Understanding Gender Differences in Leadership *

Sule Alan[†] , Seda Ertac[‡] , Elif Kubilay[§] and Gyongyi Loranth[¶]

November 2018

Abstract

Using data from a large-scale field experiment, we show that while there is no gender difference in the willingness to make risky decisions on behalf of a group in a sample of children, a large gap emerges in a sample of adolescents. The proportion of girls who exhibit leadership willingness drops by 39% going from childhood to adolescence. We explore the possible factors behind this drop and find that it is largely associated with a dramatic decline in "social confidence", measured by the willingness to perform a real effort task in public.

JEL Categories: C91, C93, D03, I28.

Keywords: leadership; gender; risk taking; social confidence; experiments.

*We would like to thank the ING Bank of Turkey, University of Vienna and TUBITAK (Grant no. 111K444) for providing funding. Ertac thanks the Turkish Academy of the Sciences (TUBA-GEBIP grant), the Science Academy (BAGEP grant); Alan thanks the British Academy for generous support. We are grateful to participants at the Conference on Economic Design in 2015, 2015 All-Istanbul Economics Workshop and the 2015 ISER conference at the University of Essex, Thomas Crossley, Armin Falk and Ragan Petrie for valuable comments and suggestions. We also thank numerous research assistants who helped us collect the data. All errors are our own.

[†]Corresponding author: Department of Economics, University of Essex, Wivenhoe Park, Colchester United Kingdom and Bilkent University, 06800 Ankara, Turkey. E-mail: salancrossley@gmail.com.

[‡]Department of Economics, Koc University.

[§]Institute for Social and Economic Research, University of Essex.

[¶]Department of Finance, University of Vienna.

1 Introduction

It is well-documented that women occupy top executive positions in politics and industry much less frequently than men. Leadership is an important component of many such careers. Rising in the hierarchy of corporations or in politics, one needs to increasingly take on leader roles, assuming the responsibility of making executive decisions. The stark scarcity of females in leadership positions persists despite much improvement in societal norms and institutional barriers in the recent years. For example, in the 2014 G20 summit, only 5 out of 58 leaders were female. Around the world, only 17% of government ministers, and only 4.6% of S&P 500 CEOs are female.[1] While explanations such as discrimination have also been put forward, self-selection, that is, differences in leadership ambition are likely a major factor behind these gender gaps. Indeed, there is evidence that women are less likely than men to *seek* to be elected into political leadership positions, and that female students are less likely to run for student government in college (Lawless and Fox (2008), New (2014), Kanthak and Woon (2015)). Consistently with this, many corporations, NGOs and colleges now implement leadership training programs targeted towards females, designed to both build women's leadership skills and get them interested in leadership in the first place.

A major component of a leader's job is to hold the power and responsibility of making decisions on behalf of others. These decisions (such as investment, financing and recruitment decisions in a corporation or campaign decisions in a political party) are often risky in nature and determine how the team, firm or party/electorate fares. In particular, they are consequential for the people who delegate the decision-making responsibility to the leader. Building decision-making skills and learning how to handle responsibility and accountability for others' outcomes are in fact major focus points of most leadership training programs (Wood and Winston (2005), Blenko et al. (2010)). Attitudes toward responsibility in social contexts can be an important factor behind observed gender differences in leadership. The recent "leader emergence" literature in psychology shows that women have lower motivation to lead and may be more concerned about whether they will harm others with the decisions

[1] Inter-Parliamentary Union and UN Women (2015), Catalyst (2016).

they will need to make as a leader (Elprana et al. (2015)). Women have been found to be less willing than men to make decisions on behalf of others in risky contexts (Ertac and Gurdal (2012), Ertac and Gurdal (2016)) and less willing to assume a position of coercive power in groups (Banerjee et al. (2015)). It is this component of leadership, taking on the responsibility of decision-making, that we focus on in this paper.[2] Over and above differences in other traits relevant to leadership, such as risk tolerance or competitiveness, differences in attitudes toward decision-making responsibility may play a distinct role in why women are less likely than men to volunteer for (and rise to) leadership roles. The implication, which is of concern not only in economic but also in social and political domains of decision-making, is that critical decisions would be mainly left to men, potentially causing inefficiencies and an over-representation of the preferences of a particular subgroup of the population.

In this paper, we study the evolution of the willingness to assume the decision-maker role in a group, a major component of leadership, from childhood to adolescence. Using unique data from a large field experiment that involves a sample of children with the average age of 10 and a sample of adolescents of average age 13 in Istanbul, Turkey, we explore factors that are associated with leadership willingness and the gender gaps therein. The rich dataset allows us to measure and study a number of factors potentially associated with the willingness to take decision making responsibility: risk attitudes, self-confidence, gender role attitudes, and a novel measure of "social confidence". Although not longitudinal, our dataset is well-suited to study the evolution of these factors from childhood to adolescence, as our samples of adolescents and children represent the same narrowly defined socio-economic segment in our study site.

To measure self-selection into a decision-making role, we use a task where subjects are placed in three-person groups, and are asked whether they would like to be the one that makes a risky decision on behalf of the group, determining everyone's payoffs. Abstracting from any pecuniary concerns (rewards or punishment) potentially associated with being a leader, the task captures pure preferences towards taking on decision-making responsibility and being

[2]Leadership may also involve other components, such as acting first and leading by example. Voluntary leadership by example has been studied in, for example, public good contribution contexts (Arbak and Villeval (2013), Rivas and Sutter (2011), Cappelen et al. (2015)).

accountable for other people's payoffs, a fundamental aspect of executive decision-making and leadership.[3] We therefore refer to the choice of whether or not to take on the decision-maker role in the group as the "leadership choice". Using this measure, we first document that while there is no gender gap in the willingness to make a decision for the group in childhood, a large gender gap (about 19 percentage points) emerges among adolescents. We then set out to understand the factors associated with the emergent gender gap in leadership, in particular, the major potential contributors such as risk tolerance, self- and social-confidence, and gender role attitudes.

Self-confidence is believed to be one of the most fundamental factors determining selection into ambitious paths in educational and occupational settings. There is a large literature that has documented gender differences in self-confidence, with women holding a less positive view of their abilities than men (see Kling et al. (1999) and Croson and Gneezy (2009) for reviews). Lack of self-confidence has also been put forward as an explanation for women's dislike of negotiation (e.g. Babcock and Laschever (2009)) and their lower willingness to self-select into competition, leading to a major source of inefficiency if such negative beliefs occur despite truly high ability. Self-confidence is also likely to be associated with who rises to leadership positions in groups (see Reuben et al. (2012), who show that women are less likely to be selected as leaders of groups in a real effort context due to lack of confidence). However, voluntary leadership usually requires a type of self-confidence that goes beyond the individual belief that one can do well, and interacts with social concerns. The decisions a leader has to make on behalf of others typically face scrutiny from the people she represents. Especially in the case of a bad outcome due to a wrong decision or bad performance, the leader may be faced with expressed disappointment or disapproval from other group members and/or may feel guilt, regret or embarrassment because of having negatively affected others' payoffs. The willingness and ability to withstand public pressure (for example, being able

[3]Responsibility has been identified as an important component of decision-making related to the allocation of payoffs as well as risk-taking on behalf of others (Charness and Jackson (2009), Trautmann and Vieider (2012), Füllbrunn and Luhan (2015)). It has also been documented that payoff commonality in groups affects individual behavior in both strategic and non-strategic contexts (Charness et al. (2007), Sutter (2009)). As related concepts, Bartling et al. (2014), Ertac et al. (2016) and Neri and Rommeswinkel (2017) study preferences for decision rights, autonomy and power.

to generate convincing arguments against dissent, being able to overrule opposition or facing the aftermath of a dismal public performance) are likely to be necessary traits to possess for a leader. Someone without such confidence may therefore not want to assume the decision-maker role in the first place.

In order to study the role of self-confidence in leadership, we develop two incentivized measures. These involve a mathematical real effort task where the subject is allowed to opt for a more difficult-higher reward or an easier-lower reward version of the same task. We use the difficulty choice as a measure of (private) self-confidence, with the conjecture that it proxies the subject's assessment of her own ability.[4] We then measure subjects' willingness to face social scrutiny. This measure involves eliciting subjects' willingness to perform the same mathematical task in public, i.e., in front of peers and experimenters. We conjecture that this measure, which we refer to as "social confidence", captures a unique aspect of self-confidence that is relevant for leadership decisions over and above what is captured by the private, individual choice of task difficulty. We document that there is about a 9 percentage-point gender gap in social confidence in childhood already, and this gap becomes very large (about 25 percentage points) in adolescence. Even after controlling for ability, risk tolerance and private self-confidence, girls are 18 percentage points less likely to accept to perform the mathematical task on the board, in front of their peers.

We find that social confidence is the single most important predictor of willingness to make decisions on behalf of others in both childhood and adolescence. The predictive power of this measure is a lot more prominent for girls and it increases significantly going from childhood to adolescence: while girls' willingness to perform under public scrutiny increases the propensity of leadership willingness by 17 percentage points in childhood, the effect almost doubles (becomes 32 percentage points) in adolescence. Our results suggest that the dramatic gender gap that emerges in social confidence in favor of boys may largely be responsible for the concurrent gender gap in leadership willingness in adolescence. Additional data from a supplementary experiment conducted on a fresh sample of students show that girls have lower

[4] Niederle and Yestrumskas (2008) show that women are less likely to seek challenges and may do so because they are less certain of their ability to succeed in a difficult task.

social confidence in spite of the fact that they can succeed in public, highlighting the inefficient nature of the gap.

We offer a theoretical mechanism which helps us understand the relationship between leadership choice and social confidence that we observe in the data. To do this, we first set up a simple expected utility model augmented with psychological costs related to social concerns. We then perform a structural estimation exercise where we estimate the cross-sectional distribution of the coefficient of relative risk aversion and the joint distribution of psychological costs of acting under public scrutiny, using an indirect estimator. With this exercise, we show that a simple expected utility model that incorporates social concerns into decision-making can successfully generate the predictive power of social confidence on leadership choice and justifies the gender gap among adolescents we observe in the data.

Gender differences in risk aversion, competitiveness and self-confidence are well-documented in individualistic performance and decision settings (see Croson and Gneezy (2009) for a review). Social performance contexts that involve accountability for others include an extra layer over and above individual decisions that may be particularly conducive to gender gaps favoring men. This paper puts forward a novel measure of "social confidence", a previously overlooked aspect of confidence, and identifies its role as a primary factor behind an individual's reluctance to rise to a decision-making position. The results point to adolescence as a period where social confidence declines more dramatically in girls, and a concurrent gender gap emerges in leadership willingness in decision-making, with boys more likely to volunteer to make decisions on behalf of others. The results offer new insight into why so few women are in decision-making positions in politics and in the business world, and implications for designing interventions to prevent these gaps from emerging in the first place.

The rest of the paper is organized as follows: Section 2 provides the background and experimental design, Section 3 presents the data and discusses the results, Section 4 concludes.

2 Background and Experimental Design

For our main analyses, we use data from two cohorts of students in a number of state-run schools in Istanbul. Our sample consists of elementary school students (children sample) who were in 4th grade, and middle school students (adolescent sample) who were in 8th grade at the time of the data collection.

The elementary school data are collected as part of a large-scale field study implemented with the aim of evaluating a series of randomized educational interventions. The experiments we conducted for the purpose of this paper were carried out in the baseline of this study. We then launched another field study that involves adolescents in middle schools, with the conjecture that social pressures that reinforce traditional gender roles may kick in around puberty when physical changes manifest, and may lead to gender gaps in behavior (as documented in Andersen et al. (2013) in the context of competitiveness). The average ages of the students are 10 and 13 for the children sample and the adolescent sample, respectively.[5]

The comparability of our children sample with the adolescent sample is facilitated by a unique feature of the Turkish education system. In Turkey, while middle and high income families mainly choose private schools, lower-SES families (our target group) tend to send their children to public schools in their catchment areas. In some districts elementary and middle schools share the same ground. Due to this locational convenience, a significant proportion of elementary school students spend their middle school years in the same school ground. We chose our sample of middle schools from among the elementary schools in our sample. Because 12 years of education is now compulsory in Turkey (with 4 years of elementary, 4 years of middle and 4 years of high school), there is no attrition at the middle school level based on gender. In addition, there is no performance-based selection into schools going from elementary to middle school. That is, students whose families sent them to state-run elementary schools stay in the state school system for the middle school as well, and stay in the same school if it has a middle school in the same ground. Therefore, we are confident that

[5]The average age of menstruation in a sample of Turkish girls in a recent study was found to be 12.2 (Bundak et al. (2008)), which is a major stage marker for girls in puberty.

our sample of children is fully comparable to our sample of adolescents. [6]

2.1 The Leadership Task

Our outcome variable, leadership willingness, is elicited using an incentivized experiment, based on Ertac and Gurdal (2012). The experiment consists of two tasks, the individual and the group decision task, one of which is randomly selected at the end for payment. In the first task, subjects make an individual decision under risk. The second task, which is the group task, involves two stages. In the first, subjects state whether they would like to be the decision-maker for the group, and in the second, one individual makes the decision that determines the payoffs for the whole group. The risky decision task, which forms the backbone of the experiment, is based on Gneezy and Potters (1997). Students have 5 tokens corresponding to gifts from a gift basket, which they can allocate between a risky and a riskless option. Tokens placed in the risky option, which is conveyed to the children as putting the tokens in a particular bowl, are either tripled or lost, with 50% chance. Tokens that are not put in the bowl are safe. Uncertainty is resolved through a draw from an opaque urn that contains one yellow and one purple ball. If the yellow ball is drawn, the good outcome occurs. If the purple ball is drawn, the tokens placed in the risky bowl are lost.

In the group decision task, children are told that they will be placed into randomly-determined groups of 3 people. The decision task is the same allocation task as in the individual case. However, everyone in the same group gets the same payoff, based on a single group member's decision. Given that different people have different preferences as to how much risk to take and these preferences are not known, taking the responsibility of the decision inherently involves "social risk" coming from the imposition of one's own preferences. Investing most of the tokens into the risky option, for example, may lead to everyone getting a low payoff in the case of a bad draw. Similarly, keeping all in the safe option may turn out to be a bad decision for everyone ex-post. Being the decision-maker in such a context is related to a major component of leadership, which is that decisions made by leaders oftentimes have

[6] A statistical comparison of teacher reported socio-economic status (SES) across our children and adolescent sample yields p-value of 0.26. We should also note that there were no public policies or interventions around the study period that specifically targeted children or adolescents.

payoff consequences for others and involve responsibility. We therefore call the decision-maker in the task the "leader" in what follows.

Who among the three people will make the actual group decision is determined based on self-selection. Specifically, each individual states whether she would like to be the one making the decision on behalf of the group. The actual decision-maker is then randomly selected from among volunteers. If there are no volunteers, one individual is selected randomly from among the three. The decision made on behalf of the group by the leader is implemented, and everyone in the group gets the same payoff based on the leader's decision.[7] Knowing this mechanism, individuals make two decisions: (1) Whether they would like to be the group decision-maker, (2) In case they are selected as the decision-maker, what their decision would be. This allows us to collect decisions from all subjects regardless of leadership willingness.

