

## UNDERSTANDING GENDER DIFFERENCES IN LEADERSHIP\*

*Sule Alan, Seda Ertac, Elif Kubilay and Gyongyi Loranth*

Using data from a large-scale field experiment, we show that while there is no gender difference in 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 willingness to perform a real effort task in public.

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. As one rises in the hierarchy of corporations or in politics, one increasingly needs to take on leadership roles, assuming responsibility for making executive decisions. The stark scarcity of females in leadership positions persists despite much improvement in societal norms and institutional barriers in recent years. For example, at the 2014 G20 summit, only five out of 58 leaders were female. Around the world, only 17% of government ministers, and only 5.2% of S&P 500 chief executive officers (CEOs) are female.<sup>1</sup> 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 to 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, non-governmental organisations (NGOs) and colleges now implement leadership training programmes 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 for 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 decision-making responsibility to the leader. Building decision-making skills and learning how to handle responsibility and accountability for others’ outcomes are in fact

\* Corresponding author: Sule Alan, Department of Economics, European University Institute, 50014 Fiesole, Italy. Email: [salancrossley@gmail.com](mailto:salancrossley@gmail.com)

*This paper was received on 30 January 2017 and accepted on 7 February 2019. The Editor was Frederic Vermeulen.*

The data and codes for this paper are available on the Journal website. They were checked for their ability to replicate the results presented in the paper.

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) and the Science Academy (BAGEP grant); Alan thanks the British Academy for generous support; Kubilay acknowledges funding from the ESRC Research Centre on Micro-Social Change at the University of Essex (ES/L009153/1). 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 to collect the data. All remaining errors are our own.

<sup>1</sup> Inter-Parliamentary Union and UN Women (2015), Catalyst (2019).

major focus points of most leadership training programmes (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 that they will need to make as leaders (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; 2019) 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 article.<sup>2</sup> 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 article we study the evolution of willingness to assume the decision-maker role in a group, which is a major component of leadership, from childhood to adolescence. Using unique data from a large field experiment that involves a sample of children of average age 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 studying 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 in which 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 accountable for other people’s payoffs, which are a fundamental aspect of executive decision making and leadership.<sup>3</sup> 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 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

<sup>2</sup> 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.*, 2016).

<sup>3</sup> 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 behaviour in both strategic and non-strategic contexts (Charness *et al.*, 2007; Sutter, 2009). As related concepts, Bartling *et al.* (2014), Neri and Rommeswinkel (2017) and Ertac *et al.* (2019) study preferences for decision rights, autonomy and power.

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, 2003) 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 that 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 to generate convincing arguments against dissent, or 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 incentivised 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.<sup>4</sup> 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 favour 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 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

<sup>4</sup> 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.

perform a structural estimation exercise in which 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 that 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 favouring men. This article 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 in which 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 article is organised as follows: Section 1 provides the background and experimental design, Section 2 presents the data and discusses the results and Section 3 concludes.

## 1. 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 randomised educational interventions. The experiments that we conducted for the purpose of this article 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 behaviour (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.<sup>5</sup>

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 socio-economic status (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 four years of

<sup>5</sup> 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.

elementary, four years of middle and four 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 send 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 our sample of children is fully comparable to our sample of adolescents.<sup>6</sup>

### 1.1. *The Leadership Task*

Our outcome variable, leadership willingness, is elicited using an incentivised 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 five tokens corresponding to gifts from a gift basket, which they can allocate between risky and a riskless options. 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 three 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 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.<sup>7</sup> Knowing this mechanism,

<sup>6</sup> A statistical comparison of teacher reported (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.

<sup>7</sup> 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

individuals make two decisions: (1) Whether they would like to be the group decision maker; (2) in the event that 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 are not 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.<sup>8</sup>

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 versus undergraduates) may have different levels of comprehension of compounded lotteries (List and Haigh, 2005). Charness *et al.* (2016) provide 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 in which 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.

### 1.2. *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 incentivised 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 in which the goal is to find pairs of numbers in a grid that add up to 100 in elementary schools and 1,000 in middle schools. The task has two versions. The four-token task brings four gift tokens whereas the one-token task brings one 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 three pairs adding up to 100 (or 1,000), within 1.5 minutes. However, the number grid in the four-token task is larger, which is why this task is more difficult. Note that mathematical tasks have been widely used in the

behavioural 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.

<sup>8</sup> Ertac and Gurdal (2012) and Ertac *et al.* (2019) 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.

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 the event that 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 is 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 who 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 one if the student was willing to perform on the board and zero otherwise. The reason for our using the decision to refrain from doing the task altogether is that this is a self-preserving strategy that absolves the individual from any social pressure or potential embarrassment.<sup>9</sup> 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.<sup>10</sup> 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 text.

In order to both familiarise 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 two minutes to find as many pairs as possible that add up to 100 (1,000 for the adolescent sample) in a large number grid. We incentivised this part of the experiment as well by offering a small gift per correct answer.

### 1.3. *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 stationery 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).<sup>11</sup> We implemented

<sup>9</sup> 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.

