Can Simple Advice Eliminate the Gender Gap in Willingness to Compete?

Dany Kessel[*]

Johanna Mollerstrom[†]

Roel van Veldhuizen[‡]

December 2019

Abstract: A recent literature has demonstrated that men and women differ in their willingness to sort into competitive environments, with men being more willing than women to compete. We investigate whether an information intervention that makes individuals aware of this fact, and advises them about potential earnings implications, decreases the gender gap in willingness to compete. We find that this simple information intervention indeed lowers the gender gap, both in a laboratory study at a German university and in a field study with Swedish high school students. Further, we find that the intervention improves the ex-post efficiency of outcomes, in the sense that it leads to high performers being more likely to enter (and win) the tournaments.

Keywords: gender differences; competitiveness; advice; experiment

JEL codes: C91, D03, J16

---

[*]Dany Kessel: Södertörn University and Research Institute of Industrial Economics, Stockholm, Sweden, email: dany.kessel@sh.se. [†]Johanna Mollerstrom: Interdisciplinary Center for Economics Science and the Economics Department, George Mason University, Fairfax, VA, and Research Institute of Industrial Economics, Stockholm, Sweden, email: jmollers@gmu.edu. [‡]Corresponding author: Roel van Veldhuizen: Lund University, PO Box 7082, 22007 Lund, Sweden, email: rrvanveldhuizen@gmail.com.

We thank Jordi Brandts, Thomas Buser, Marie-Pierre Dargnies, Anna Dreber, Emma von Essen, Karin Hederos, Noemi Peter, Christina Rott, Anna Sandberg, Eva Ranehill and participants at the SNEE workshop at the Copenhagen Business School and the Behavioral Workshop at the Swedish Institute for Social Research for very valuable comments. The laboratory experiment reported here was pre-registered at https://www.socialscienceregistry.org/trials/3487.

1. Introduction

Why do men and women differ in their labor market outcomes? A recent literature in experimental economics suggests that these differences may be partly driven by gender differences in psychological traits (c.f. Blau and Kahn, 2017). In particular, a series of experiments demonstrate that men have a higher willingness than women to enter competitions. This finding was first documented by Niederle and Vesterlund (2007), and has since been replicated numerous times (see Croson and Gneezy, 2009; Niederle and Vesterlund, 2011, and Niederle, 2015 for overviews). Since activities like job applications and promotion contests can be thought of as competitive, this may explain some of the observed gender gap in labor market outcomes – and differences in willingness to compete measured in laboratory settings have indeed been found to predict important career choices outside the lab (see e.g. Buser et al, 2014; 2017).

Given that men and women differ in their willingness to compete and that these differences appear to predict important career choices, it is important to see whether these differences can be reduced or eliminated. In this study, we test whether this can be done by implementing a simple information intervention. For this purpose, we run two separate experiments with a similar design: one in the laboratory with German university students and one in the field with Swedish high school students. Both experiments closely follow Niederle and Vesterlund (2007)'s seminal paradigm in which participants solve addition problems and can choose between receiving tournament and piece-rate pay. The main difference with previous studies is that, prior to their final decision, we inform participants of the existence of a gender gap in willingness to compete in previous work, and tell them that men and women could, on average, have earned more by more frequently choosing piece rate or tournament pay respectively. In the laboratory experiment, we also vary the framing of the information intervention to emphasize the role of competitiveness, risk preferences or overconfidence.

These are typically thought to be the main drivers of the gender gap in willingness to compete (Niederle and Vesterlund, 2011; Van Veldhuizen, 2018), and we therefore test whether emphasizing one driver over another changes the effect of the information intervention on willingness to compete.

While our hypothesis going into this work, as also described in the pre-registration, was that the advice (i.e., the information intervention)[1] would reduce or eliminate the gender gap in the willingness to compete, this is not a foregone conclusion. It could, for example, be argued that the information intervention puts an additional emphasis on a gender stereotype (that women compete less than men), and that this means that advice could lead to behavior aligning even more with the stereotype (c.f. Eagly, 1987; Akerlof and Kranton, 2000; Rudman and Glick, 2001). If this is true, then being presented with information that triggers the stereotype could lead men to compete more and women to compete less, increasing the gender gap in willingness to compete. This is, however, not what we find. Instead, we report three main results. First, our simple advice intervention decreases the gender gap in willingness to compete in both experiments. Second, our intervention achieves this while increasing the efficiency of overall choices in the sense that it makes high performers more likely to enter (and win) the competitions. Third, our laboratory data demonstrate that the advice effect is similar regardless of whether the advice emphasizes the role of risk preferences, overconfidence or competitiveness respectively. Overall, the results imply that simple advice can be a powerful and effective tool for reducing the gender difference in the willingness to compete and increasing willingness to compete among high-ability individuals.