We interpret saying yes to the question of whether one would like to be the decision-maker as leadership willingness. Notice that in this task, there is no payoff-related reason to say no to being the group decision-maker. Since leaders do not get monetarily punished for decisions that lead to low payoffs, someone who cares only about their own monetary payoff should always take the opportunity to implement her own preference. An individual who declines the opportunity to be a leader may be unwilling to impose her own preferences on the group or may not want to take the risk of causing a bad outcome that may not be liked by other group members.[8]

One concern that may come to mind with this design is whether the use of a random payment scheme creates an issue, if children and adolescents understand random payment schemes differently, given existing results that different subject pools (e.g. professional traders vs. undergraduates) may have different levels of comprehension of compounded lotteries (List

[7] How the uncertainty is resolved was a treatment variable in the elementary school sample. Specifically, in one treatment the decision-maker was also responsible for drawing the ball that determines what happens to tokens invested into the risky option. In another treatment, an assistant would be asked to draw the ball rather than the decision-maker, to test whether potential effects come from perceptions of individual bad luck. We do not find any differences in any behavioral measure (p-value=0.42 for leadership choices and p-value=0.78 for allocation decisions) with respect to this treatment variable, and therefore pool the data. In the adolescent sample, the decision-maker also had the responsibility of drawing the ball.

[8] Ertac and Gurdal (2012) and Ertac et al. (2016) show that (adult) women are much less likely than men to give an affirmative answer to the question of whether they would like to be the decision-maker for their group in this task.

and Haigh (2005)). Charness et al. (2016) provides a methodological discussion of the use of random payment schemes in experiments. While random payment has advantages such as the avoidance of cross-task contamination, hedging and wealth effects, it may create problems in terms of diluted incentives and the introduction of background risk. In our specific context, given that gender gaps within each cohort are our main focus, and given that there is no reason to expect differences in the way adolescent boys and girls (and younger boys and girls) react to the incentive structure, the random payment design is unlikely to confound our main results.

2.2 The Self- and Social Confidence Tasks

As mentioned above, self-selection into a leadership position is likely to be related to self-confidence, particularly in the face of social scrutiny. Someone who has a tendency to feel regret, guilt or embarrassment after making a decision that disappoints or is disapproved by others may decline the leadership position in the first place. Similarly, being able to withstand public dissent after a failed decision or dismal performance is likely a necessary trait to possess for a leader.

We propose an incentivized measure that aims to elicit this type of strength in the context of a real effort task, which we refer to as "social confidence". We conjecture that this measure will capture an important aspect of self-confidence that should be especially relevant for predicting leadership willingness. We use this measure along with a measure of "private" self-confidence in own performance that will not be subject to public scrutiny. To elicit both types of confidence we use a real effort task. Specifically, students are presented with a task where the goal is to find pairs of numbers in a grid that add up to 100 in elementary schools and 1000 in middle schools. The task has two versions. The 4-token task brings 4 gift tokens whereas the 1-token task brings 1 gift token in the case of success, with both types of task giving zero payoff in the case of failure. In both tasks, the goal is to find at least 3 pairs adding up to 100 (or 1000), within 1.5 minutes. However, the number grid in the 4-token task is larger, which is why this task is more difficult. Note that mathematical tasks have been widely used

in the literature documenting gender differences in competitiveness and self-confidence, and are useful for measuring differences that may have implications for educational and labour market choices.

For the private self-confidence measure, we ask the students whether they would like to do the difficult or the easy task, in case they will do the task by themselves, anonymously. The idea here is that individuals who are more confident in their ability to do well will be more likely to choose the more difficult task. To elicit "social confidence", we elicit students' willingness to perform this task in public, that is, on the board, in front of their classmates. Students are asked to decide which task they would like to perform, in case they are selected to do the task in front of the class. They also have the option to refrain from doing the task altogether. After everyone makes their decision, one student is selected at random, and her choice is implemented. If she chose to do the task on the board, she gets paid according to her performance. If she chose to opt out, another student is randomly selected to do the task (only the randomly selected student that does the task on the board is paid). In what follows, our measure of social confidence is a binary variable that takes the value of 1 if the student was willing to perform on the board and zero otherwise. The reason why we use the decision to refrain from doing the task altogether is that this is a self-preserving strategy that absolves the individual of any social pressure or potential embarrassment.[9] Although the probability of success is higher, doing the easy task on the board still involves (even stronger) social risk. This is because failure in the easy task can lead to social ridicule, and having chosen the easy task may not be appreciated by others even in the case of success. Refraining from doing the task altogether protects the individual from such risks, albeit at the cost of forgoing gifts.[10] Note also that we refer to the individual self-confidence measure as "private self-confidence" and occasionally refer to the social confidence measure as the "board task" throughout the

[9] Ludwig et al. (2017) show that women tend to downgrade their self-assessments if these assessments will be observed, that is, they are averse to overestimating themselves and others seeing this.

[10] In unreported regressions we find that using difficult task-board, easy task-board, and refraining as three separate categories does not change the results, in the sense that once the subject chooses to do the task on the board, it does not matter whether she chose the easy or difficult version, for predicting leadership (p-value=0.43). This confirms that refraining from doing the task altogether captures the social aspect of the task better than the version chosen once the individual accepts to perform in public. Tables A.9 and A.10 in the Appendix document the cohort and gender differences in the choice of doing the task on the board, respectively.

text.

In order to both familiarize students with the general task and have a measure of mathematical ability, before making the private difficulty choice and whether to perform on the board, students are given 2 minutes to find as many pairs as possible that add up to 100 (1000 for the adolescent sample) in a large number grid. We incentivized this part of the experiment as well by offering a small gift per correct answer.

2.3 Experimental Procedures

All experiments were conducted in-class, with pencil and paper, during the allotted class time for extracurricular projects; see sample instructions provided in the Online Appendix. Rewards were in the form of gifts for the elementary school children—each token that was earned in the selected tasks corresponded to one gift item that children could take from a gift basket that included attractive toys and stationary items. We took care to ensure that the gifts were of value to the children, and that the basket included adequate numbers of each type of gift. In the adolescent sample, tokens corresponded to coupons worth 1TL (about \$0.5 at the time).[11] We implemented both the individual and the group decision tasks in a single class hour, and one task was selected at random for payment at the end of the session.

Children first made a decision in the individual investment task, and then proceeded to the group task. To collect decisions, children were (randomly) distributed choice sheets that had their group's ID number. At the time of decision, children did not know with whom they were in a group. After the leadership decision and the group investment decision were made, we collected the sheets and sorted them according to group ID. At the end of the session, either the individual part or the group part was randomly selected for payment. If the individual decision was selected, each child received gifts based on her individual risk allocation decision and the outcome of the random draw. If the group decision was randomly selected for payment, we determined the group decision-makers according to the mechanism of random selection among volunteers. Each choice sheet had a letter in small print (A, B

[11] It is common in the literature examining the evolution of economic behavior and related gender gaps over age to use gifts for younger children and money for adolescents (e.g. Sutter and Glätzle-Rützler (2015), Kosse et al. (2018)).

or C). In case of ties (more than one person or no one willing to decide), letters earlier in the alphabet took precedence. This procedure achieves randomness, since choice sheets were distributed randomly. At this stage, the identity of the group decision-maker and his/her decision was revealed to everyone in the group, which amplifies the social risks associated with being the group decision-maker. Based on the decision-maker's choice of tokens invested and the random draw, everyone in the group received the same number of gifts.

In the elementary school sample, the self-confidence tasks and the individual-group decision tasks were conducted on two separate days because of logistical constraints, while in the middle school sample all were done on the same day. The individual and group decision tasks came before the self-confidence task in both children and adolescents. In addition to the main experiments, we report results from an additional (smaller) field study conducted on a fresh sample of children and adolescents, in Section 3.4.

3 Data and Results

In addition to our incentivized social and private self-confidence measures, our data contain a number of other variables. We utilize these variables as potential predictors of leadership choice. One such predictor is risk attitude. As explained in Section 2.1, we elicit risk attitudes using the Gneezy-Potters investment task in the context of the individual decision-making part of the leadership task. In this task, children choose how many of their 5 gift tokens to invest into a risky option where invested tokens are either tripled or lost, with a lower number of tokens invested into the risky option indicating higher risk aversion; see Charness et al. (2013) for a review of the use of this task for eliciting risk preferences. As a measure of mathematical skill, we use the number of pairs found in the initial piece-rate number task that was conducted before choices were made.

We also use a battery of survey questions with which we construct a summary score that measures grit, a non-cognitive skill that has been shown to correlate with academic achievement as well as competitiveness (see Duckworth et al. (2007), Duckworth and Quinn (2009), Alan and Ertac (2018)). We conjecture that in this context grit may play a role as one

might expect that gritty individuals, i.e., those who set challenging goals and are perseverant, are more likely to self-select into leadership positions. Finally, using a large number of survey questions, we construct a summary score that measures how traditional students' beliefs on gender roles are, with the conjecture that these beliefs may play a role in volunteering to become the group leader. We provide the translation of all survey questions used to construct the grit and gender stereotype scores in the Online Appendix. All survey data were collected after experimental measures, in order to prevent potential priming effects on behavior.

While all data on adolescents were collected in a single visit to participating middle schools, data on children were collected in different sessions (days) as this effort was part of a bigger field study with a much larger sample. This created a moderate missing data problem for our elementary school sample because on a given day, about 20% of the students do not attend school for various reasons such as common viral infections. This non-attendance is likely to be random, and consistently with this, we see that girls and boys do not have significantly different likelihood of missing school (p=0.422), and children with missing values for covariates have the same leadership willingness as those who have full data (p=0.280). In the adolescent sample, we also have some students with missing covariates, in this case not due to non-attendance but incomplete questionnaire data (e.g. on gender roles, grit). Here, boys are more likely to have missing covariates (p=0.001) but reassuringly, the leadership willingness and social confidence of those students with and without missing covariates are similar (p=0.969 and p=0.841, respectively). For our main analyses, we restrict our data to those for whom we have the non-missing leadership indicator and impute missing values of our covariates. We provide our main results without imputation in the Online Appendix (see Tables A.14).

Our main sample consists of 769 children and 625 adolescents who participated in the leadership task. These data come from a total of 18 schools (25 classrooms in elementary schools and 21 in middle schools). All data were collected using pencil and paper by physically visiting the classrooms. In all analyses, we cluster standard errors over classroom to account for intra-cluster correlations.

3.1 Descriptive Statistics

Table 1 provides the sample statistics of some of the key variables used in our analyses for boys and girls separately, in the children and adolescent samples. Empirical distributions of all non-binary variables, i.e. math ability, risk tolerance, self reported grit and self reported gender roles measures are depicted in figures A.2, A.3, A.4 and A.5 in the Online Appendix. The very first row documents the statistics that motivate the paper: the proportion of students who state their willingness to be the decision-maker for the group. Here, we note two observations: First, the willingness to decide on behalf of a group is much higher in the elementary school sample (75% in the whole sample with both girls and boys) than in the adolescent sample (56% in the whole sample). Second, while leadership willingness declines going from childhood to adolescence for both girls and boys, a large gender gap of 19 percentage points emerges in favor of boys. Specifically, while boys' willingness to lead declines too as they become teens (by 10 percentage points), the proportion of girls who exhibit leadership willingness drops by 30 percentage points (39%) going from childhood to teen years, resulting in a significant gender gap in leadership willingness.

Table 1 also shows the differences between boys and girls in each age group with respect to a number of other attitudes and outcomes, which are potential factors associated with leadership willingness. It is clear from this table that some stark differences between genders are present even in childhood, and most of these differences persist into adolescence. A notable gap is in mathematical ability, as measured by initial performance in our real effort task. It appears that boys perform better in this context, both in childhood and in adolescence (see Fryer and Levitt (2010), Golsteyn and Schils (2014), Hyde et al. (1990)). Consistently with some of the previous findings in the literature, girls appear to be more risk averse than boys, although this gender difference seems to disappear in adolescence in our sample; see Croson and Gneezy (2009), Cárdenas et al. (2012), Sutter et al. (2013), Harbaugh et al. (2002), Khachatryan et al. (2015), Almås et al. (2016) for related evidence. They also exhibit higher self-reported grit and more progressive beliefs regarding gender roles.

An important finding in this table is the gender difference in self-confidence measures.

Note first that while there is no gender difference in private self-confidence in childhood, a significant gap emerges in adolescence. In terms of social confidence, a significant gender gap in favor of boys is already present in childhood and this gap significantly widens in adolescence. While girls are 9 percentage points less likely to state a willingness to perform the real-effort task on the board than boys in childhood (which is statistically significant), the gap becomes 25 percentage points in adolescence. In what follows, we will show that social confidence is the major predictor of leadership decisions. In particular, the change in social confidence favoring boys largely predicts the emerging gender gap in leadership willingness going from childhood to adolescence.

3.2 Leadership Willingness and its Determinants

Figure 1 shows the percentage of children and adolescents who exhibit leadership willingness. The two panels present the finding in the first row of Table 1 in visual clarity. The willingness to lead a group is quite high among children, with no statistically significant gender gap. Specifically, about 76% percent of girls and 75% of boys state that they want to be the leader. The picture changes dramatically when we look at our adolescent sample (Panel 2). Here, we see that the willingness to lead declines significantly and that a significant (19 percent) gender gap emerges going from childhood to adolescence.[12]

The first analysis we carry out aims to pin down the factors associated with leadership willingness. Table 2 presents the predictive power of the variables in Table 1 in determining leadership willingness in childhood and adolescence. Our measure of social confidence (board task) appears as the major predictor of leadership willingness in both childhood and adolescence: While children who elect to perform a mathematical task in front of their peers are about 16 percentage points more likely to exhibit willingness to make a risky decision on behalf of a group, the impact of the social confidence measure increases in size in adolescence (about 25 percentage points). Compared to a model without social confidence, adding in

[12] In both elementary and middle schools, students willing to be leaders take significantly more risk on behalf of their groups than students unwilling to be leaders (2.85 tokens vs. 2.42 tokens invested in the risky option in elementary school, with p-value=0.02, and 3.03 vs. 2.77 tokens in middle school, with p-value=0.01). This suggests that the decisions made in leadership positions depend on the type of selection into these positions.

social confidence increases R-square by almost 107% in childhood and 62% in adolescence, higher than increases due to any of the other covariates. Note also that self-reported grit is significantly and positively correlated with leadership willingness in childhood and adolescence.[13] Specifically, a one standard deviation increase in the grit score is associated with about a 4 (6) percentage point increase in leadership willingness in childhood (adolescence). Risk tolerance and private self-confidence emerge as significant predictors only in adolescence.

Given that we are interested in understanding the factors behind the gender gap in the leadership decision, it would be informative to analyze the predictive power of these covariates separately for boys and girls. Table 3 presents this analysis for our full specification (columns 2 and 4 in Table 2). A number of interesting findings should be noted here. First, social confidence is the strongest predictor for both boys and girls, especially in adolescence, but its impact is higher for girls than boys within both age groups. In particular, going from childhood to adolescence, the impact of this measure almost doubles for girls, although we cannot reject the equality of coefficients for either cohort (p-values of 0.44 and 0.16 for the children and adolescent samples, respectively). Second, risk tolerance is an important predictor for girls in childhood and boys in adolescence. Third, grit seems to be an important predictive factor for the leadership choice only for girls in both childhood and adolescence. Finally, private self-confidence is positively associated with leadership decisions for both genders in adolescence, albeit lacking statistical significance when we look at subgroups, possibly due to the smaller sample size.

So far, our findings highlight an emergent gender gap in leadership willingness going from childhood to puberty and a number of important factors that seem to determine this attitude, whose predictive powers are different across gender and age groups. Can changes in these underlying predictive factors explain the gap that emerges in adolescence? In the next section, we attempt to identify the changes in these predictive factors and explore how these changes contribute to the gap in leadership willingness going from childhood to adolescence.