<sup>10</sup> 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 performing in public. Tables A.9 and A.10 in the Online Appendix document the cohort and gender differences in the choice of doing the task on the board, respectively.

<sup>11</sup> It is common in the literature examining the evolution of economic behaviour 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).

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 or C). In the event of a tie (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 Subsection 2.4.

## 2. Data and Results

In addition to our incentivised social and private self-confidence measures, our data contain a number of other variables. We utilise these variables as potential predictors of leadership choice. One such predictor is risk attitude. As explained in Subsection 1.1, 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 five 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, 2019). 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 behaviour.

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 due not to non-attendance but to 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 result without imputation in the Online Appendix (see Table 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.

### 2.1. 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., maths 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 article: 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 favour 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 Hyde, Fennema and Lamon, 1990; Fryer and Levitt, 2010; Golsteyn and Schils, 2014). Consistently with some of the previous findings in the literature, girls appear to be more risk averse than boys, although

Table 1. *Summary of Experimental Results, Self-reported Measures and Demographics.*

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

*Notes:* 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 1 if willing to be a leader; 0 otherwise. Maths 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 five tokens (privately made, prior to the leadership task). Private self-confidence: binary choice of task difficulty, equals 1 if task is 4TL; 0 otherwise. Social confidence: binary choice of performing the task at the board, equals 1 if willing to perform the task on the board; 0 otherwise. Self-reported grit: standardised summary score constructed using survey questions adapted from the Duckworth grit scale. Self-reported gender roles: standardised 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 in childhood sample and it is self-reported in adolescent sample based on a 1–4 item scale.

this gender difference seems to disappear in adolescence in our sample; see Harbaugh *et al.* (2002), Croson and Gneezy (2009), Cárdenas *et al.* (2012), Sutter *et al.* (2013), 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 favour 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 favouring boys largely predicts the emerging gender gap in leadership willingness going from childhood to adolescence.

## 2.2. 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,

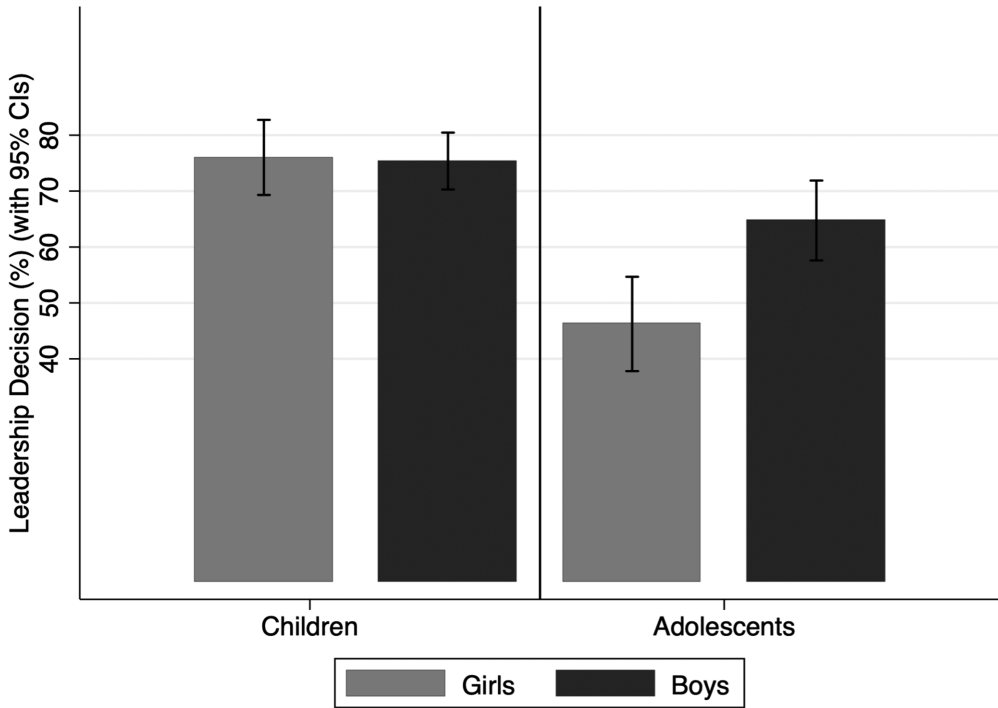


Fig. 1. *Gender Gap in Leadership Decision.*

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%) gender gap emerges going from childhood to adolescence.<sup>12</sup>

The first analysis that 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 with a model without social confidence, adding in social confidence increases  $R^2$  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.<sup>13</sup> Specifically, a one standard deviation

<sup>12</sup> 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.

<sup>13</sup> 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.