Our study relates to other work that focuses on implementing institutional and other changes to reduce the gender gap in willingness to compete. Examples of institutional changes

[1] Our information intervention can be thought of as a minimal type of advice without an explicit directive. In what follows we will therefore use the terms ‘advice’ and ‘information intervention’ interchangeably.

include switching to within-gender competitions (Datta Gupta et al., 2013) or team-based competitions (Dargnies, 2012), instituting affirmative action policies such as gender quota (Balafoutas and Sutter, 2012; Niederle et al, 2013), and changing the nature of the competition from being against others to being against one's own performance (Apicella et al., 2017; Bönte et al., 2017; Klinowski, 2017; Carpenter et al., 2018).

Most closely related to our work are studies that provide relative performance feedback or highly personalized tournament entry advice (Wozniak et al., 2014; Ertac and Szentes, 2011, Brandts et al., 2015). While most of these interventions have proven at least somewhat effective at reducing the gender gap in willingness to compete, they are quite drastic and potentially expensive and may therefore not be practical or even feasible to implement. The policy intervention that we are studying – advising people about the existence of a gender difference in willingness to compete – is, on the other hand, exceptionally simple and cost-efficient, and could be easily implemented in a wide variety of contexts.[2]

Finally, our work also relates to recent studies that seek to understand (and quantify) experimenter demand effects (e.g., De Quidt et al., 2018). In particular, in addition to conveying information about the expected outcomes of different actions, advice may also serve as a cue to what the advisor regards as appropriate behavior, which is similar to an experimenter demand effect. Our design aims to capture both the 'information' channel and the 'demand' channel through which advice may affect behavior.

The work described here originated in two separate projects with two of the authors (Mollerstrom and Van Veldhuizen) designing and implementing the laboratory study in Germany and the third author (Kessel) conducting the field study in Sweden. After learning about the other study, and discussing the results, we decided to write the work up as one joint

[2] More generally, our information intervention also relates to previous work examining the effect of advice on economic decision making in the laboratory and in the field. This includes Schotter and Sopher (2003, 2006, 2007), Chaudhuri, Schotter and Sopher (2009), Çelen, Kariv and Schotter (2010), Hoxby and Turner (20.13), Braun et al., (2014), Castleman and Page, 2015, Carpio and Guadalupe (2018), and Guillen and Hakimov (2018).

paper. We provide this information here as there are some differences between the two studies (for example the fact that the laboratory experiment was pre-registered and the field experiment was not) that may not seem very natural without knowing that the two projects were merged into one only after data collection had taken place.

From here, the paper proceeds as follows. Section 2 describes the design, analysis and results of the laboratory study, and Section 3 does the same for the field study. Section 4 concludes.

2. Laboratory Study

We conducted the laboratory study at the experimental economics laboratory of the Technical University Berlin in Germany. We pre-registered this study in the AEA registry; the pre-analysis plan is reprinted in Online Appendix A, and the analysis below follows this plan. For each session, we invited 28 participants (14 men and 14 women). In total 374 people participated – 185 men and 189 women. Participants' average age was 23. We programmed the experiment in zTree (Fischbacher, 2007) and recruited participants using ORSEE (Greiner, 2015). Average earnings were 14.20 Euros (including a 7 Euro show-up fee). Treatment assignment was done at the individual level.

2.1 Design

Participants were randomly assigned to a computer upon entering the laboratory and notified that the experiment consisted of six parts. They were also informed that out of the first four parts, one would be randomly selected for payment whereas parts five and six were always paid out. We presented all payments in the experiment using experimental currency units

(ECUs), which were converted to Euros at a rate of 10 ECUs per Euro.[3] The experimental instructions for both studies are available in Online Appendix B.