[13] Our self-reported grit measures are factors extracted from a survey that contain statements related to grit. The following questions are found to have the highest factor loadings, i.e. explanatory power: questions 6, 8 and 10 in childhood and in questions 6, 7 and 10 adolescence. Survey questions are provided in the Online Appendix.

Before moving on to what explains the gender gap, it is worthwhile to discuss whether the fact that girls enter puberty earlier than boys confounds our results. Puberty is a transformation process rather than a single event, and the onset of puberty has been found to occur at a mean age of 10.1 for girls in Turkey and a mean age of 11.6 for boys (Bundak et al. (2007), Bundak et al. (2008)), suggesting that all students in our adolescent sample are likely to have at least begun the process. Table A.12 in the Online Appendix shows that if we separate age into three groups in the adolescent sample and take the oldest group where both boys and girls are likely to have entered puberty, we still have the result that boys are more willing to become leaders.

3.3 Explaining the Emerging Gender Gap in Leadership Willingness

In this section, we explore the relative contributions of the "change" in the aforementioned predictive factors to the "change" in the gender gap in leadership willingness between childhood to adolescence. Figure 2 depicts the changes in the gender gap in leadership and changes in the gender gap in the predictive factors we examine in earlier sections, by presenting difference-in-difference estimates of the gender gaps with 95% confidence bands.[14] The top line shows the "change" in the gender gap in leadership choice, that is, the gap we observe in adolescence minus the gap we observe in childhood (approximately 19% with p-value=0.00). Coefficients plotted on the right hand side of the zero line represent the change in gap estimates in favor of boys, while the left hand side depicts those in favor of girls. This figure clearly shows that the only factors for which the gender gap goes in the same direction as that in leadership willingness are private self-confidence and social confidence.

These results suggest that the dramatic decline in self-confidence and in particular, social confidence, may explain a significant portion of the emergent gap in leadership willingness. Interestingly, the gender gaps in risk tolerance and progressive beliefs on gender roles seem to shift in favor of girls, while we do not observe any significant change in gender differences in math ability or grit. While remaining the same in levels, the contribution of the latter

[14]The coefficients plotted are obtained from the empirical model: $y_i = \alpha + \beta_1 Male + \beta_2 Elementary + \beta_3 MaleXElementary + \varepsilon_i$. The plotted coefficient is β_3, which shows the change in the gender gap going from childhood to adolescence.

two factors may become differentially larger going from childhood to adolescence. This in turn could contribute to the emerging gender gap. Going back to the estimates provided in Table 3 can provide some clues in this regard. Testing the equality of the coefficients across samples for each gender, we find no evidence of changing contribution of math ability for girls going from childhood to puberty (p-value=0.53). The coefficient estimate increases and turns positive for boys in adolescence but this increase does not represent a significant change in contribution (p-value=0.29). Similarly for grit, we see no evidence of changing contribution in a way that is different across genders. The predictive power of grit increases for both genders in a similar magnitude going from childhood to adolescence. Overall, it appears that only the gender-differential decline in social confidence stands out as a prominent factor in explaining the emerging gender gap in leadership.

A couple of caveats are in order here. First, even after controlling for social confidence and other factors, a large gender gap of about 12 percentage points remains (see the last column of Table 2). While this may suggest that pure preference change may be a major reason for the observed gap, it may also point to omitted factors. Second, without exogenous variation in social confidence (or a valid instrument), the documented relationship cannot be given causal interpretation. In what follows, we will try to shed more light on these issues with the help of supplementary data and a simple theoretical model.

3.4 Discussion

The above analysis establishes that social confidence, as measured by the willingness to perform a mathematical task in front of peers, is strongly associated with the willingness to assume a decision-making role. The reason why we take decisiveness as the dependent variable is conceptual: given that leaders are often faced with decision-making responsibility and this is a central aspect of leadership, the unwillingness to take on decision-making responsibility may be a major reason behind women's self-selection away from leadership. In this sense, decision-making on behalf of others (a potentially difficult social situation) is the central behavioral aspect of leadership we focus on, and social confidence is the level of ease with which

one can face such social situations. In our conceptualization, the level of social fear constitutes a reason why and determines the extent to which individuals shy away from making decisions on behalf of others. Having said that, it is likely that a number of unobserved confounds govern both social confidence and decisiveness simultaneously. Without a credible instrument for social confidence, we cannot give causal interpretation to the coefficient estimates presented in Tables 2 and 3.

Table 4 presents the coefficient estimates from a bivariate probit regression and as such, the extent to which unobserved confounds may be associated with both decisions. The last two rows in this table provide the estimates (95% confidence intervals) of the cross-equation correlation coefficients across two equations for each sample. As can be seen from this table, our data decisively reject the no correlation restriction for both children and adolescent samples. This finding suggests the presence of unobserved confounds governing both decisions.

Table 5 examines the social confidence variable in isolation. As shown in the table, a significantly higher portion of female students refrain from this task. Even after controlling for mathematical ability and risk tolerance, girls are about 7 (19) percentage points less likely to opt for the board task in childhood (adolescence).[15] Not surprisingly, private self-confidence is significantly associated with social confidence: willingness to attempt the difficult version of the task privately is associated with a 10 (12) percentage point increase in the willingness to do the task on the board in childhood (adolescence). Note that risk tolerance is significantly associated with the board task choice only in adolescence, which may suggest that the social risk involved in performing the task on the board may come into play especially in this period.

Why is it the case that girls shy away from this task? It may be that even if they are equally able, girls may be less likely than boys to succeed when they perform the task under public pressure, and they are aware of this issue. Put differently, if girls were asked to do the board task regardless of their willingness, perhaps they would not perform as well as boys

[15] It may be that when forming beliefs, even after controlling for own performance, children in each gender group might give some weight to the perceived group mean of their gender to make predictions about own performance. In order to account for this, we use (1) the ratio of the average math grade of girls to the average math grade of boys in a particular class, (2) the actual ability level of girls with respect to boys in our specific task in a particular class, (3) the question from the gender roles survey, which captures beliefs about girls' general math ability with respect to boys. Our result that girls are less socially confident is robust to controlling for these factors (regression results available upon request).

of the same ability level. This may be particularly relevant given the mathematical task, in which girls may experience stereotype threat (Spencer et al. (1999)).[16] One cannot test this idea by simply comparing the performance of girls who performed the task on the board with that of boys due to the obvious selection problem. Understanding whether social concerns have any direct impact on one's actual performance or whether such concerns are limited to beliefs and choices is important for mitigating gender-achievement gaps.

In order to compare performances in front of peers purged of selection, we organized an additional field study and supplemented our main data with a small fresh sample of students, a significant proportion of which were asked to perform the board task regardless of their initial choices. Contrary to the procedures followed in the collection of main data, we informed the students at the outset that they would make a choice, and while this choice would count with some chance, with some chance they would be asked to perform the task on the board regardless of what they chose.[17] In each class, after everyone made their decision, a random set of students were picked one by one and they were asked to do the (difficult) task on the board (or with very low probability, their own choice was implemented). We continued this procedure until we reached the end of the allotted time for our experiment. This gives us a sample of board performances that is largely free of self-selection. Children also did the leadership in decision-making task, which allows us to observe whether the data patterns regarding leadership replicate in this sample.

These supplementary data consist of 300 students. Among these, 155 constitute our supplementary elementary school sample (children), and 145 our middle school sample (adolescents). These students were recruited from one elementary and one middle school, about 2 years after the initial field experiment. These schools were new schools (not in our original sample) but students were the same grades and ages as in the original sample, from the same socioeconomic status. Therefore, our supplementary sample is expected to have similar demographic characteristics to our main sample. Table A.11 in the Online Appendix compares the

[16] However, we should note that even if this is true, we would expect one to at least choose the easy task on the board and get the one gift, since the probability of success is almost 100% in the easy task.

[17] The probability that the students would be asked to do the task was set to 90%. This ensures that while the decision to perform or not perform on the board is incentivized, a large majority of students would actually be imposed the board task.

key variables used in the paper for the main and supplementary samples. While math ability in both children and adolescents and private self-confidence in only adolescents are lower in the supplementary sample, there are no differences in the gender gaps in these variables across the main and supplementary samples (for math ability, p=0.261 for children and p=0.12 for adolescents; for self-confidence, p=0.841). Nevertheless, we caution that the purpose of this exercise is not to replicate our main results, rather, to provide some evidence on the rationale behind the decisions we observe.[18]

A total of 139 students performed the task on the board; 60 children and 79 adolescents. In this sample, a total of 106 students had chosen not to perform the task on the board (35% of the whole supplementary sample), similar in proportion to our main data (39%). Consistently with the results from the main data, we find that there is a significant gender gap in the willingness to perform on the board, with girls exhibiting lower willingness both in childhood and adolescence (13 percentage points and 16 percentage points differences in childhood and adolescence, respectively).

Table 6 presents marginal effects from a logit model of the probability of success in the board task. Looking at the unconditional proportions (columns 1 and 3), we see that there is no gender difference in performance, either in childhood or in adolescence.[19] These results do not change when we control for private self-confidence, social confidence, risk tolerance, and math ability for the children sample but a 19 percentage point gender gap in favor of girls appears in the adolescent sample. This result makes the observed gender gap in the willingness to perform the board task all the more concerning from an efficiency perspective. It provides strong evidence that despite the fact that they would do well if they are asked to attempt them, females shy away from rewarding tasks that are to be performed under public pressure. Interestingly, social confidence has no predictive power on actual success on the board.

[18] Despite a small sample, however, our results on the determinants of leadership willingness are largely replicated in these supplementary data. In Figure A.1 in the Online Appendix, we again see an emerging gender gap in leadership willingness going from childhood to adolescence. We also replicate the strong relationship between leadership willingness and the willingness to perform the board task for adolescents.

[19] For one child in elementary school, the performance record is missing. Therefore, we have 59 observations instead of 60 in Column 1.

In this supplementary fieldwork, in order to better understand the role of social concerns in jointly determining leadership and board task choices, we conducted a survey in addition to the incentivized experiments. This survey involves a battery of questions that aim to elicit fear of embarrassment, assertiveness, anxiousness and fear of disappointing others, behaviors and attitudes which are likely to drive both leadership willingness and willingness to do the board task.[20] Using these questions, we construct standardized summary scores.

Table 7 shows how these summary scores correlate with leadership and board task choices. The signs of these correlations are quite intuitive. We find that leadership choice is strongly positively associated with assertiveness and negatively correlated with anxiousness: a one standard deviation increase in the assertiveness score increases the probability of leadership choice by 5 percentage points. Similar intuitive correlations are present in the board task choice as well: while a one standard deviation increase in the anxiousness score lowers the probability of leadership choice by about 9 percentage points, it lowers the probability of willingness to perform the board task by 13 percentage points. What is important in this table is that similar social concerns appear to influence both choices in the same direction, an observation we will exploit when we discuss our proposed mechanism via a simple expected utility model that might help in interpreting our results.

3.4.1 A Qualitative Analysis of Leadership Willingness and Social Confidence

In the supplementary fieldwork, we also asked those students who declined to decide on behalf of a group and those who opted out of the board task to give us the reason(s) for their decisions. For this, we gave students a large number of options to choose from.[21]

Figure 3 presents the distribution of the answers to the question "why did you not want to be the decision-maker for your group?" for the sample that said no to leadership, in children and adolescents. In general, 42% of children and 53% of adolescents express at least one "social concern" such as the fear of letting others down and not wanting to take the responsibility

[20] All questions are provided in the Online Appendix.

[21] Students were allowed to state multiple reasons for both questions. They were also allowed to write down their own answer if they did not think any of the options provided was applicable to them. Among the 26 (61) students who decline to be a leader in elementary schools (middle schools), none (1) wrote down their own reason. We do not include this student in this analysis.

for a bad outcome as reasons for their unwillingness to be a leader.

Figure 4 presents the distribution of reasons given by students who chose not to perform the board task. Here, in the children sample, social anxiety is the major reason stated. In adolescents, believing one is not good at math emerges as an important predictor as well as a dislike of performing in public. Overall, the analysis in this section gives us qualitative evidence on the importance of social concerns that are likely to influence both leadership willingness and board task choice.

3.4.2 Leadership Choice and Social Confidence: A Simple Model

In order to further facilitate the interpretation of our results, we stipulate a simple expected utility model augmented with social concerns in decision making. Suppose that subjects have a concave utility function that is defined over experimental rewards, separable from other consumption bundles. The expected payoff (π) of subject i who wants to invest x tokens into the risky option is:

$$E(\pi_i) = p(W + \alpha x_i) + (1 - p)(W - x_i)$$

where α is the gross return from investment, W is the initial endowment given to the subjects, x is the amount bet and p is the probability of winning. Assuming expected utility and a CRRA utility function, the solution for the optimal amount of investment in the risky option x^* for subject i is proportional to her endowment:

$$x_i^* = \left(\frac{1 - \Gamma_i}{1 + \alpha\Gamma_i}\right) W$$

with

$$\Gamma_i = \left(\frac{1 - p}{\alpha p}\right)^{1/\rho_i}$$

where ρ_i is the coefficient of relative risk aversion of subject i. Because the endowment and the return offered are the same for all subjects, what determines the differences in x across subjects is their risk aversion, which is captured by the coefficient of relative risk aversion ρ in this specification.

In the standard model, a rational individual i maximizes her expected utility so we would not expect her to prefer a suboptimal allocation of x_j^*, since

$$U(x_i^*) > U(x_j^*)$$

as long as $i \neq j$.

However, if one departs from the standard model and considers the fact that individuals also concern themselves with what others think and incorporate these social concerns into their decisions (as we document above using our supplementary data), the above relationship may take a more complicated form. These concerns may come into play in contexts where the individual's decision is consequential for others, as in our leadership task. These concerns may be modeled as psychological costs of self-image damage or fear of peer backlash in the case of a bad outcome. Such costs can justify why a rational subject may choose to delegate decision making in our context by essentially waiving the opportunity to implement her optimal allocation. In such a model a subject will choose to decide for the group if she thinks that such costs are worth bearing:[22]

$$U(x_i^*) - V(s_i) > U(x_j^*)$$

where $V(s_i)$ can be modeled as the psychic cost of imposing one's will on others, where $V(s_i) > 0$, $V'(s_i) > 0$, and $V''(s_i) > 0$. The argument s itself can depend on ρ, possibly with $\frac{\partial s}{\partial \rho} > 0$, on p with $\frac{\partial s}{\partial p} < 0$, and certainly on age, with $\frac{\partial s}{\partial Age} > 0$. Specifically, if decisions involve a risk of social retribution (e.g. investing all into the risky option and losing, leading everyone to get a low payoff), which is costly for the decision-maker, the impact of such social concerns may be higher for a more risk-averse individual. The link between risk-tolerance and social fears may also come into play only after a certain level of maturity, i.e. adolescence. That is, not only social concerns can increase with age, but the interaction of social concerns

[22] In the actual experiment, there is randomness coming from the incentive structure, in the sense that (1) the group task may or may not be chosen, (2) the individual may or may not be selected as the decision-maker. However, this randomness should not change the decision of whether to volunteer or not.

with other characteristics such as risk-tolerance may depend on age as well. For simplicity, we assume no such relationships in our structural model.