Table 2. *Gender Gap in the Leadership Decision.*

|                            | Children         |                    | Adolescents        |                    |
|----------------------------|------------------|--------------------|--------------------|--------------------|
|                            | (1)<br>Leader    | (2)<br>Leader      | (3)<br>Leader      | (4)<br>Leader      |
| Male                       | -0.006<br>(0.04) | -0.023<br>(0.04)   | 0.185***<br>(0.05) | 0.122**<br>(0.05)  |
| Maths ability              |                  | -0.007<br>(0.02)   |                    | 0.004<br>(0.02)    |
| Risk tolerance             |                  | 0.029<br>(0.02)    |                    | 0.038**<br>(0.02)  |
| Private self-confidence    |                  | -0.008<br>(0.04)   |                    | 0.097*<br>(0.05)   |
| Public confidence          |                  | 0.155***<br>(0.04) |                    | 0.253***<br>(0.04) |
| Self-reported grit         |                  | 0.040*<br>(0.02)   |                    | 0.060***<br>(0.02) |
| Self-reported gender roles |                  | -0.027<br>(0.02)   |                    | -0.007<br>(0.02)   |
| Class size                 |                  | -0.002<br>(0.00)   |                    | -0.005<br>(0.01)   |
| High SES                   |                  | 0.018<br>(0.05)    |                    | -0.148<br>(0.12)   |
| Low SES                    |                  | -0.078*<br>(0.04)  |                    | -0.064<br>(0.04)   |
| Age                        |                  | -0.059<br>(0.04)   |                    | -0.021<br>(0.03)   |
| Household size             |                  | 0.019<br>(0.02)    |                    | -0.031<br>(0.04)   |
| Observations               | 769              | 769                | 625                | 625                |
| R <sup>2</sup>             | 0.000            | 0.053              | 0.034              | 0.150              |
| Adjusted R <sup>2</sup>    | -0.001           | 0.038              | 0.033              | 0.133              |

Notes: 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. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

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 analyse 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

Table 3. *Gender Gap in the Leadership Decision.*

|                            | Children           |                  | Adolescents         |                    |
|----------------------------|--------------------|------------------|---------------------|--------------------|
|                            | (1)<br>Girls       | (2)<br>Boys      | (3)<br>Girls        | (4)<br>Boys        |
| Maths ability              | -0.005<br>(0.03)   | -0.005<br>(0.03) | -0.034<br>(0.04)    | 0.031<br>(0.02)    |
| Risk tolerance             | 0.038**<br>(0.02)  | 0.022<br>(0.03)  | 0.023<br>(0.03)     | 0.052**<br>(0.02)  |
| Private self-confidence    | 0.015<br>(0.05)    | -0.026<br>(0.05) | 0.094<br>(0.07)     | 0.095<br>(0.06)    |
| Public confidence          | 0.174***<br>(0.05) | 0.137*<br>(0.07) | 0.315***<br>(0.06)  | 0.203***<br>(0.06) |
| Self-reported grit         | 0.055*<br>(0.03)   | 0.025<br>(0.03)  | 0.071*<br>(0.03)    | 0.052<br>(0.03)    |
| Self-reported gender roles | -0.051*<br>(0.03)  | -0.008<br>(0.03) | -0.042<br>(0.04)    | 0.021<br>(0.02)    |
| Class size                 | -0.004<br>(0.00)   | -0.001<br>(0.00) | -0.004<br>(0.01)    | -0.007<br>(0.01)   |
| High SES                   | -0.037<br>(0.07)   | 0.066<br>(0.06)  | -0.422***<br>(0.09) | -0.124<br>(0.13)   |
| Low SES                    | -0.061<br>(0.07)   | -0.104<br>(0.06) | -0.033<br>(0.07)    | -0.088<br>(0.07)   |
| Age                        | -0.060<br>(0.06)   | -0.060<br>(0.04) | -0.036<br>(0.06)    | -0.003<br>(0.03)   |
| Household size             | 0.039<br>(0.03)    | 0.009<br>(0.02)  | -0.027<br>(0.07)    | -0.044<br>(0.05)   |
| Observations               | 367                | 402              | 279                 | 346                |
| $R^2$                      | 0.078              | 0.052            | 0.149               | 0.125              |
| Adjusted $R^2$             | 0.049              | 0.026            | 0.113               | 0.096              |

Notes: 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. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

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.

Before moving on to what explains the gender gap, it is worthwhile discussing 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; 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, in which both boys and girls are likely to have entered puberty, we still have the result that boys are more willing to become leaders.

### 2.3. Explaining the Emerging Gender Gap in Leadership Willingness

In this subsection, 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 that we examine in earlier sections, by presenting