Parts 1-3 of the experiment closely followed the design of Niederle and Vesterlund (2007). In each part, participants had four minutes available to solve problems consisting of adding five randomly generated two-digit numbers. The only difference between the three parts was how participants were paid. In the first part, participants received a piece-rate pay of 5 ECUs per problem solved correctly (and no payment for incorrect answers). In the second part, participants took part in a two-person winner-takes-all tournament where they were paid 10 ECUs per problem if their performance exceeded the performance of a random participant from the same session, and zero otherwise (in case of a tie, participants had a 50% chance of being paid the 10 ECUs per correct problem). They were, however, not informed about their performance and about whether they won or lost until the very end of the experiment. Ahead of the third part, participants were instructed to choose if they wanted to apply the piece-rate or the tournament pay to their performance in part 3. In case they chose tournament pay, their performance in part 3 was compared to the performance of their competitor in part 2, as is the custom in the literature.[4]

Immediately following their initial tournament entry decision in part 3, we presented participants with information about the gender gap in willingness to compete in previous work (while making the initial decision, participants were not aware that they would subsequently receive advice and would be given the opportunity to revise their decision). The aim of the information was to advise women to compete more often and men to compete less.

[3] EUR/USD ≈ 0.89 at the time when the experiment where conducted.

[4] Note that the fact that participants are compared to their competitors' past performance also implies that participants do not need to contemplate whether other participants are reacting to the advice given in a way that could impact that their own tournament entry decision (e.g. if the advice leads to more high quality competitors). Our implementation also differs from Niederle and Vesterlund (2007) in two other ways. First, participants competed against only one competitor instead of three. Second, they had four minutes for the task in each round instead of five. Similar changes implemented in previous work (e.g., Niederle, Segal and Vesterlund 2013; Buser et al. 2014 or Buser, Ranehill and Van Veldhuizen, 2018) did not substantively affect the gender gap in willingness to compete.

We ran three different treatments that framed the information in three different ways. In the treatment "Competitiveness", we informed participants that previous research has found that men compete too much and women compete too little, in the sense that men and women on average would have increased their earnings by competing less and competing more respectively. In the treatment "Risk", we instead informed participants that women would have increased their earnings by taking more risk in similar experiments (whereas men took on too much risk), whereas in the treatment "Confidence" we informed participants that women could have earned more if they had been more confident (and the reverse for men). These statements are based on the findings of Niederle and Vesterlund (2007), where the average woman would have increased her earnings by being more willing to compete, whereas the opposite was true for men. Since tournaments are also riskier than the piece rate and confidence positively predicts tournament entry, these results also imply that women (men) could have increased their earnings by taking on more (less) risk or being more (less) confident. Similar results have been reported in subsequent work, see e.g., Croson and Gneezy (2009), Niederle and Vesterlund (2011) and Niederle (2015) for surveys of the literature. The text we used when giving the advice (translated from German) was:

> *"In many previous experiments similar to this one, it has been documented that women are, on average, [too reluctant to compete/too reluctant to take risks/not confident enough]. This means that in those experiments women, on average, would have earned more money if they had been [more willing to compete/more willing to take risk/more confident].*
>
> *Men, however, in these previous experiments have been found, on average, to be [too eager to compete/too eager to take risks/too confident]. This*

means that men, on average, would have earned more money if they had been [less willing to compete/less willing to take risk/less confident].”

After participants had seen this information, we reminded them of their initial decision of whether to apply the piece-rate or the tournament pay in part 3, asked them whether they would like to confirm or change their decision, and notified them that this decision would be final. After having confirmed or revised their choice, they then moved on to the task, where they had four minutes to solve addition problems, as in parts 1 and 2.

This design allows us to identify the effect of advice as a within-subject treatment effect by comparing tournament entry decisions pre-advice to post-advice. We opted to use a within-subject design in part because it increases statistical power and decreases the minimum detectable effect size; we refer to the power calculations presented in Online Appendix A for more details.

In the fourth part of the experiment, participants were presented with a price list for which they had to make 20 choices between a fixed payment and a lottery. Unbeknownst to the participants, the size of the payment in both options depended on their own performance in part 2. Specifically, each participant’s fixed payment was always equal to 5 ECUs times their performance in part 2, whereas the lottery would pay out either 10 ECUs times performance in part 2 or zero. This procedure ensures that the stakes in part 4 are similar to the stakes in part 3 for each individual in the experiment. The only difference between the 20 items in the choice list was the probability of winning the lottery, which varied from 5% to 100% in increments of 5 percentage points. Participants were informed that if part 4 of the experiment was selected for payment, one of the 20 items would be randomly chosen to be paid out. We use the decisions in part 4 as a measure of risk preferences directly tailored to tournament entry decisions.[5]