Given the empirical results we document using our supplementary data, it is plausible that these psychological costs also influence decision making in other contexts, such as the context we utilize to measure self-confidence. In our board task, the expected payoff for subject i is straightforward:

$$E(Payoff_i) = q_i(success)(R) + (1 - q_i(success))(0) \geq 0$$

where $q(.)$ is the subjective probability of finding three pairs within the allotted time on the board and R is the payoff in case of success, which is 4 gifts in our context. From this expression, a rational, payoff-maximizing subject who attaches a positive probability to her success is expected to exhibit a willingness to do this task. However, similar psychic costs may be at work in this context as well. In particular, the subject may decide to do the board task if

$$E(U(\bar{x} + Payoff)) - c_i > E(U(\bar{x}))$$

where $\bar{x}$ is the subject's expected payoff from the risk game and c is the cost of performing on the board. Here, the argument c can be the level of psychic cost of social pressure when performing the task (anxiousness, fear of embarrassment/being the center of attention etc., as also highlighted in the post-experiment questionnaire) and can very well be related to the subjective probability of success q, and age. All else equal, the psychic cost might be lower for a subject whose subjective probability of success is high. Note that an alternative would be to assume that the psychological costs come in only in the case of failure. Changing the model to reflect this does not change our results significantly.[23]

[23]The *actual* probability of success may also depend on psychological factors. For our main elementary school sample, we have some additional data that can shed some light onto private vs. public performance levels. Specifically, a random sample of these children were forced to do the difficult task privately. When we compare these children's private task performance with the performance on the board of children in the supplementary sample, controlling for observable characteristics, we find that performance on the board is significantly higher, with no differences across boys and girls. This suggests that children may have extra motivation when asked to do the task on the board, and points to the fact that a myriad of psychological factors (anxiety, extra motivation) may be involved when one is engaged in public performance.

The idea is that similar psychological costs can drive different behaviors and choices, as documented empirically in Table 7. Equally plausibly, different types of costs may govern different behaviors and choices but these costs may be correlated, generating a correlation between choices ex-post. For example, in our context, subjects' unwillingness to face their friends in the case of a bad outcome may primarily govern the decision of not becoming a leader. Alternatively, fear of being ridiculed by peers may govern the decision of not performing the board task. As long as these two concerns are correlated within individuals, the above model would yield a positive correlation between the two choices. As it is plausible to think that the importance of these concerns increases with age, the correlation may also become stronger in older ages.

In order to show that the above simple model can justify our empirical results, we perform a structural estimation exercise using its most stripped-down, fully parameterized form. We perform the matching exercise separately for boys and girls in the children and adolescent samples. After fitting the model (estimating the structural parameters via a simulated minimum distance estimator), we check whether the fitted model is able to generate the statistics we do not use for matching (a goodness of fit exercise), notably the correlation between leadership willingness and willingness to do the board task. Table 8 presents the structural estimation results, which are not of direct interest. Table 9 presents the fit of the model. Most statistics are matched quite closely, especially for the adolescent sample. The fit for the excluded statistic is very good in general for both samples and both gender, i.e., this simple and very restricted model is able to generate the positive correlation between the two experimental choices quite well. Combined with the empirical evidence, the results of this exercise suggest that incorporating social concerns into decision making is important for understanding choices that subjects perceive as consequential for others and those related to performance in public contexts. The details of the estimation procedure is given in the Online Appendix.

4 Conclusion

Understanding the forces behind self-selection into leadership positions is an important step toward designing effective policies that can mitigate inefficient gender gaps in labor markets as well as in corporate or political decision-making. This paper focuses on decision-making responsibility in groups and social performance, which are central aspects of a leader's job. The results highlight aversion to social scrutiny as a novel factor behind why women are less frequently observed in leadership positions. In particular, our results suggest that shying away from contexts that involve social pressure and/or scrutiny by others might explain why women often do not seek to rise to decision-making positions in groups, which require accountability for outcomes. Our results show that in a task performance context as well as a context where ability/effort is irrelevant and only preferences matter, sensitivity to social scrutiny arises as an important common thread that affects girls' behavior, i.e. leads them to refrain from situations that expose them to others' scrutiny. Differences across girls' and boys' leadership willingness are particularly strong in adolescence, when gender may become more salient and sex-typed behavior may be more likely to manifest (Hill and Lynch (1983)).

Given that many positions of leadership require social decision-making or social performance, the results suggest that being comfortable with potential public failure as a result of decisions or performance can be seen as a non-cognitive skill that may be conducive to rising to top positions and earning high rewards. Policies and interventions such as exposure to female role models in leadership positions or in occupations subject to public scrutiny (as in Beaman et al. (2012)) may be especially effective for girls in adolescence, which is when social fears seem to arise and contribute to gender gaps in choices. It may be especially important to target early puberty to ensure that worries about public self-image do not culminate in permanent damage in self-confidence and prevent girls from seeking and assuming decision-making roles in groups, committees or organizations.

Two caveats are worth mentioning here. First, while our results are strongly suggestive of the role of social confidence in explaining the gender gap in leadership, our data do not allow us to make any causal claims. Further research is needed to pin down this relationship

in a causal manner. Second, our sample represent a lower socio-economic segment of Turkey, therefore our results are not generalizable to the Turkish population. However, while Turkey is a muslim country with strong gender norms all across, these norms are particularly prominent in the country's low socio-economic segments. Hence, relevant policies may be more effective if they specifically target this sub-population.

References

Alan, S. and Ertac, S. (2018). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association*. Forthcoming.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., and Tungodden, B. (2016). Willingness to compete: Family matters. *Management Science*, 62(8):2149–2162.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., and Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95(4):1438–1443.

Arbak, E. and Villeval, M. C. (2013). Voluntary leadership: motivation and influence. *Social Choice and Welfare*, 40(3):635–662.

Babcock, L. and Laschever, S. (2009). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Banerjee, D., Ibañez, M., Riener, G., and Wollni, M. (2015). *Volunteering to take on power: Experimental evidence from matrilineal and patriarchal societies in India*. Number 204. DICE Discussion Paper.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, pages 2005–2039.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science*, 335(6068):582–586.

Blenko, M. W., Mankins, M. C., and Rogers, P. (2010). *Decide & deliver: 5 steps to breakthrough performance in your organization*. Harvard Business Press.

Bundak, R., Darendeliler, F., Gunoz, H., Bas, F., Saka, N., and Neyzi, O. (2007). Analysis of

puberty and pubertal growth in healthy boys. *European Journal of Pediatrics*, 166(6):595–600.

Bundak, R., Darendeliler, F., Günöz, H., Baş, F., Saka, N., and Neyzi, O. (2008). Puberty and pubertal growth in healthy turkish girls: no evidence for secular trend. *J Clin Res Pediatr Endocrinol*, 1(1):8–14.

Cappelen, A. W., Reme, B.-A., Sørensen, E. Ø., and Tungodden, B. (2015). Leadership and incentives. *Management Science*.

Cárdenas, J. C., Dreber, A., Von Essen, E., and Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in colombia and sweden. *Journal of Economic Behavior & Organization*, 83(1):11–23.

Catalyst (2016). Pyramid: Women in S&P 500 companies. New York: Catalyst.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51.

Charness, G. and Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, 69(3):241–247.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. *The American Economic Review*, 97(4):1340–1352.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087.

Duckworth, A. L. and Quinn, P. D. (2009). Development and validation of the short grit scale (grit–s). *Journal of Personality Assessment*, 91(2):166–174.

Elprana, G., Felfe, J., Stiehl, S., and Gatzka, M. (2015). Exploring the sex difference in affective motivation to lead. *Journal of Personnel Psychology*, 14(3):142–152.

Ertac, S., Gumren, M., and Gurdal, M. Y. (2016). Demand for decision autonomy and willingness to take responsibility in risky environments: Experimental evidence. Working paper.

Ertac, S. and Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, 83(1):24–30.

Ertac, S. and Gurdal, M. Y. (2016). Preference communication and leadership in group decision-making. Working paper.

Fryer, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–240.

Füllbrunn, S. and Luhan, W. J. (2015). Am I my peer's keeper? Social Responsibility in Financial Decision Making. NiCE Working Paper 15-03.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, pages 631–645.

Golsteyn, B. H. and Schils, T. (2014). Gender gaps in primary school achievement: A decomposition into endowments and returns to iq and non-cognitive factors. *Economics of Education Review*, 41:176–187.

Gouriéroux, C., Phillips, P. C., and Yu, J. (2010). Indirect inference for dynamic panel models. *Journal of Econometrics*, 157(1):68–77.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, 5(1):53–84.

Hill, J. P. and Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In *Girls at puberty*, pages 201–228. Springer.

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, 107(2):139.

Inter-Parliamentary Union and UN Women (2015). Facts and figures: Leadership and political participation. Retrieved from http://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures.

Kanthak, K. and Woon, J. (2015). Women don't run? election aversion and candidate entry. *American Journal of Political Science*, 59(3):595–612.

Khachatryan, K., Dreber, A., Von Essen, E., and Ranehill, E. (2015). Gender and preferences at a young age: Evidence from armenia. *Journal of Economic Behavior & Organization*, 118:318–332.

Kling, K. C., Hyde, J. S., Showers, C. J., and Buswell, B. N. (1999). Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin*, 125(4):470.

Kosse, F., Deckers, T., Pinger, P. R., Schildberg-Hörisch, H., and Armin, F. (2018). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy*. Forthcoming.

Lawless, J. L. and Fox, R. L. (2008). *Why Are Women Still Not Running for Public Office?* Brookings Institution.

List, J. A. and Haigh, M. S. (2005). A simple test of expected utility theory using professional traders. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3):945–948.

Ludwig, S., Fellner-Röhling, G., and Thoma, C. (2017). Do women have more shame than men? an experiment on self-assessment and the shame of overestimating oneself. *European Economic Review*, 92:31–46.

Neri, C. and Rommeswinkel, H. (2017). Decision rights: freedom, power, and interference.

New, J. (2014). Getting women to run. Retrieved from https://www.insidehighered.com/news/2014/11/04/female-students-still-scarce-student-government-executive-positions.

Niederle, M. and Yestrumskas, A. H. (2008). Gender differences in seeking challenges: The role of institutions. NBER Working paper.

Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1):111–117.

Rivas, M. F. and Sutter, M. (2011). The benefits of voluntary leadership in experimental public goods games. *Economics Letters*, 112(2):176–178.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1):4–28.

Sutter, M. (2009). Individual behavior and group membership: Comment. *The American Economic Review*, 99(5):2247–2257.

Sutter, M. and Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, 61(10):2339–23354.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, 103(1):510–531.

Trautmann, S. T. and Vieider, F. M. (2012). Social influences on risk attitudes: Applications in economics. In *Handbook of Risk Theory*, pages 575–600. Springer.

Wood, J. and Winston, B. E. (2005). Toward a new understanding of leader accountability: Defining a critical construct. *Journal of Leadership & Organizational Studies*, 11(3):84–94.

Table 1: Summary of experimental results, self-reported measures and demographics

	Children					Adolescents				
	Girls	Boys	p-value	SE	N	Girls	Boys	p-value	SE	N
Leadership	**0.76 (0.43)**	**0.75 (0.43)**	**0.87**	**0.04**	**769**	**0.46 (0.50)**	**0.65 (0.48)**	**0.00**	**0.05**	**625**
Math Ability	4.46 (2.34)	5.22 (2.30)	0.00	0.19	638	1.94 (1.72)	2.33 (1.86)	0.02	0.15	615
Risk Tolerance	2.50 (1.45)	2.70 (1.45)	0.02	0.08	760	2.83 (1.12)	2.67 (1.19)	0.21	0.13	620
Private Self Confidence	0.72 (0.45)	0.74 (0.44)	0.52	0.04	655	0.66 (0.47)	0.80 (0.40)	0.01	0.05	621
Social Confidence	0.70 (0.46)	0.79 (0.41)	0.01	0.03	654	0.33 (0.47)	0.58 (0.49)	0.00	0.05	622
Self Reported Grit	0.05 (0.99)	-0.19 (1.01)	0.00	0.08	618	0.12 (0.99)	-0.09 (1.01)	0.02	0.08	543
Self Reported Gender Roles	0.25 (0.93)	-0.38 (0.99)	0.00	0.08	659	0.47 (0.87)	-0.40 (0.94)	0.00	0.09	545
Age	9.96 (0.37)	10.1 (0.53)	0.00	3.53	712	13.3 (0.67)	13.4 (0.77)	0.10	0.06	601
SES	2.81 (0.97)	2.89 (0.98)	0.33	0.08	559	2.26 (0.51)	2.27 (0.64)	0.79	0.096	608
Household Size	2.28 (0.83)	2.34 (0.96)	0.33	0.05	724	2.09 (0.53)	2.03 (0.54)	0.24	0.05	605

Presented variables are constructed as follows: Leadership: a binary outcome variable that indicates whether the student chose to decide on behalf of the group (leadership choice); equals to 1 if willing to be a leader; 0 otherwise. Math ability: Number of pairs found in the number task implemented prior to the choice of task difficulty and choice of performing the task at the board. Risk tolerance: Number of tokens invested in the Gneezy-Potters task allocation of 5 tokens (privately made, prior to the leadership task). Private self-confidence: Binary choice of task difficulty, equals to 1 if task is 4TL; 0 otherwise. Social confidence: Binary choice of performing the task at the board, equals to 1 if willing to perform the task on the board; 0 otherwise. Self reported grit: Standardized summary score constructed using survey questions adapted from the Duckworth grit scale. Self reported gender roles: Standardized summary score constructed using survey questions targeting gender stereotypes. Grit and gender roles scores were constructed using a principal-component factor method. Higher values mean that individuals become more perseverant, and they tend to have more progressive gender role beliefs. SES is reported by the teacher based on a 1-5 item scale.