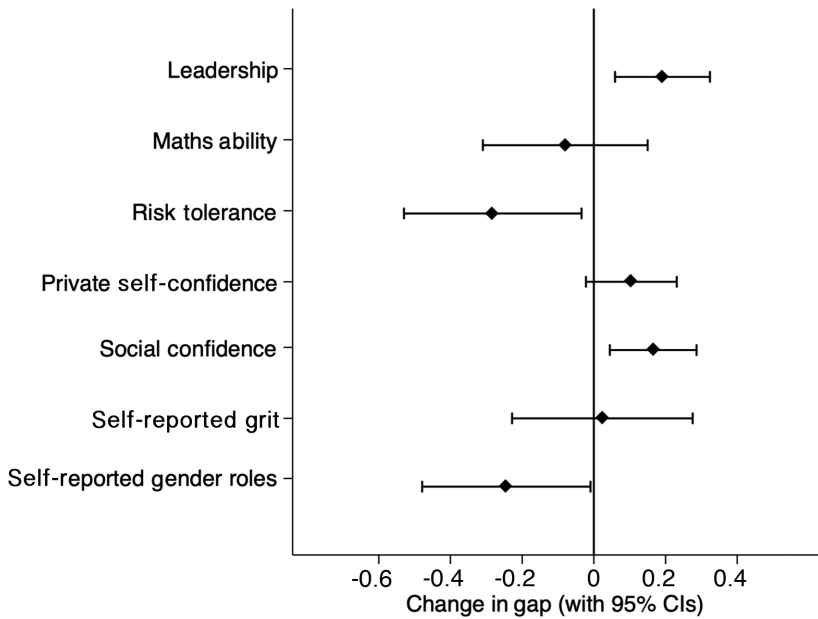


Fig. 2. *Change in Willingness to Lead and Its Determinants from Childhood to Adolescence.*

difference-in-difference estimates of the gender gaps with 95% confidence bands.<sup>14</sup> 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 favour of boys, while the left-hand side depicts those in favour 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 favour of girls, while we do not observe any significant change in gender differences in maths ability or grit. While remaining the same in levels, the contribution of the latter 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

<sup>14</sup> The coefficients plotted are obtained from the empirical model:  $y_i = \alpha + \beta_1 \text{Male} + \beta_2 \text{Elementary} + \beta_3 \text{Male} \times \text{Elementary} + \varepsilon_i$ . The plotted coefficient is  $\beta_3$ , which shows the change in the gender gap going from childhood to adolescence.

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.

#### 2.4. 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 for our taking 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, 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 behavioural aspect of leadership that we focus on, and social confidence is the level of ease with which one can face such social situations. In our conceptualisation, the level of social fear constitutes a reason for individuals shying away from making decisions on behalf of others and determines the extent to which they do so. 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 proportion 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).<sup>15</sup> 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.

<sup>15</sup> 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 maths grade of girls to the average maths 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 maths 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).

Table 4. *Joint Estimation of Leadership Decision and Social Confidence.*

|                            | (1)<br>Children    | (2)<br>Adolescents |
|----------------------------|--------------------|--------------------|
| <i>Leadership</i>          |                    |                    |
| Male                       | -0.041<br>(0.14)   | 0.461***<br>(0.15) |
| Maths ability              | 0.007<br>(0.07)    | 0.065<br>(0.05)    |
| Risk tolerance             | 0.108*<br>(0.06)   | 0.137***<br>(0.04) |
| Self-reported grit         | 0.137*<br>(0.07)   | 0.193***<br>(0.05) |
| Self-reported gender roles | -0.072<br>(0.07)   | -0.053<br>(0.07)   |
| Class size                 | -0.008<br>(0.01)   | -0.011<br>(0.01)   |
| Private self-confidence    | 0.020<br>(0.13)    | 0.333**<br>(0.15)  |
| High SES                   | 0.059<br>(0.16)    | -0.396<br>(0.31)   |
| Low SES                    | -0.218*<br>(0.13)  | -0.147<br>(0.12)   |
| Age                        | -0.181<br>(0.11)   | -0.062<br>(0.07)   |
| Household size             | 0.065<br>(0.05)    | -0.100<br>(0.10)   |
| <i>Social confidence</i>   |                    |                    |
| Male                       | 0.248**<br>(0.11)  | 0.512***<br>(0.16) |
| Maths ability              | 0.186***<br>(0.07) | 0.215***<br>(0.05) |
| Risk tolerance             | 0.063<br>(0.05)    | 0.124**<br>(0.05)  |
| Self-reported grit         | 0.003<br>(0.07)    | 0.119<br>(0.08)    |
| Self-reported gender roles | 0.105<br>(0.07)    | -0.147**<br>(0.06) |
| Class size                 | -0.002<br>(0.00)   | 0.016<br>(0.02)    |
| Private self-confidence    | 0.305***<br>(0.11) | 0.353**<br>(0.15)  |
| High SES                   | 0.046<br>(0.13)    | 0.174<br>(0.36)    |
| Low SES                    | 0.185<br>(0.16)    | 0.161<br>(0.14)    |
| Age                        | 0.003<br>(0.15)    | -0.041<br>(0.12)   |
| Household size             | 0.023<br>(0.06)    | -0.068<br>(0.08)   |
| Observations               | 769                | 625                |

Notes: 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.

