher Urs (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10:171–178.

Flory JA, Leibbrandt A, List J (2015) Do competitive workplaces deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions. *Rev. Econom. Stud.* 82:122–155.

Freeman NK (2007) Preschoolers' perceptions of gender appropriate toys and their parents' beliefs about genderized behaviors: Miscommunication, mixed messages, or hidden truths? *Early Childhood Ed. J.* 34:357–366.

Gneezy U, Rustichini A (2004) Gender and competition at a young age. *Amer. Econom. Rev., Papers Proc.* 94:377–381.

Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77:1637–1664.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118:1049–1074.

Günther C, Ekinci NA, Schwieren C, Strobel M (2010) Women can't jump?—An experiment on competitive attitudes and stereotype threat. *J. Econom. Behav. Organ.* 75:395–401.

Hyde JS, Lindberg SM, Linn MC, Ellis AB, Williams CC (2008) Gender similarities characterize math performance. *Science* 321: 494–495.

Kail RV, Cavanaugh JC (2010) *Human Development: A Life-Span View*, 5th ed. (Wadsworth Publishing, Belmont, CA).

Khachatryan K (2012) Gender differences in preferences at a young age? Experimental evidence from Armenia. Working paper, Stockholm School of Economics, Stockholm.

Meier S, Sprenger CD (2015) Temporal stability of time preferences. *Rev. Econom. Statist.* 97:273–286.

Miller AR, Segal C (2012) Does temporary affirmative action produce persistent effects? A study of black and female employment in law enforcement. *Rev. Econom. Statist.* 94:1107–1125.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122: 1067–1101.

Niederle M, Vesterlund L (2010) Gender and competition. *Annual Rev. Econom.* 3:601–630.

Niederle M, Segal C, Vesterlund L (2013) How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Sci.* 59:1–16.

Papaiakovou G, Giannakos A, Michailidis C, Patikas D, Bassa E, Kalopisis V, Anthrakidis N, Kotzamanidis C (2009) The effect of chronological age and gender on the development of sprint performance during childhood and puberty. *J. Strength Conditioning Res.* 23:2568–2573.

Sutter M, Kocher MG, Glätzle-Rützler D, Trautmann ST (2013) Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *Amer. Econom. Rev.* 103:510–531.

Villeval M-C (2012) Ready, steady, compete. *Science* 335:544–545.

Volk S, Thöni C, Ruigrok W (2012) Temporal stability and psychological foundations of cooperation preferences. *J. Econom. Behav. Organ.* 81:664–676.

Weichselbaumer D, Winter-Ebmer R (2007) The effects of competition and equal treatment laws on gender wage differentials. *Econom. Policy* 22:236–287.

Weinberger N, Stein K (2008) Early competitive game playing in same- and mixed-gender peer groups. *Merrill-Palmer Quart.* 54:499–514.

Wozniak D, Harbaugh WT, Mayr U (2014) The menstrual cycle and performance feedback alter gender differences in competitive choices. *J. Labor Econom.* 32:161–198.

Zhang J (2013) Can experimental economics explain competitive behavior outside the lab? Working paper, Hong Kong University of Science and Technology, Clear Water Bay.