Table 2: Gender Gap in the Leadership Decision

	Children		Adolescents	
	(1)	(2)	(3)	(4)
	Leader	Leader	Leader	Leader
Male	-0.006	-0.023	0.185***	0.122**
	(0.04)	(0.04)	(0.05)	(0.05)
Math Ability		-0.007		0.004
		(0.02)		(0.02)
Risk Tolerance		0.029		0.038**
		(0.02)		(0.02)
Private Self-Confidence		-0.008		0.097*
		(0.04)		(0.05)
Social Confidence		0.155***		0.253***
		(0.04)		(0.04)
Self Reported Grit		0.040*		0.060***
		(0.02)		(0.02)
Self Reported Gender Roles		0.027		0.007
		(0.02)		(0.02)
Class Size		-0.002		-0.005
		(0.00)		(0.01)
High SES		0.018		-0.148
		(0.05)		(0.12)
Low SES		-0.078*		-0.064
		(0.04)		(0.04)
Age		-0.059		-0.021
		(0.04)		(0.03)
Household Size		0.019		-0.031
		(0.02)		(0.04)
Observations	769	769	625	625
R-Square	0.000	0.053	0.034	0.150
Adjusted R-Square	-0.001	0.038	0.033	0.133

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Gender Gap in the Leadership Decision

	Children		Adolescents	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys
Math Ability	-0.005	-0.005	-0.034	0.031
	(0.03)	(0.03)	(0.04)	(0.02)
Risk Tolerance	0.038**	0.022	0.023	0.052**
	(0.02)	(0.03)	(0.03)	(0.02)
Private Self-Confidence	0.015	-0.026	0.094	0.095
	(0.05)	(0.05)	(0.07)	(0.06)
Social Confidence	0.174***	0.137*	0.315***	0.203***
	(0.05)	(0.07)	(0.06)	(0.06)
Self Reported Grit	0.055*	0.025	0.071*	0.052
	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles	0.051*	0.008	0.042	-0.021
	(0.03)	(0.03)	(0.04)	(0.02)
Class Size	-0.004	-0.001	-0.004	-0.007
	(0.00)	(0.00)	(0.01)	(0.01)
High SES	-0.037	0.066	-0.422***	-0.124
	(0.07)	(0.06)	(0.09)	(0.13)
Low SES	-0.061	-0.104	-0.033	-0.088
	(0.07)	(0.06)	(0.07)	(0.07)
Age	-0.060	-0.060	-0.036	-0.003
	(0.06)	(0.04)	(0.06)	(0.03)
Household Size	0.039	0.009	-0.027	-0.044
	(0.03)	(0.02)	(0.07)	(0.05)
Observations	367	402	279	346
R-Square	0.078	0.052	0.149	0.125
Adjusted R-Square	0.049	0.026	0.113	0.096

Reported estimates are average marginal effects from from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Joint Estimation of Leadership Decision and Social Confidence

	(1) Children	(2) Adolescence
Leadership		
Male	-0.041	0.461***
	(0.14)	(0.15)
Math Ability	0.007	0.065
	(0.07)	(0.05)
Risk Aversion	0.108*	0.137***
	(0.06)	(0.04)
Self Reported Grit	0.137*	0.193***
	(0.07)	(0.05)
Self Reported Gender Roles	0.072	0.053
	(0.07)	(0.07)
Class Size	-0.008	-0.011
	(0.01)	(0.01)
Private Self-Confidence	0.020	0.333**
	(0.13)	(0.15)
High SES	0.059	-0.396
	(0.16)	(0.31)
Low SES	-0.218*	-0.147
	(0.13)	(0.12)
Age	-0.181	-0.062
	(0.11)	(0.07)
Household Size	0.065	-0.100
	(0.05)	(0.10)
Social Confidence		
Male	0.248**	0.512***
	(0.11)	(0.16)
Math Ability	0.186***	0.215***
	(0.07)	(0.05)
Risk Aversion	0.063	0.124**
	(0.05)	(0.05)
Self Reported Grit	0.003	0.119
	(0.07)	(0.08)
Self Reported Gender Roles	-0.105	0.147**
	(0.07)	(0.06)
Class Size	-0.002	0.016
	(0.00)	(0.02)
Private Self-Confidence	0.305***	0.353**
	(0.11)	(0.15)
High SES	0.046	0.174
	(0.13)	(0.36)
Low SES	0.185	0.161
	(0.16)	(0.14)
Age	0.003	-0.041
	(0.15)	(0.12)
Household Size	0.023	-0.068
	(0.06)	(0.08)
Observations	769	625
RHO- Elementary	(0.14-0.42)	
RHO -Middle	(0.28-0.52)	

Reported estimates are from a bivariate probit regression where the dependent variables are binary leadership choice and social confidence. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01. The last two rows give the 95% confidence bands for the correlation coefficient between the errors of two equations.

Table 5: Gender Gap in Social Confidence (Board Choice)

	Children		Adolescents	
	(1)	(2)	(3)	(4)
	Board	Board	Board	Board
Male	0.079**	0.074**	0.244***	0.185***
	(0.03)	(0.03)	(0.05)	(0.06)
Math Ability		0.056***		0.076***
		(0.02)		(0.02)
Risk Tolerance		0.018		0.043**
		(0.02)		(0.02)
Private Self-Confidence		0.100**		0.119**
		(0.04)		(0.05)
Self Reported Grit		0.000		0.040
		(0.02)		(0.03)
Self Reported Gender Roles		-0.029		0.049**
		(0.02)		(0.02)
Class Size		-0.000		0.006
		(0.00)		(0.01)
High SES		0.015		0.060
		(0.04)		(0.13)
Low SES		0.054		0.053
		(0.05)		(0.05)
Age		0.002		-0.013
		(0.05)		(0.05)
Household Size		0.005		-0.029
		(0.02)		(0.03)
Observations	769	769	625	625
R-Square	0.008	0.055	0.059	0.136
Adjusted R-Square	0.007	0.041	0.058	0.121

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary board task choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Gender Gap in Board Success

	Children		Adolescents	
	(1)	(2)	(3)	(4)
	Board Success	Board Success	Board Success	Board Success
Male	0.116	0.007	-0.101	-0.193*
	(0.11)	(0.10)	(0.13)	(0.09)
Observations	59	56	79	75

Reported estimates are average marginal effects from logit regressions where the dependent variable is the binary success at the board (supplementary data). The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Correlations

	Leadership Choice				Board Task Choice			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Leader	Leader	Leader	Leader	Board	Board	Board	Board
Fear of Embarrassment	-0.047				-0.092***			
	(0.03)				(0.03)			
Assertiveness		0.050***				0.026		
		(0.02)				(0.02)		
Anxiousness			-0.087**				-0.131***	
			(0.03)				(0.03)	
Fear of Disappointment				0.010				0.064*
				(0.02)				(0.03)
Observations	286	293	292	288	285	292	291	287

Coefficients presented are OLS coefficients obtained by running regressions of leadership and board task choices on the respective summary score. Presented standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table 8: Estimated Structural Parameters

	Children Sample		Adolescent Sample	
Estimated Structural Parameters	Girls	Boys	Girls	Boys
Mean Coefficient of Relative Risk Aversion ($ln(\rho)$)	-0.35	-0.46	-0.48	-0.38
Std. of Coefficient of Relative Risk Aversion (σ_ρ)	-0.07	-0.04	-0.64	-0.54
Mean psychic cost $ln(s)$	-3.07	-1.92	-1.22	-2.38
Mean psychic cost $ln(c)$	0.26	-0.10	1.31	0.19
Std. of psychic costs $ln(c)$ and $ln(s)$	2.31	0.52	2.05	2.15
Correlation between $ln(s)$ and $ln(c)$ (set)	0.5	0.5	0.5	0.5

Structural parameters are estimated by matching 5 ap's obtained from the model with those obtained from the main data by minimizing the criterion $\chi = (a^{sim} - a^{data})' \Omega^{-1} (a^{sim} - a^{data})$, where Ω is the variance-covariance matrix of data ap's .

Table 9: Goodness of Fit

PANEL 1: CHILDREN				
	Girls		Boys	
	Actual Data	Simulated Data	Actual Data	Simulated Data
Mean G&P Allocation	2.46	2.62	2.66	2.80
STD of G&P Allocation	1.49	1.51	1.52	1.53
Prop. of Leadership Choice	0.76	0.78	0.76	0.78
Prop. of Board Choice	0.72	0.53	0.81	0.73
Corr. G&P and Leadership	0.10	0.16	0.06	0.24
Corr. Leadership and Board	**0.21**	**0.22**	**0.10**	**0.07**
PANEL 2: ADOLESCENTS				
	Girls		Boys	
	Actual Data	Simulated Data	Actual Data	Simulated Data
Mean G&P Allocation	2.85	2.83	2.68	2.63
STD of G&P Allocation	1.13	1.12	1.19	1.18
Prop. of Leadership Choice	0.46	0.51	0.65	0.67
Prop. of Board Choice	0.34	0.36	0.57	0.53
Corr. G&P and Leadership	0.06	0.12	0.15	0.15
Corr. Leadership and Board	**0.34**	**0.29**	**0.24**	**0.26**

Table presents data and simulated ap's. Simulated ap's are calculated at the estimated structural parameters in Table 8. The last row in each panel presents the correlation coefficients between leadership and board choice. This ap was not used for fitting the model so presents a goodness of fit.

Figures

Figure 1: Gender Gap in Leadership Decision

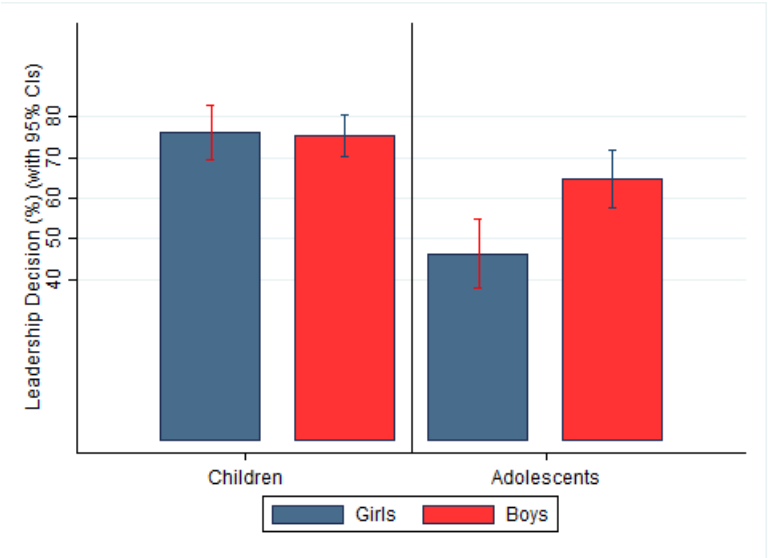

Figure 2: Change in the Willingness to Lead and Its Determinants from Childhood to Adolescence

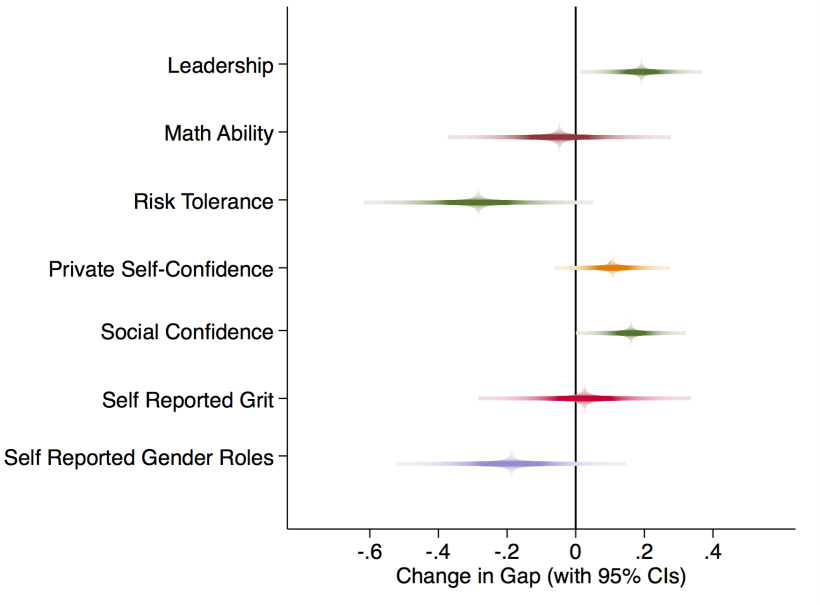

Figure 3: Frequency of Stated Reasons for Avoiding Leadership

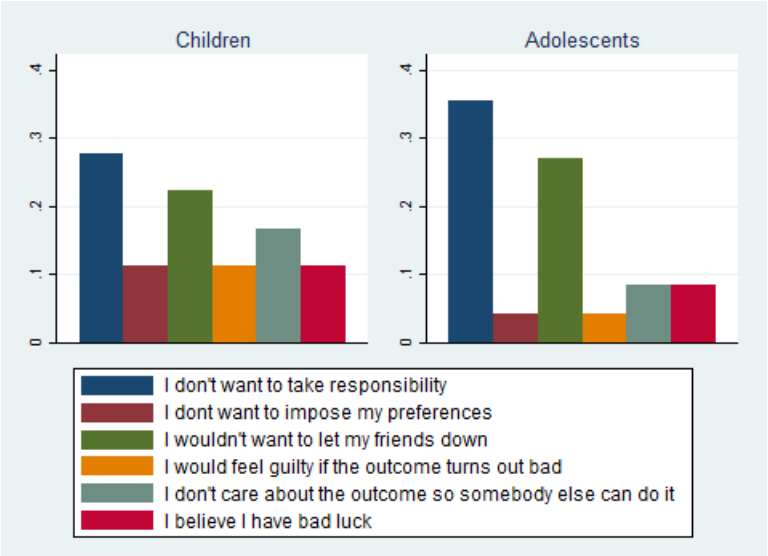

Figure 4: Frequency of Stated Reasons for Avoiding the Board Task

Appendix (For Online Publication)

1 Leadership Choice and Social Confidence: A Simple Model

Based on the simple model laid out in Section 3.4.2, we conduct a structural estimation exercise using a Minimum Distance estimator. Below, we provide the details of the estimation procedures and present the results.

1.1 Simulated Minimum Distance Estimation

Assume that the utility function is iso-elastic (CRRA) and the psychological cost function V is convex in its argument s.

$$U(x_i) = p\frac{(W + \alpha x_i)^{1-\rho_i}}{1-\rho_i} + (1-p)\frac{(W - x_i)^{1-\rho_i}}{1-\rho_i}$$

$$V(s_i) = \frac{s_i^2}{2}$$

The subscript i denotes an individual subject therefore we only assume heterogeneity in ρ, c and s. The other parameters are set to the values used in the actual experiments; $W = 5$, $\alpha = 3$, $p = 0.5$. We estimate this model using indirect inference. Gouriéroux et al. (2010) provide evidence in favor of using indirect inference in the context of estimating fully parametric models. Indirect inference requires the specification of a set of statistics which are known as auxiliary parameters (ap's). Estimation proceeds by comparing the ap's based on the sample with those based on the simulated data from the model. The estimated distribution parameters are determined by minimizing the weighted distance between the two sets of ap's. The ap's can be moments or functions of moments or any other statistics as long as each ap does have a probability limit as the number of cross-section units becomes large.[24]

We first fully parameterize our model. For simplicity, we assume that ρ is independent

[24]These statistics do not have to be statistics of interest, nor do they need to be an unbiased estimator of anything. All we need is that they have a monotonic relationship with some of the structural parameters.

of c and s, but we allow for a positive correlation between c and s. We assume that ρ_i is distributed lognormally with

$$ln\rho \sim N(\mu_\rho, \sigma_\rho)$$

Similarly, we assume that c and s are joint lognormally distributed with:

$$\begin{bmatrix} \mu_s \\ \mu_c \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{s,c} \\ \sigma_{c,s} & \sigma^2 \end{bmatrix}$$

In order to restrict the parameter space and to show that even a very restrictive model can generate the results we obtain, we set $\sigma_s = \sigma_c = \sigma$ and the correlation coefficient to 0.5. Given these restrictions we reduce the number of structural parameters to estimate, which we denote with vector θ_n, to 5, $\theta_n = \{\mu_\rho, \sigma_\rho, \mu_s, \mu_c, \sigma\}$.

Given this model and starting values of the 5 structural parameters, we first simulate x_i^*, and a vector of leadership indicator l_i as well as board task choice indicator b_i[25] for each subject M times. [26] We set the number of subjects to 769 for children and 625 for adolescents, as in data. Using these simulated data we generate k ap's, which describe the above structural environment. In principle, one can generate as many auxiliary parameters as possible, as long as the number is at least as many as the number of structural parameters to estimate. The model is just identified when $n = k$, and overidentified when $n < k$. The latter situation is useful to test the model by testing overidentifying restrictions. For this exercise we prefer a just-identified model as the purpose is not to test a particular model but rather show that the choices we observe in the data can be justified with a simple model.

To match the distribution of the coefficient of relative risk aversion, we use the vector x_i^* for the simulated data and Gneezy and Potters (GP) task choices in the actual data, we can generate two auxiliary parameters to identify μ_ρ and σ_ρ^2 :

[25] Each individual i is compared to a random individual j to get the leadership decision. An alternative would be to compare it to the, say, classroom average. These choices do not change the results qualitatively.