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 of the



Table 5. *Gender Gap in Social Confidence (Board Choice).*

|                            | Children          |                    | Adolescents        |                    |
|----------------------------|-------------------|--------------------|--------------------|--------------------|
|                            | (1)<br>Board      | (2)<br>Board       | (3)<br>Board       | (4)<br>Board       |
| Male                       | 0.079**<br>(0.03) | 0.074**<br>(0.03)  | 0.244***<br>(0.05) | 0.185***<br>(0.06) |
| Maths ability              |                   | 0.056***<br>(0.02) |                    | 0.076***<br>(0.02) |
| Risk tolerance             |                   | 0.018<br>(0.02)    |                    | 0.043**<br>(0.02)  |
| Private self-confidence    |                   | 0.100**<br>(0.04)  |                    | 0.119**<br>(0.05)  |
| Self-reported grit         |                   | 0.000<br>(0.02)    |                    | 0.040<br>(0.03)    |
| Self-reported gender roles |                   | 0.029<br>(0.02)    |                    | −0.049**<br>(0.02) |
| Class size                 |                   | −0.000<br>(0.00)   |                    | 0.006<br>(0.01)    |
| High SES                   |                   | 0.015<br>(0.04)    |                    | 0.060<br>(0.13)    |
| Low SES                    |                   | 0.054<br>(0.05)    |                    | 0.053<br>(0.05)    |
| Age                        |                   | 0.002<br>(0.05)    |                    | −0.013<br>(0.05)   |
| Household size             |                   | 0.005<br>(0.02)    |                    | −0.029<br>(0.03)   |
| Observations               | 769               | 769                | 625                | 625                |
| $R^2$                      | 0.008             | 0.055              | 0.059              | 0.136              |
| Adjusted $R^2$             | 0.007             | 0.041              | 0.058              | 0.121              |

Notes: 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. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

same ability level. This may be particularly relevant given the mathematical task, in which girls may experience stereotype threat (Spencer *et al.*, 1999).<sup>16</sup> 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 organised an additional field study and supplemented our main data with a small fresh sample of students, a significant proportion of whom 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.<sup>17</sup> 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

<sup>16</sup> 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.

<sup>17</sup> The probability that the students would be asked to do the task was set at 90%. This ensures that while the decision to perform or not perform on the board is incentivised, a large majority of students would actually be imposed the board task.

Table 6. *Gender Gap in Board Success.*

|              | Children                   |                            | Adolescents                |                            |
|--------------|----------------------------|----------------------------|----------------------------|----------------------------|
|              | (1)<br>Successful on board | (2)<br>Successful on board | (3)<br>Successful on board | (4)<br>Successful on board |
| Main         |                            |                            |                            |                            |
| Male         | 0.116<br>(0.11)            | 0.007<br>(0.10)            | -0.934<br>(1.16)           | -0.193*<br>(0.09)          |
| Observations | 59                         | 56                         | 79                         | 75                         |

*Notes:* 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. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

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 two 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 socio-economic 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 key variables used in the article for the main and supplementary samples. While maths 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 maths 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, it is to provide some evidence on the rationale behind the decisions that we observe.<sup>18</sup>

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), which was 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.<sup>19</sup> These results do not change when we control for private self-confidence, social confidence, risk tolerance and maths ability for the children sample but a 19 percentage point gender gap in favour of girls appears in the

<sup>18</sup> 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.

<sup>19</sup> For one child in elementary school, the performance record is missing. Therefore, we have 59 observations instead of 60 in column 1.

Table 7. *Correlations.*

|                        | Leadership choice |                    |                    |                 | Board task choice           |                             |                             |                             |
|------------------------|-------------------|--------------------|--------------------|-----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|                        | (1)<br>Leader     | (2)<br>Leader      | (3)<br>Leader      | (4)<br>Leader   | (5)<br>Social<br>confidence | (6)<br>Social<br>confidence | (7)<br>Social<br>confidence | (8)<br>Social<br>confidence |
| Fear of embarrassment  | -0.047<br>(0.03)  |                    |                    |                 | -0.092***<br>(0.03)         |                             |                             |                             |
| Assertiveness          |                   | 0.050***<br>(0.02) |                    |                 |                             | 0.026<br>(0.02)             |                             |                             |
| Anxiousness            |                   |                    | -0.087**<br>(0.03) |                 |                             |                             | -0.131***<br>(0.03)         |                             |
| Fear of disappointment |                   |                    |                    | 0.010<br>(0.02) |                             |                             |                             | 0.064*<br>(0.03)            |
| Observations           | 286               | 293                | 292                | 288             | 285                         | 292                         | 291                         | 287                         |

Notes: Coefficients presented are ordinary least squares (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. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

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.

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 incentivised experiments. This survey involves a battery of questions that aim to elicit fear of embarrassment, assertiveness, anxiousness and fear of disappointing others, behaviours and attitudes which are likely to drive both leadership willingness and willingness to do the board task.<sup>20</sup> Using these questions, we construct standardised 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 that we will exploit when we discuss our proposed mechanism via a simple expected utility model that might help in interpreting our results.

#### 2.4.1. *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.<sup>21</sup>

<sup>20</sup> All questions are provided in the Online Appendix.