[26] Consistency is achieved for any number, M, of simulated data sets. The issue is that the use of simulation inflates asymptotic standard errors by the factor $(1+M^{-1})^{\frac{1}{2}}$; for M $\geq$ 10, this factor is negligible. Therefore we set the number of simulations to 10. It should also be noted that the asymptotic standard errors are unreliable in these highly non-linear models.

$$ap_1^{sim} = mean(x_i)$$

$$ap_1^{data} = mean(GP_i)$$

and

$$ap_2^{sim} = std(x_i)$$

$$ap_2^{data} = std(GP_i)$$

To match the joint distribution of s and c we use the proportion of subjects who want to be the leader and who want to perform the task on the board:

$$ap_3^{sim} = prop(l_i)$$

$$ap_3^{data} = prop(Leader_i)$$

$$ap_4^{sim} = prop(b_i)$$

$$ap_4^{data} = prop(Board)$$

Finally we use the correlation between x and l.

$$ap_5^{sim} = corr(x, l)$$

$$ap_5^{data} = corr(GP, Leader)$$

Note that we have 5 auxiliary statistics to estimate 5 structural parameters (the model is just-identified) so the match should be exact, i.e, the below criterion function should be driven to zero:

$$\chi = (a^{sim} - a^{data})' \Omega^{-1} (a^{sim} - a^{data})$$

where Ω is the weighting matrix, which is the variance-covariance matrix of $(a^{sim} - a^{data})$.

1.2 Structural Estimation Results

We perform the matching exercise separately for boys and girls in the children and adolescent samples. After fitting the model (estimating the structural parameters via a simulated minimum distance estimator), we check whether the fitted model is able to generate the statistics we do not use for matching (a goodness of fit exercise), notably the correlation between leadership willingness and willingness to do the board task. The estimated structural parameters, which are not of direct interest, are presented in Table 8 in the main text. Table 9 in the main text presents the statistic of interest that we do not use in the matching algorithm: the correlation coefficient between leadership and board task choice.

2 Survey Questions

2.1 Survey Questions for Grit Score

4-point item scale: completely agree, agree, disagree, completely disagree

Children:

1. Your intelligence is something very basic about you that you can't change very much.
2. If I study sufficiently, I could be the most successful student in the class.
3. I like school work best which makes me think hard, even if I make a lot of mistakes.
4. Truly smart people do not need to try hard.

5. When I receive a bad result on a test I work harder in this class from now on.
6. When I receive a bad result on a test I spend less time on this subject and focus on other subjects that I'm actually good at.
7. If you are not good at a subject, working hard won't make you good at it.
8. If I set a goal and see that it is harder than I thought, I easily lose interest.
9. Setbacks discourage me.
10. If I think that I will lose in a game, I do not want to continue playing.
11. When I receive a bad result on a test it's because I'm just not good at this subject.
12. I am confident that I will be successful in my work in my future life.
13. When I receive a bad result on a test it's because the test was unfairly hard.
14. When I receive a bad result on a test it's because I didn't really like the subject.
15. When I receive a bad result on a test it's because I am not smart enough.
16. I don't like it when people (my teacher, parents, friends) make comments and suggestions about how to improve my performance in a class, game or task that I am not very good at.
17. No matter how much intelligence you have, you can always change it quite a bit.
18. Music or drawing talent can be learned by anyone.

Adolescents:

1. Your intelligence is something very basic about you that you can't change very much.
2. If I study sufficiently, I could be the most successful student in the class.
3. I like school work best which makes me think hard, even if I make a lot of mistakes.

4. Truly smart people do not need to try hard
5. When I receive a bad result on a test I work harder.
6. When I receive a bad result on a test I spend less time on this subject and focus on other subjects
7. If I'm having difficulty in a task, it is a waste of time to keep trying. I move on to things which I am better at doing
8. If you are not good at a subject, working hard won't make you good at it.
9. I prefer easy homework where I can easily answer all questions correctly.
10. If I set a goal and see that it is harder than I thought, I easily lose interest.
11. Setbacks discourage me.

2.2 Survey Questions for Gender Stereotypes

4-point item scale: completely agree, agree, disagree, completely disagree

1. Boys are more talented in math than girls.
2. Boys have a better understanding of monetary/financial matters than girls.
3. It is natural for boys to be better at school than girls.
4. In a family, it is the father's responsibility to make money and the mother's responsibility to take care of the children.
5. It is natural for boys to be the leaders in a group rather than girls.
6. It is natural for girls to help with household chores more than boys.
7. Being an astronomer or an astronaut is not a proper occupation for a woman.
8. [Girls only] I prefer being a housewife to working outside of the home.

9. Being a nurse is not a proper occupation for a man.
10. Girls are as intelligent as boys.
11. If there are no financial difficulties in a family, it is more appropriate for the wife to stay home and raise the children rather than have a job outside.
12. My mother would prefer me to be a housewife rather than work outside of the home.
13. My father would prefer me to be a housewife rather than work outside of the home.

2.3 Supplementary Survey

4-point item scale: absolutely true for me, true for me, not true for me, absolutely not true for me

1. I shy away from talking in public because I am afraid of embarrassing myself.
2. During class hours I refrain from raising my hand and speaking because I don't want to embarrass myself in front of my teacher and my friends.
3. I don't feel uncomfortable when people criticize me.
4. It is never hard for me to read a poem or solve a problem on the blackboard while everyone is looking at me.
5. I don't shy away from expressing to others what I want.
6. Since I don't know the class material well, I refrain from raising my hand and talking in class.
7. It is important for me to decide on my own, I don't like others interfering with my decisions.
8. I don't get nervous when I talk to my teachers or other adults, and I never shy away from talking to them.
9. I get sad when the game that I prefer is not played in a group.

10. Sometimes I get nervous and forget what I know when I am on the board.

11. I am scared of not being able to do what my friends, family and teachers expect me to do, and of letting them down.

12. I don't care about what others think about me, I do things as I like.

13. When my teacher asks a question, I refrain from raising my hand or coming up to the board even if I know the answer.

14. I am scared of making others sad (as a result of my behavior).

15. I frequently feel guilty.

16. I refrained from becoming a leader, because:

    (a) I did not want to take the responsibility for a bad outcome.

    (b) I did not want to impose my decision since I did not know what others want.

    (c) I was afraid that the outcome would be bad and I'd have let my friends down.

    (d) I was afraid that the outcome would be bad and my friends would be angry with me.

    (e) If the outcome is bad, I would feel guilty.

    (f) I do not care about the result so I let someone else do it.

    (g) I am unlucky, so I thought we would lose.

    (h) h. Other:_____

17. I did not want to play the number game on the board, because:

    (a) I am not good at math.

    (b) I am good at math but I was scared of getting anxious and forgetting what I know.

    (c) I was scared of losing and making a fool of myself in front of my friends.

    (d) I am good at math but I do not like performing in public.

    (e) Other:_____

3 Instructions[27]

3.1 Instructions for the Individual Risk Task[28]

We are going to play different games with you today. Depending on your decisions in these games, you will earn different amounts of "coupons" [show coupons]. These coupons, as you know, can be used for buying items in many different stores (e.g. restaurants, grocery stores, stationary shops etc.), that is, they have monetary value. The rewards in our games will correspond to coupons. There will be tokens, where each token corresponds to 1 TL worth of coupons.

Now, in this lecture hour, we will have three different games. At the end of the lecture, we will select one randomly (by drawing the number 1, 2 or 3 out of a bag). Whichever game is selected, you will earn rewards according to the decisions you made in that game and only that game. That is, your rewards from different games do not accumulate. For example, suppose someone earned 8 tokens in Game 1, 5 in Game 2 and 6 in Game 3. If Game 2 is drawn, what does that person get at the end? Only 5. So please consider the games in isolation and make all your decisions very carefully.

Now we will explain the rules of the first game. Your rewards will be based on the decisions you make in this game, if this game is randomly selected at the end of the lecture. Please be very quiet while the rules are being explained. If you have a question, please raise your hand. Also note that there are no right or wrong decisions in the games we will play today. In this game, each one of you will have 5 tokens. Each token corresponds to 1 TL (worth of coupons). For example, one token corresponds to 1 TL, two tokens correspond to 2 TL, three tokens to 3 TL etc. How many tokens you have will determine how many coupons you will get at the end of the game.

Now here is a bowl [draw a bowl on the board]. You can put as many tokens as you want

[27] Same instructions were used for the individual risk task and leadership task for the main and supplementary samples. For the private and social confidence task, there were minor changes in the instructions which are provided in detail in Section 3.3.

[28] For brevity, the instructions provided are the ones used for adolescents. As explained in the text, the children's procedures differed in terms of reward medium (gifts rather than coupons) and in the overall difficulty level of the number task (adding up to 100 rather than 1000).

in this bowl. The tokens you do not put in the bowl are yours to keep. What will happen to the tokens you put in the bowl depends on chance. These tokens will either multiply or they will be lost. How? Here is a bag with two balls in it, one of them is yellow and the other one is purple [show bag and balls]. If this game is selected, you will draw a ball without looking. The yellow ball is the good ball: If you draw this ball, the tokens you put in the bowl will triple. The purple ball is the bad ball: If you draw this ball, all of the tokens you put in the bowl will be lost. That is, depending on the color of the ball you draw, you have a 50-50 chance of losing or winning. If this game is selected at the end, you will draw the ball and the color of the ball along with how many tokens you put in the bowl will determine how many coupons you will get. Now we will go over some examples to make sure that everyone understood the rules:

Assume that you did not put any tokens in the bowl [Draw all 5 tokens outside of the bowl, on the board]. Then, since you kept all of your five tokens you get 5 TL for sure. Assume that you put one token in the cup and kept 4 [Draw one token in the bowl on the board, draw the remaining ones outside]. Assume that you draw the purple ball. You lose all of your tokens in the bowl. Since you had kept 4 of your tokens, you get 4 TL. Now assume that you draw the yellow ball, then the one token in the cup triples and becomes three tokens [Draw two more tokens in the cup]. You had already kept 4 tokens, so in total you have 7 tokens. Therefore, you will get 7 TL.

Assume that you put 4 tokens in the bowl and kept one of them [Draw on the board]. Assume that you draw the purple ball. You lose all of your tokens in the bowl. Since you had kept one of your tokens, you get 1 TL. Now assume that you draw the yellow ball. The 4 tokens in the bowl triple and become 12 tokens. You had kept one token, so in total you have 13 tokens which correspond to 13 TL.

Assume that you put all of your tokens in the bowl [Draw on the board]. Assume that you draw the purple ball. Then you lose all the tokens in the bowl and since you did not keep any, you get 0 TL. Now, assume that you draw the yellow ball. Then your tokens in the bowl triple and you get 15 tokens in total, which correspond to 15 TL.

Did you understand the rules of the game? Any questions? [The decision-making will not start until the students answer the following questions correctly]

Assume that you put two tokens in the bowl and keep three tokens. Assume that you draw the yellow ball. How many TLs would you get? [Correct answer is 9]. Assume that you draw the purple ball; how many TLs would you get? [Correct answer is 3].

Assume that you put three tokens in the bowl and you keep two tokens. Assume that you draw the yellow ball. How many TLs would you get? [Correct answer is 11]. Assume that you draw the purple ball; how many TLs would you get? [Correct answer is 2].

Now, each one of you will get a decision sheet. You will mark the number of tokens that you want to put in the bowl on your decision sheet. If this game is selected, the rewards you will get will be determined based on this decision and the color of the ball that you draw. Make your decision quietly and do not show your decision sheet to anyone. [Decision sheets are distributed, students write their names and make their decisions, sheets are collected]

3.2 Instructions for the Leadership Task

Now we will play our second game. If this game is selected at the end of the lecture, you will earn rewards based on the rules we will explain now. This game is similar to the one you just played, but there are some differences. In this game, everyone will be assigned to a 3-person group. We determined your groups randomly, before coming here. For now, you will not know who is in your group. The decision to be made in this game, again, is to decide how many of the 5 tokens to put in the risky bowl, and how many to keep. However, this time, a single person will make a decision for all 3 people in the same group. How many gifts each person gets will be determined according to the decision made by this person. Everyone in the same group will get the same number of gifts. For example, suppose that the person making the decision on behalf of the group put 2 tokens in the bowl (and kept 3). If the yellow ball is drawn, each person in the group gets 9 gifts. If the purple ball is drawn, each person in the group gets 3 gifts. Notice that individuals in a group do not share the gifts. Each person in the same group individually and separately receives the same amount of gifts. Another

example: Suppose the person making the decision on behalf of the group put 5 tokens in the bowl. If the yellow ball is drawn, everyone in the group individually gets 15 gifts. If the purple ball is drawn, no one in the group gets any gifts. Is this clear? Any questions?

Now, who will be the person deciding on behalf of the 3 people in the group? First off, each of you will individually answer the question of "Would you like to be the one making the decision on behalf of your group?". If you want to be the decision-maker, you will mark YES on your decision sheets. If you do not want to be the decision-maker, you will mark NO. Each of the three people in the same group will have said either YES or NO. If only one person in the group said "Yes, I'd like to be the decision-maker", that person will make the decision for the group [Write on the board as an example, NO, YES, NO. Ask the students: who decides? Person 1, 2 or 3?]. Now, if more than one person in the group said yes, we will randomly select one of the people who said yes. That person will make the group decision [Write on the board as an example: YES, NO, YES. Who decides? Either Person 1 or Person 3 will make the decision, based on a random draw]. Now, suppose nobody wanted to make the decision in the group. In that case, we will randomly select one person among the three. That person will make the decision on behalf of the group [Write on the board as an example: NO, NO, NO. Person 1, 2 or 3 will make the decision, based on a random draw]. Is this clear? Any questions?

Now, how many gifts will the group members get? This depends on how many tokens the decision-maker puts in the bowl. For example, suppose the person who makes the group decision put 3 tokens in the bowl and the yellow ball is drawn. In that case, everyone in the group gets 11 gifts each. If the purple ball was drawn, then everyone in the group gets 2 gifts. [Ask children] If the group decision-maker put 2 tokens and the yellow ball is drawn, how many gifts does each person in the group get? 9. If the purple ball is drawn? 3.

The group decision-maker will also draw the ball on behalf of the group. Based on the group decision-maker's decision and the ball he/she draws, the payoffs of everyone in the group will be determined. If he/she draws the purple ball, everyone in the group loses the tokens put in the bowl. If he/she draws the yellow ball, the tokens in the bowl are tripled

for everyone. You do not know who you are in a group with now, but after decisions are made and when rewards are being determined, everyone will end up knowing whom they were in a group with, who was the decision-maker and what decision he/she made as well as the outcome.

[Distribute decision sheets] Now, everyone please write your name on your decision sheets and mark whether you would like to be the person making the decision on behalf of your group. OK? Did everyone make their selection?

OK. Now remember, in fact it is possible for anyone to be selected as the decision-maker. Why? Because even if someone said no, if nobody in their group wanted to make the decision, he/she may need to do it. This is why each person, regardless of whether they said YES or NO, will answer the question of "if you become the decision-maker for the group, how many tokens will you put in the bowl?". Now please turn your decision sheets over. There you will again see choices from 0 to 5. Mark your answer. Don't forget: if you are selected as the decision-maker, the decision you are making now will determine the rewards for everyone in your group! [29]

3.3 Instructions for the Private and Social Confidence Tasks

Now, in this lecture we are going to play a different game, a number puzzle. In this game the goal is to find numbers that add up to 1000 [Give examples on the below number grid, using different markers for different number pairs adding up to 1000]. Now, before we start the main game, we will have an initial task. In this task the goal is to find as many number pairs as possible in a large grid (within 2 minutes). For this part of the game, you will be rewarded 0.5 TL for each pair of numbers you find, adding up to 1000 [Sheets are distributed and collected when time is up].

[29] After this game, there was a 3rd and final game, which included group decision-making with communication (implemented for another research question). Since that part always came after the leadership task, we do not expect it influence any behaviors we report on in this paper.

300	523	420	807
215	578	477	355
645	193	700	785

Now, in the main part of the game you will again find numbers. However, we have two types of number task here. We have the 4-TL task, in which the reward is 4 TL if you are successful. If you fail, you get 0 TL [Write on board]. We have another type of number task too. In the 1-TL task, the reward is 1 TL if you are successful. If you fail you get 0 TL [Write on board]. But why does one task pay more than the other in the case of success? Because that task is more difficult than the other. In both type of task, the goal is to find at least 3 numbers that add up to 1000, within 1 minute 45 seconds. However, in the 4-TL task the grid is larger and there are less pairings [show], while in the 1-TL task the grid is smaller.

[FOR THE MAIN SAMPLE] Now, there are two possibilities about how the game will be played–individually at your desk, or on the board. If Game Individual is randomly drawn [write on board], everyone will do the number task on their own, at their desk. If Game Board is randomly drawn, a single person will do the task on the board, in front of everyone. The person who will be called to the board will be randomly selected. Now we ask you: if the game is to be played individually and privately, do you prefer to do the 4-TL, more difficult task or the 1-TL, less difficult task? Now we will distribute decision sheets and you will mark your decision along with your name. [Distribute sheets] OK, now please turn the sheet over. If the game is to be played on the board and you are the one selected to do the task on the board, do you prefer to do the 4-TL, more difficult task on the board, or the 1-TL, less difficult task on

the board, or do you prefer not to go on the board at all? Please select one on your decision sheet. [Collect sheets]. Now, we will have a draw. [Write Game Individual, Game Board on two pieces of paper, put in a bag, call a class representative to draw]. (If Game Individual is drawn) Everyone will do their preferred task at their desk, privately. Now, we will distribute both tasks, the 4-TL and the 1-TL task to everyone. Based on what you marked on your decision sheet, take the one that you will do, and leave the other one. Is everyone ready? Start. [The game is timed]. (If Game Board is drawn) Now, we will randomly select one person using your decision sheets. If this person wanted to do the task on the board, he/she will do his/her preferred task on the board. If not, we will draw another person.

[FOR THE SUPPLEMENTARY SAMPLE] Now, there are two possibilities about how the game will be played–individually at your desk, or on the board. With some chance you will play the game of your choice and with some chance you will play Game Board, i.e. you will do the task on the board, in front of everyone. Now, if the game is played individually at your desk [Game Individual], do you prefer to do the 4-TL, more difficult task or the 1-TL, less difficult task? If Game Individual is drawn, with some chance we will implement your choice, with some chance you will do the 4-TL task. Now we will distribute decision sheets and you will mark your decision along with your name. [Distribute sheets] OK, now please turn the sheet over. If the game is to be played on the board and you are the one selected to do the task on the board, do you prefer to do the 4-TL, more difficult task on the board, or the 1-TL, less difficult task on the board, or do you prefer not to go on the board at all? Again, if Game Board is drawn, with some chance we will implement your choice, with some chance you will do the 4-TL task. Please select one on your decision sheet. [Collect sheets]. [Write Game Individual on one piece of paper and Game Board on nine pieces of paper, put in a bag, call a class representative to draw]. (If Game Individual is drawn, determine whether their choice will be implemented or they will play the 4-TL task). [If their choice is implemented] Everyone will do their preferred task at their desk, privately. Now, we will distribute both tasks, the 4-TL and the 1-TL task to everyone. Based on what you marked on your decision sheet, take the one that you will do, and leave the other one. Is everyone ready? Start. [The

game is timed]. [If they are imposed to play 4-TL task]Everyone will do the 4-TL task at their desk, privately. Now, we will distribute the 4-TL task to everyone. Is everyone ready? Start. [The game is timed]. (If Game Board is drawn, determine whether their choice will be implemented or they will play the 4-TL task) [If their choice is implemented] Now, we will randomly select people using your decision sheets. If you wanted to do the task on the board, you will do your preferred task on the board. If not, we will draw another person. [If they are imposed to play the 4-TL task] Now, we will randomly select people. If your name is called you will do the 4-TL task on the board.

4 Extra Tables

Table A.1: Gender Gap in Leadership Decision for Children

	Children										
	(1) Leader	(2) Leader	(3) Leader	(4) Leader	(5) Leader	(6) Leader	(7) Leader	(8) Leader	(9) Leader	(10) Leader	(11) Leader
Male	-0.006	-0.009	-0.014	-0.014	-0.023	-0.014	-0.025	-0.023	-0.028	-0.022	-0.023
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Math Ability		0.008	0.011	0.010	0.001	-0.004	-0.004	-0.005	-0.006	-0.006	-0.007
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance			0.033*	0.033*	0.030*	0.029*	0.029*	0.029*	0.029*	0.029*	0.029
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Private Self-Confidence				0.010	-0.006	-0.009	-0.006	-0.010	-0.009	-0.010	-0.008
				(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Social Confidence					0.150***	0.148***	0.151***	0.150***	0.155***	0.155***	0.155***
					(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self Reported Grit						0.035*	0.047*	0.046*	0.041*	0.041*	0.040*
						(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Self Reported Gender Roles							0.025	0.025	0.028	0.027	0.027
							(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Class Size								-0.001	-0.002	-0.002	-0.002
								(0.00)	(0.00)	(0.00)	(0.00)
High SES									0.016	0.016	0.018
									(0.05)	(0.05)	(0.05)
Low SES									-0.079*	-0.075*	-0.078*
									(0.04)	(0.04)	(0.04)
Age										-0.058	-0.059
										(0.04)	(0.04)
Household Size											0.019
											(0.02)
Observations	769	769	769	769	769	769	769	769	769	769	769
R-Square	0.000	0.001	0.007	0.007	0.029	0.036	0.038	0.039	0.047	0.051	0.053
Adjusted R-Square	-0.001	-0.002	0.003	0.002	0.023	0.028	0.029	0.029	0.035	0.038	0.038

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.2: Gender Gap in Leadership Decision for Adolescents

	Adolescents										
	(1) Leader	(2) Leader	(3) Leader	(4) Leader	(5) Leader	(6) Leader	(7) Leader	(8) Leader	(9) Leader	(10) Leader	(11) Leader
Male	0.185***	0.175***	0.183***	0.168***	0.113**	0.126**	0.124**	0.119**	0.121**	0.124**	0.122**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Math Ability		0.046**	0.048**	0.031	0.012	0.004	0.004	0.005	0.004	0.004	0.004
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance			0.055***	0.053***	0.042**	0.039**	0.039**	0.038**	0.038**	0.038**	0.038**
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Private Self-Confidence				0.136**	0.101*	0.091*	0.091*	0.098*	0.100*	0.098*	0.097*
				(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Social Confidence					0.254***	0.247***	0.246***	0.252***	0.255***	0.254***	0.253***
					(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self Reported Grit						0.057***	0.057***	0.059***	0.061***	0.061***	0.060***
						(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Self Reported Gender Roles							0.003	0.007	0.009	0.008	0.007
							(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Class Size								-0.005	-0.005	-0.006	-0.005
								(0.01)	(0.01)	(0.01)	(0.01)
High SES									-0.144	-0.145	-0.148
									(0.12)	(0.12)	(0.12)
Low SES									-0.066	-0.068	-0.064
									(0.04)	(0.04)	(0.04)
Age										-0.020	-0.021
										(0.03)	(0.03)
Household Size											-0.031
											(0.04)
Observations	625	625	625	625	625	625	625	625	625	625	625
R-Square	0.034	0.043	0.056	0.069	0.127	0.139	0.140	0.144	0.148	0.149	0.150
Adjusted R-Square	0.033	0.040	0.051	0.063	0.120	0.131	0.130	0.133	0.134	0.134	0.133

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.3: Gender Gap in Leadership Decision for Children (Girls)

	Children (Girls)									
	(1) Leader	(2) Leader	(3) Leader	(4) Leader	(5) Leader	(6) Leader	(7) Leader	(8) Leader	(9) Leader	(10) Leader
Math Ability	0.012	0.016	0.014	0.002	-0.005	-0.002	-0.002	-0.002	-0.002	-0.005
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Risk Tolerance		0.043**	0.042**	0.040**	0.037*	0.036*	0.037**	0.038**	0.038**	0.038**
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Private Self-Confidence			0.026	0.010	0.001	0.011	0.004	0.004	0.005	0.015
			(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Social Confidence				0.175***	0.171***	0.172***	0.171***	0.174***	0.173***	0.174***
				(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Self Reported Grit					0.035	0.058*	0.057*	0.055*	0.055*	0.055*
					(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles						0.049*	0.049*	0.050*	0.050*	0.051*
						(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class Size							-0.003	-0.003	-0.003	-0.004
							(0.00)	(0.00)	(0.00)	(0.00)
High SES								-0.039	-0.037	-0.037
								(0.07)	(0.07)	(0.07)
Low SES								-0.058	-0.054	-0.061
								(0.07)	(0.07)	(0.07)
Age									-0.053	-0.060
									(0.06)	(0.06)
Household Size										0.039
										(0.03)
Observations	367	367	367	367	367	367	367	367	367	367
R-Square	0.002	0.012	0.013	0.046	0.053	0.062	0.064	0.070	0.072	0.078
Adjusted R-Square	-0.001	0.006	0.004	0.036	0.040	0.046	0.046	0.046	0.046	0.049

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.4: Gender Gap in Leadership Decision for Children (Boys)

	Children (Boys)									
	(1) Leader	(2) Leader	(3) Leader	(4) Leader	(5) Leader	(6) Leader	(7) Leader	(8) Leader	(9) Leader	(10) Leader
Math Ability	0.005	0.006	0.007	0.001	-0.004	-0.004	-0.004	-0.006	-0.005	-0.005
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Risk Tolerance		0.024	0.025	0.022	0.021	0.021	0.021	0.022	0.022	0.022
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Private Self-Confidence			-0.004	-0.019	-0.018	-0.018	-0.019	-0.023	-0.026	-0.026
			(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Social Confidence				0.123*	0.124*	0.125*	0.124*	0.136*	0.137*	0.137*
				(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Self Reported Grit					0.033	0.035	0.035	0.026	0.025	0.025
					(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles						0.004	0.004	0.009	0.008	0.008
						(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Class Size							-0.000	-0.001	-0.001	-0.001
							(0.00)	(0.00)	(0.00)	(0.00)
High SES								0.067	0.064	0.066
								(0.06)	(0.06)	(0.06)
Low SES								-0.107	-0.104	-0.104
								(0.06)	(0.06)	(0.06)
Age									-0.059	-0.060
									(0.04)	(0.04)
Household Size										0.009
										(0.02)
Observations	402	402	402	402	402	402	402	402	402	402
R-Square	0.000	0.003	0.004	0.017	0.023	0.024	0.024	0.047	0.052	0.052
Adjusted R-Square	-0.002	-0.002	-0.004	0.007	0.011	0.009	0.007	0.025	0.028	0.026

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.5: Gender Gap in Leadership Decision for Adolescents (Girls)

	Adolescents (Girls)									
	(1) Leader	(2) Leader	(3) Leader	(4) Leader	(5) Leader	(6) Leader	(7) Leader	(8) Leader	(9) Leader	(10) Leader
Math Ability	0.034	0.033	0.006	-0.028	-0.039	-0.032	-0.032	-0.032	-0.033	-0.034
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Risk Tolerance		0.027	0.025	0.019	0.020	0.020	0.021	0.021	0.023	0.023
		(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Private Self-Confidence			0.168**	0.097	0.092	0.099	0.105	0.098	0.095	0.094
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Social Confidence				0.341***	0.332***	0.318***	0.320***	0.320***	0.316***	0.315***
				(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Self Reported Grit					0.056	0.066*	0.068*	0.072*	0.072*	0.071*
					(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles						0.040	0.043	0.046	0.043	0.042
						(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Class Size							-0.003	-0.003	-0.004	-0.004
							(0.01)	(0.01)	(0.01)	(0.01)
High SES								-0.416***	-0.424***	-0.422***
								(0.09)	(0.09)	(0.09)
Low SES								-0.032	-0.037	-0.033
								(0.07)	(0.07)	(0.07)
Age									-0.035	-0.036
									(0.06)	(0.06)
Household Size										-0.027
										(0.07)
Observations	279	279	279	279	279	279	279	279	279	279
R-Square	0.004	0.007	0.030	0.122	0.134	0.139	0.140	0.146	0.148	0.149
Adjusted R-Square	0.001	0.000	0.019	0.109	0.118	0.120	0.118	0.117	0.116	0.113

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.6: Gender Gap in Leadership Decision for Adolescents (Boys)

	Adolescents (Boys)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Leader	Leader	Leader	Leader	Leader	Leader	Leader	Leader	Leader	Leader
Math Ability	0.054**	0.060***	0.050**	0.038*	0.031	0.029	0.031*	0.029	0.029	0.031
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance		0.076***	0.075***	0.063***	0.057***	0.056***	0.053**	0.053**	0.053**	0.052**
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Private Self-Confidence			0.111	0.101	0.086	0.088	0.093	0.096	0.096	0.095
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)
Social Confidence				0.196***	0.190***	0.193***	0.201***	0.204***	0.204***	0.203***
				(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Self Reported Grit					0.057*	0.050	0.051	0.053	0.053	0.052
					(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles						-0.027	-0.022	-0.021	-0.021	-0.021
						(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Class Size							-0.007	-0.007	-0.007	-0.007
							(0.01)	(0.01)	(0.01)	(0.01)
High SES								-0.118	-0.118	-0.124
								(0.14)	(0.14)	(0.13)
Low SES								-0.094	-0.094	-0.088
								(0.07)	(0.07)	(0.07)
Age									-0.002	-0.003
									(0.03)	(0.03)
Household Size										-0.044
										(0.05)
Observations	346	346	346	346	346	346	346	346	346	346
R-Square	0.014	0.041	0.049	0.089	0.103	0.106	0.115	0.123	0.123	0.125
Adjusted R-Square	0.011	0.035	0.041	0.078	0.090	0.090	0.096	0.099	0.096	0.096

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.7: Social Confidence (Board Choice) in Leadership Decision for Children

	Children									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Board	Board	Board	Board	Board	Board	Board	Board	Board	Board
Male	0.079**	0.059*	0.055	0.056	0.058	0.072**	0.072**	0.074**	0.074**	0.074**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Math Ability		0.066***	0.068***	0.058***	0.056***	0.055***	0.055***	0.056***	0.056***	0.056***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance			0.023	0.019	0.019	0.019	0.019	0.018	0.018	0.018
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Private Self-Confidence				0.107***	0.105***	0.101**	0.099**	0.100**	0.100**	0.100**
				(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self Reported Grit					0.011	-0.003	-0.003	-0.000	-0.000	0.000
					(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Self Reported Gender Roles						-0.029	-0.029	-0.030	-0.029	-0.029
						(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Class Size							-0.001	-0.000	-0.000	-0.000
							(0.00)	(0.00)	(0.00)	(0.00)
High SES								0.015	0.015	0.015
								(0.04)	(0.04)	(0.04)
Low SES								0.055	0.055	0.054
								(0.05)	(0.05)	(0.05)
Age									0.002	0.002
									(0.05)	(0.05)
Household Size										0.005
										(0.02)
Observations	769	769	769	769	769	769	769	769	769	769
R-Square	0.008	0.032	0.034	0.046	0.047	0.050	0.051	0.054	0.054	0.055
Adjusted R-Square	0.007	0.029	0.031	0.041	0.041	0.043	0.042	0.043	0.042	0.041