<sup>21</sup> 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

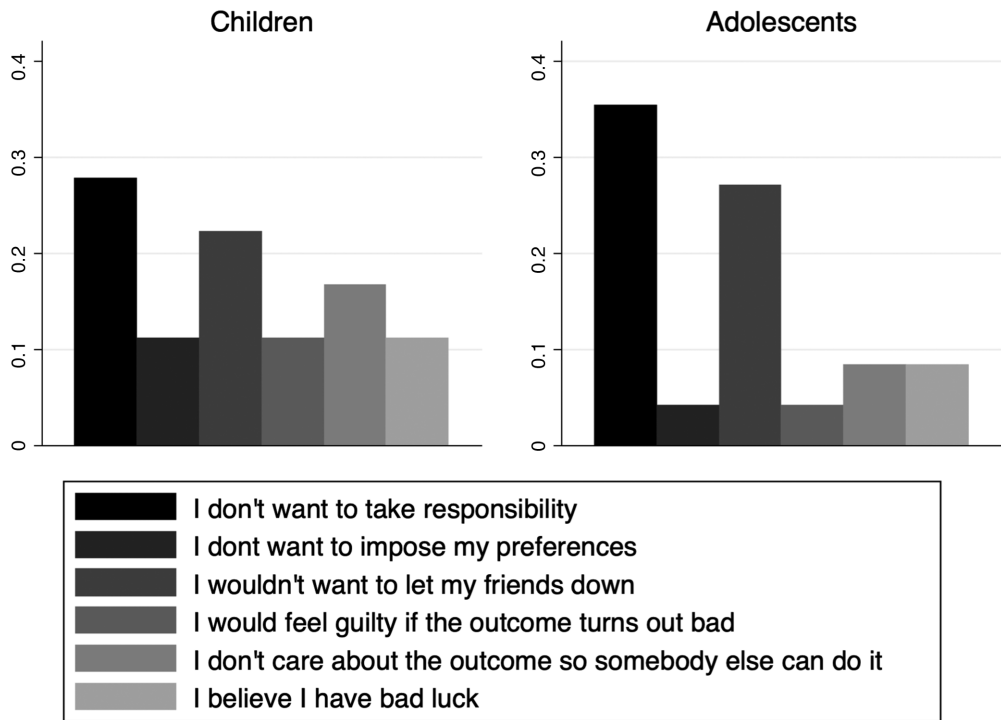


Fig. 3. Frequency of Stated Reasons for Avoiding Leadership.

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 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 that one is not good at maths 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.

#### 2.4.2. 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

declined 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.

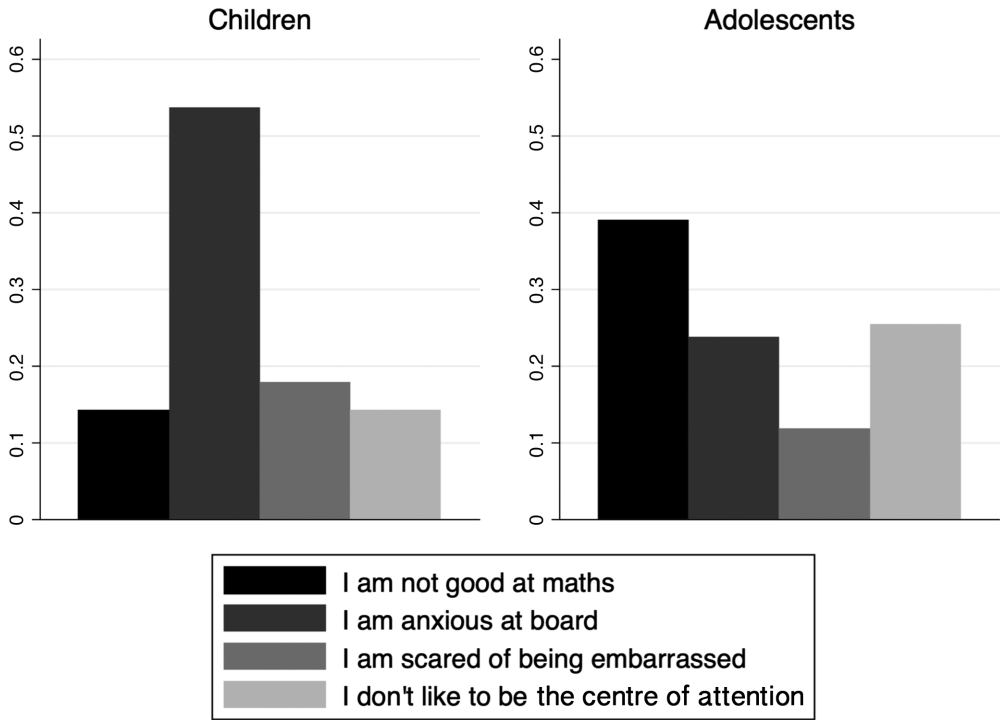


Fig. 4. Frequency of Stated Reasons for Avoiding the Board Task.

bundles. The expected payoff ( $\pi$ ) 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  $\alpha$  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 Constant Relative Risk Aversion 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  $\rho_i$  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  $\rho$  in this specification.

In the standard model, a rational individual  $i$  maximises 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 modelled 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:<sup>22</sup>

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

where  $V(s_i)$  can be modelled 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  $\rho$ , possibly with  $\frac{\partial s}{\partial \rho} > 0$ , on  $p$  with  $\frac{\partial s}{\partial p} < 0$ , and certainly on age, with  $\frac{\partial s}{\partial \text{Age}} > 0$ . 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 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 utilise to measure self-confidence. In our board task, the expected payoff for subject  $i$  is straightforward:

$$E(\text{Payoff}_i) = q_i(\text{success})(R) + (1 - q_i(\text{success}))(0) \geq 0,$$

where  $q(\cdot)$  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-maximising 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

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

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

<sup>22</sup> 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.

Table 8. *Estimated Structural Parameters.*

| Estimated structural parameters                               | Children sample |       | Adolescent sample |       |
|---|-----------------|-------|-------------------|-------|
|   | Girls           | Boys  | Girls             | Boys  |
| Mean coefficient of relative risk aversion ( $\ln(\rho)$ )    | -0.35           | -0.46 | -0.48             | -0.38 |
| SD of coefficient of relative risk aversion ( $\sigma_\rho$ ) | -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  |
| SD 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   |

Notes: Structural parameters are estimated by matching five access points (APs) obtained from the model with those obtained from the main data by minimising the criterion  $\chi = (a^{sim} - a^{data})' \Omega^{-1} (a^{sim} - a^{data})$ , where  $\Omega$  is the variance-covariance matrix of data APs.

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.<sup>23</sup>

The idea is that similar psychological costs can drive different behaviours and choices, as documented empirically in Table 7. Equally plausibly, different types of costs may govern different behaviours 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 parameterised 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

<sup>23</sup> 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.

Table 9. *Goodness of Fit.*

| <i>Panel 1: Children</i>    |             |                |             |                |
|-----------------------------|-------------|----------------|-------------|----------------|
|                             | Girls       |                | Boys        |                |
|                             | Actual data | Simulated data | Actual data | Simulated data |
| Mean G&P allocation         | 2.46        | 2.62           | 2.66        | 2.80           |
| SD 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           |
| <i>Panel 2: Adolescents</i> |             |                |             |                |
|                             | Girls       |                | Boys        |                |
|                             | Actual data | Simulated data | Actual data | Simulated data |
| Mean G&P allocation         | 2.85        | 2.83           | 2.68        | 2.63           |
| SD 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           |

*Notes:* Table presents data and simulated APs. Simulated APs 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.

for others and those related to performance in public contexts. The details of the estimation procedure are given in the Online Appendix.

### 3. Conclusion

Understanding the forces behind self-selection to leadership positions is an important step toward designing effective policies that can mitigate inefficient gender gaps in labour markets as well as in corporate or political decision making. This article 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' behaviour, 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 behaviour 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 organisations.

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 generalisable 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.

*European University Institute and Bilkent University  
Koc University  
University of Essex  
University of Vienna*

Additional Supporting Information may be found in the online version of this article:

### **Online Appendix Replication Package**

### **References**

- Alan, S. and Ertac, S. (2019). 'Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment', *Journal of the European Economic Association*, vol. 17(4), pp. 1147–85.
- Almás, I., Cappelen, A.W., Salvanes, K.G., Sørensen, E.Ø. and Tungodden, B. (2016). 'Willingness to compete: Family matters', *Management Science*, vol. 62(8), pp. 2149–62.
- 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*, vol. 95(4), pp. 1438–43.
- Arbak, E. and Villeval, M.C. (2013). 'Voluntary leadership: motivation and influence', *Social Choice and Welfare*, vol. 40(3), pp. 635–62.
- Babcock, L. and Laschever, S. (2003). *Women Don't Ask: Negotiation and the Gender Divide*, Princeton and Oxford: 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', DICE Discussion Paper, No. 204, <http://www.dice.hhu.de> (last accessed: 11 December 2019).
- Bartling, B., Fehr, E. and Herz, H. (2014). 'The intrinsic value of decision rights', *Econometrica*, vol. 82(6), pp. 2005–39.
- 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*, vol. 335(6068), pp. 582–6.
- Blenko, M.W., Mankins, M.C. and Rogers, P. (2010). *Decide & Deliver: 5 Steps to Breakthrough Performance in your Organization*, USA: Bain&Company, Inc.
- 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*, vol. 166(6), pp. 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*, vol. 1(1), pp. 8–14.
- Cappelen, A.W., Reme, B.A., Sørensen, E.Ø. and Tungodden, B. (2016). 'Leadership and incentives', *Management Science*, vol. 62(7), pp. 1944–53.
- 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*, vol. 83(1), pp. 11–23.
- Catalyst. (2019). 'Pyramid: Women in S&P 500 companies', New York: Catalyst, <https://www.catalyst.org/research/women-in-sp-500-companies/> (last accessed: 11 December 2019).