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary board task choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.8: Social Confidence (Board Choice) in Leadership Decision for Children

	Adolescents									
	(1) Board	(2) Board	(3) Board	(4) Board	(5) Board	(6) Board	(7) Board	(8) Board	(9) Board	(10) Board
Male	0.244***	0.225***	0.231***	0.216***	0.222***	0.180***	0.185***	0.185***	0.186***	0.185***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Math Ability		0.090***	0.092***	0.075***	0.071***	0.076***	0.075***	0.075***	0.075***	0.076***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance			0.045**	0.044**	0.042**	0.043**	0.043**	0.044**	0.044**	0.043**
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Private Self-Confidence				0.136**	0.130**	0.131**	0.122**	0.121**	0.119**	0.119**
				(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Self Reported Grit					0.030	0.044	0.041	0.040	0.040	0.040
					(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Self Reported Gender Roles						0.056**	0.051**	0.050**	0.049**	0.049**
						(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Class Size							0.006	0.006	0.006	0.006
							(0.01)	(0.01)	(0.01)	(0.01)
High SES								0.063	0.063	0.060
								(0.13)	(0.13)	(0.13)
Low SES								0.050	0.049	0.053
								(0.05)	(0.05)	(0.05)
Age									-0.012	-0.013
									(0.05)	(0.05)
Household Size										-0.029
										(0.03)
Observations	625	625	625	625	625	625	625	625	625	625
R-Square	0.059	0.092	0.101	0.114	0.117	0.127	0.133	0.135	0.135	0.136
Adjusted R-Square	0.058	0.089	0.097	0.108	0.110	0.119	0.123	0.122	0.121	0.121

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary board task choice. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Table A.9: Cohort Differences in Board Choice

	Children (%)	Adolescents (%)	p-value
No Board	25	53	0.00
Easy Task	32	17	0.00
Difficult Task	43	30	0.00

Table presents the share of children for each possible board choice category (not being on board, solving the easy task on board and solving the difficult task on board) broken down by cohort. p–values for cohort differences for each board choice category are reported.

Table A.10: Gender Differences in Board Choice

	Children			Adolescents		
	Girls (%)	Boys (%)	p-value	Girls (%)	Boys (%)	p-value
No Board	30	21	0.00	67	42	0.00
Easy Task	36	29	0.07	11	22	0.00
Difficult Task	34	50	0.00	22	36	0.01

Table presents the share of children for each possible board choice category (not being on board, solving the easy task on board and solving the difficult task on board) broken down by gender and cohort. p–values for gender differences in each cohort for each board choice category are reported.

Table A.11: Key Variables for Main and Supplementary Sample

	Children			Adolescents		
	Main	Supplementary	p-value	Main	Supplementary	p-value
Math Ability	4.85	2.91	0.00	2.16	1.61	0.00
Risk Tolerance	2.60	2.76	0.36	2.74	2.63	0.49
Private Self-Confidence	0.73	0.78	0.36	0.74	0.53	0.00
Social Confidence	0.75	0.75	0.90	0.47	0.55	0.27

Table presents the means of key variables for the main and the supplementary samples across cohorts. p-values for sample differences in each cohort for each variable are reported.

Table A.12: Leadership Willingness: Oldest Age Group

	(1) Leader	(2) Leader
Male	0.234**	0.198**
	(0.09)	(0.08)
Math Ability	-0.002	-0.009
	(0.03)	(0.03)
Risk Tolerance	0.009	0.015
	(0.04)	(0.02)
Private Self-Confidence	0.092	0.087
	(0.10)	(0.07)
Social Confidence	0.275***	0.288***
	(0.06)	(0.05)
Self Reported Grit	0.072**	0.061***
	(0.03)	(0.02)
Self Reported Gender Roles	-0.018	0.023
	(0.05)	(0.03)
Class Size	-0.003	-0.004
	(0.01)	(0.00)
High SES	-0.105	-0.186
	(0.20)	(0.15)
Low SES	-0.093	-0.059
	(0.06)	(0.05)
Age	0.032	-0.040
	(0.05)	(0.05)
Household Size	0.042	0.004
	(0.08)	(0.06)
Observations	184	374
R-Square	0.203	0.166
Adjusted R-Square	0.147	0.138

Reported estimates are average marginal effects from a linear probability model where the dependent variable is the binary leadership choice. The first specification includes the oldest age category in our adolescent sample and the second specification includes all girls and only boys who belong to the oldest age category in our adolescent sample. The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

Share of Missing Observations

Table A.13: Share of Missing Observations

	Children (%)	Adolescents (%)
Math Ability	17.04	1.60
Risk Tolerance	1.17	0.80
Private Self-Confidence	14.82	0.64
Social Confidence (Board)	14.95	0.48
Self Reported Grit	19.64	13.12
Self Reported Gender Roles	14.30	12.80
Age	7.41	3.84
SES	27.31	2.72
Household Size	5.85	3.20

Table presents the share of missing observations for each covariate separately for children and adolescents.

Main Results without Missing Data Imputation

Table A.14: Gender Gap in Leadership Decision

	Children		Adolescents	
	(1)	(2)	(3)	(4)
	Leader	Leader	Leader	Leader
Male	-0.006	-0.009	0.185***	0.109*
	(0.04)	(0.05)	(0.05)	(0.06)
Math Ability		-0.028		0.005
		(0.02)		(0.03)
Risk Tolerance		0.015		0.030
		(0.02)		(0.02)
Private Self-Confidence		0.064		0.096
		(0.06)		(0.06)
Social Confidence		0.166**		0.233***
		(0.07)		(0.04)
Self Reported Grit		0.076**		0.053**
		(0.03)		(0.02)
Self Reported Gender Roles		-0.051		-0.004
		(0.04)		(0.02)
Class Size		0.001		-0.004
		(0.00)		(0.01)
High SES		0.038		-0.152
		(0.07)		(0.15)
Low SES		-0.022		-0.071
		(0.07)		(0.05)
Age		-0.012		-0.028
		(0.06)		(0.04)
Household Size		0.037		-0.033
		(0.02)		(0.05)
Observations	769	355	625	460
R-Square	0.000	0.052	0.034	0.121
Adjusted R-Square	-0.001	0.018	0.033	0.098

Reported estimates are average marginal effects from logit regressions where the dependent variable is the binary leadership choice (no missing data imputation). The standard errors are clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01.

5 Extra Figures

Figure A.1: Gender Gap in the Willingness to Lead: Supplementary Data

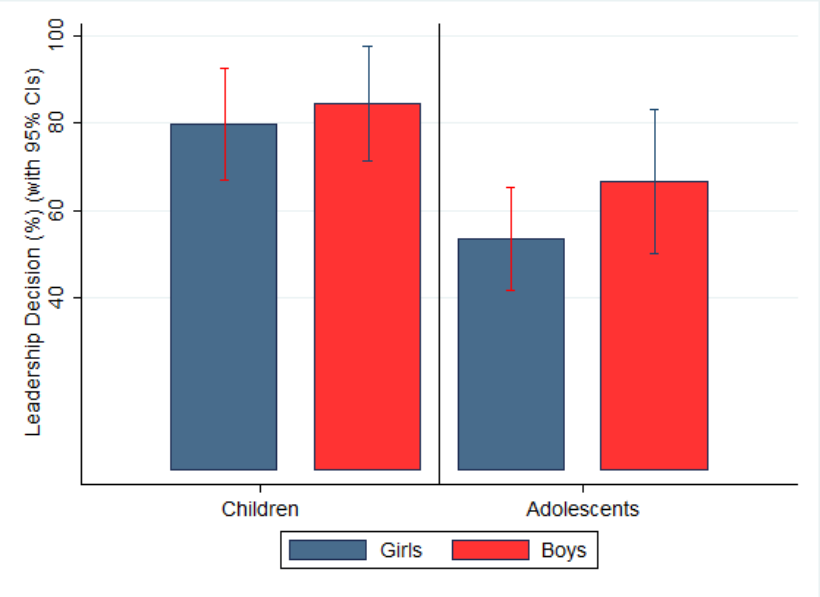

Figure A.2: Math Ability by Gender for Children and Adolescents

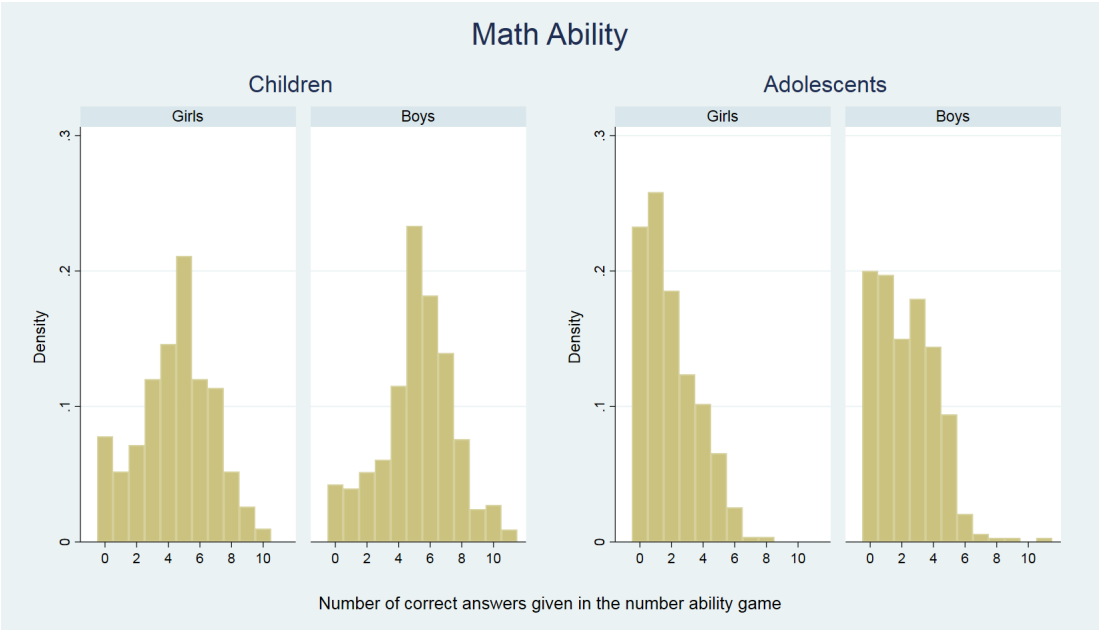

Figure A.3: Risk Tolerance by Gender for Children and Adolescents

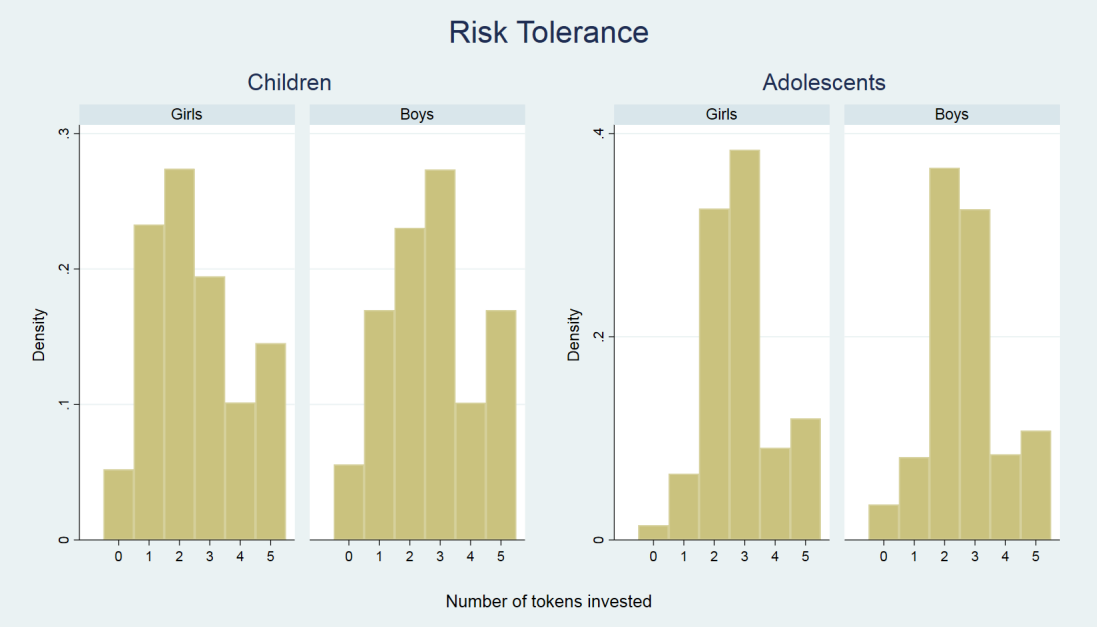

Figure A.4: Self Reported Grit by Gender for Children and Adolescents

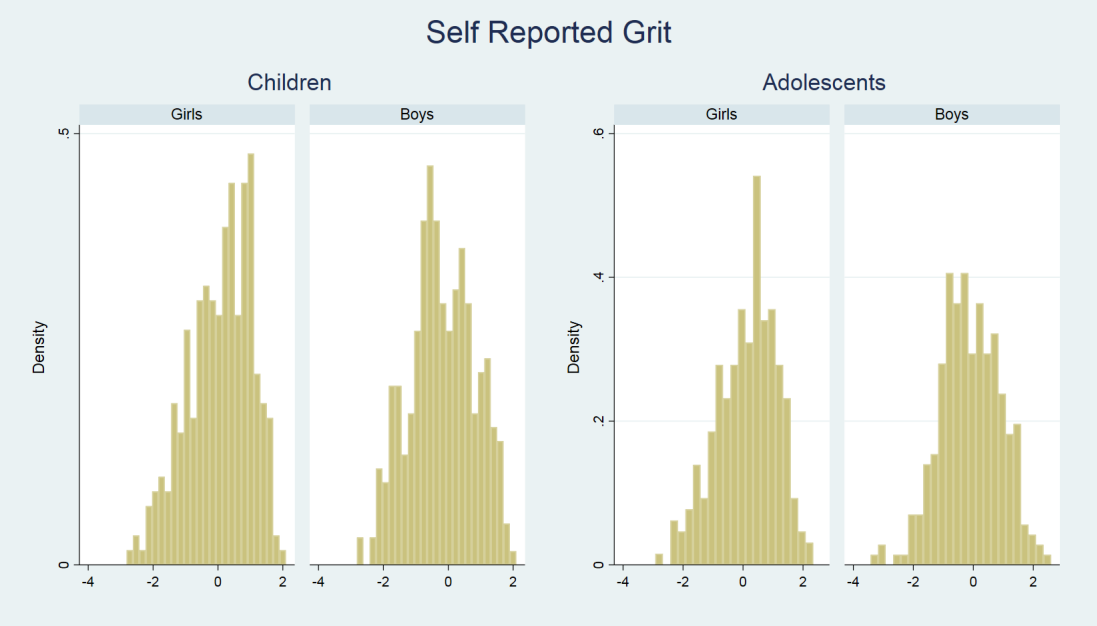

Figure A.5: Self Reported Gender Roles by Gender for Children and Adolescents