- Charness, G., Gneezy, U. and Halladay, B. (2016). 'Experimental methods: Pay one or pay all', *Journal of Economic Behavior & Organization*, vol. 131, pp. 141–50.
- Charness, G., Gneezy, U. and Imas, A. (2013). 'Experimental methods: Eliciting risk preferences', *Journal of Economic Behavior & Organization*, vol. 87, pp. 43–51.
- Charness, G. and Jackson, M.O. (2009). 'The role of responsibility in strategic risk-taking', *Journal of Economic Behavior & Organization*, vol. 69(3), pp. 241–7.
- Charness, G., Rigotti, L. and Rustichini, A. (2007). 'Individual behavior and group membership', *The American Economic Review*, vol. 97(4), pp. 1340–52.
- Crosron, R. and Gneezy, U. (2009). 'Gender differences in preferences', *Journal of Economic Literature*, vol. 47(2), pp. 448–74.
- 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*, vol. 92(6), p. 1087.
- Duckworth, A.L. and Quinn, P.D. (2009). 'Development and validation of the short grit scale (grit-s)', *Journal of Personality Assessment*, vol. 91(2), pp. 166–74.
- Elprana, G., Felfe, J., Stiehl, S. and Gatzka, M. (2015). 'Exploring the sex difference in affective motivation to lead', *Journal of Personnel Psychology*, vol. 14(3), pp. 142–52.
- Ertac, S., Gumren, M. and Gurdal, M.Y. (2019). 'Demand for decision autonomy and willingness to take responsibility in risky environments: Experimental evidence', *Journal of Economic Psychology*, 102200.
- Ertac, S. and Gurdal, M.Y. (2012). 'Deciding to decide: Gender, leadership and risk-taking in groups', *Journal of Economic Behavior & Organization*, vol. 83(1), pp. 24–30.
- Ertac, S. and Gurdal, M.Y. (2019). 'Preference communication and leadership in group decision-making', *Journal of Behavioral and Experimental Economics*, vol. 80, pp. 130–40.
- Fryer, R.G. and Levitt, S.D. (2010). 'An empirical analysis of the gender gap in mathematics', *American Economic Journal: Applied Economics*, vol. 2(2), pp. 210–40.
- Füllbrunn, S. and Luhan, W.J. (2015). 'Am I my peer's keeper? Social Responsibility in Financial Decision Making', *Ruhr Economic Paper No. 551*, <https://ssrn.com/abstract=2618439> (last accessed: 11 December 2019).
- Gneezy, U. and Potters, J. (1997). 'An experiment on risk taking and evaluation periods', *The Quarterly Journal of Economics*, vol. 112(2), pp. 631–45.
- 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*, vol. 41, pp. 176–87.
- 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*, vol. 5(1), pp. 53–84.
- Hill, J.P. and Lynch, M.E. (1983). 'The intensification of gender-related role expectations during early adolescence', in (J. Brooks-Gunn and A.C. Petersen, eds.), *Girls at Puberty*, pp. 201–28, Boston, MA: Springer.
- Hyde, J.S., Fennema, E. and Lamon, S.J. (1990). 'Gender differences in mathematics performance: a meta-analysis', *Psychological Bulletin*, vol. 107(2), p. 139.
- Inter-Parliamentary Union and UN Women. (2015). 'Facts and figures: Leadership and political participation', <http://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures> (last accessed: 11 December 2019).
- Kanthak, K. and Woon, J. (2015). 'Women don't run? Election aversion and candidate entry', *American Journal of Political Science*, vol. 59(3), pp. 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*, vol. 118, pp. 318–32.
- Kling, K.C., Hyde, J.S., Showers, C.J. and Buswell, B.N. (1999). 'Gender differences in self-esteem: a meta-analysis', *Psychological Bulletin*, vol. 125(4), p. 470.
- Kosse, F., Deckers, T., Pinger, P.R., Schildberg-Hörisch, H. and Armin, F. (2019). '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?', *Issues in Governance Studies*, vol. 16, pp. 1–20.
- 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*, vol. 102(3), pp. 945–8.
- 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*, vol. 92, pp. 31–46.
- Neri, C. and Rommeswinkel, H. (2017). 'Decision rights: freedom, power, and interference', <https://ssrn.com/abstract=2485107> (last accessed: 11 December 2019).
- New, J. (2014). 'Getting women to run', <https://www.insidehighered.com/news/2014/11/04/female-students-still-scarce-student-government-executive-positions> (last accessed: 11 December 2019).
- Niederle, M. and Yestrumskas, A.H. (2008). 'Gender differences in seeking challenges: The role of institutions', (No. w13922), NBER Working paper, <http://core.ac.uk/download/pdf/6606644.pdf> (last accessed: 11 December 2019).
- Reuben, E., Rey-Biel, P., Sapienza, P. and Zingales, L. (2012). 'The emergence of male leadership in competitive environments', *Journal of Economic Behavior & Organization*, vol. 83(1), pp. 111–7.
- Rivas, M.F. and Sutter, M. (2011). 'The benefits of voluntary leadership in experimental public goods games', *Economics Letters*, vol. 112(2), pp. 176–8.

- Spencer, S.J., Steele, C.M. and Quinn, D.M. (1999). 'Stereotype threat and women's math performance', *Journal of Experimental Social Psychology*, vol. 35(1), pp. 4–28.
- Sutter, M. (2009). 'Individual behavior and group membership: Comment', *The American Economic Review*, vol. 99(5), pp. 2247–57.
- Sutter, M. and Glätzle-Rützler, D. (2015). 'Gender differences in the willingness to compete emerge early in life and persist', *Management Science*, vol. 61(10), pp. 2339–54.
- 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*, vol. 103(1), pp. 510–31.
- Trautmann, S.T. and Vieider, F.M. (2011). 'Social influences on risk attitudes: Applications in economics', in (S. Roeser, R. Hillerbrand, P. Sandin and M. Peterson, eds.), *Handbook of Risk Theory*, pp. 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*, vol. 11(3), pp. 84–94.