In part five of the experiment we asked participants to estimate their ability relative to a comparison sample of 20 participants who took part in an earlier competitiveness experiment at the same laboratory (we created the comparison sample using data from Buser et al., 2018; the comparison sample consisted of 20 individuals drawn randomly from their data). Specifically, we asked participants to indicate their rank on a scale from 1 (better than everyone in the comparison sample) to 21 (worse than everyone in the comparison sample). We informed participants that they would earn either 20 ECUs or 0 ECUs in this part, and that they would maximize their earnings by truthfully reporting their rank. We ensured that this was the case by incentivizing responses using the crossover method (Karni, 2009; Mobius, et al, 2013). A more detailed description of the crossover mechanism and some of its properties is presented in Online Appendix D.

In the sixth part of the experiment, we again elicited participants' risk preferences, this time using the investment game (Gneezy and Potters 1997; Charness and Gneezy 2012). Participants were given an endowment of 10 ECUs and were asked to allocate it between two options. The safe option returned the money invested at a rate of 1:1, whereas the risky investment returned the money at a rate of 2.5:1 with a 50% chance, but would otherwise not return any money. The more risk averse the participant, the less she should invest.

After part six had ended, participants saw an overview of their earnings from the experiment. They subsequently went through a questionnaire that elicited their age, gender and field of study. We also elicited their risk preferences and self-confidence on Likert-scales, and included several questions asking them for their reasons for making a choice in part 3, as well

[5] Part 4 is based on Van Veldhuizen (2018) and decisions in part 4 can also be used as a way to differentiate between risk preferences, competitiveness and confidence as the drivers of the gender gap in willingness to compete. See Online Appendix C for additional details.

as asking them what reasons they thought would play a role in others' decisions in part 3. We then paid participants individually and dismissed them from the experiment.

2.2 Results

Table 1 presents pre-advice choices and outcomes, separated by gender. In line with previous research, we see that men are more likely to choose the tournament scheme in part 3 prior to receiving advice (54 percent of men do so, versus 25 percent of women). Men score slightly higher on the task in all three parts, though the difference is only significant at the 10% level in part 2. Also in line with previous research, we find that men are more confident (they report lower ranks) and take significantly more risk in both the price list task and in the investment game.

Assuming that a participant's performance in part 2 is a good predictor for how they would have performed in the competitive setting in part 3, it is possible to calculate the expected pay-off of tournament and piece rate pay for each participant. We can then define the fraction of participants that "should compete" as the fraction of participants who would maximize their expected payment by choosing tournament pay. For men, the share is similar to the actual share, whereas for women it is much higher (row 4 of Table 1). Indeed, 36% of women would benefit from changing their decision to competing after receiving advice (row 6), whereas only 9% of women would benefit from changing in the opposite direction (row 7). For men, the corresponding percentages are 18% and 15% respectively, suggesting that men in our study would not, on average, benefit from changing their decisions about whether or not to compete. In other words, while women do indeed "compete too little", in contrast to Niederle and Vesterlund (2007) men in our laboratory study do not "compete too much" even prior to receiving advice, and would not, on average, increase their earnings by switching from tournament to piece rate pay.

Table 1: Pre-advice Results for the Laboratory Study

	Men	Women	Difference
(1) Score, Part 1	7.124	6.730	0.394
	(3.246)	(3.218)	(0.334)
(2) Score, Part 2	8.465	7.873	0.592+
	(3.452)	(3.223)	(0.345)
(3) Score, Part 3	8.876	8.492	0.384
	(3.366)	(3.361)	(0.348)
(4) Should compete (based on part 2 score)	0.568	0.529	0.038
	(0.497)	(0.500)	(0.052)
(5) Compete, before advice	0.535	0.254	0.281***
	(0.500)	(0.436)	(0.049)
(6) Should compete but does not, before advice	0.184	0.360	-0.176***
	(0.388)	(0.481)	(0.045)
(7) Should not compete but does, before advice	0.151	0.085	0.067*
	(0.359)	(0.279)	(0.033)
(8) Number of risky choices, part 4	7.076	5.905	1.171**
	(3.316)	(3.564)	(0.356)
(9) Beliefs about rank, part 5	8.000	10.444	-2.444***
	(3.907)	(3.854)	(0.401)
(10) Investment, part 6	7.665	6.122	1.543***
	(3.050)	(3.120)	(0.319)
Observations	185	189	374

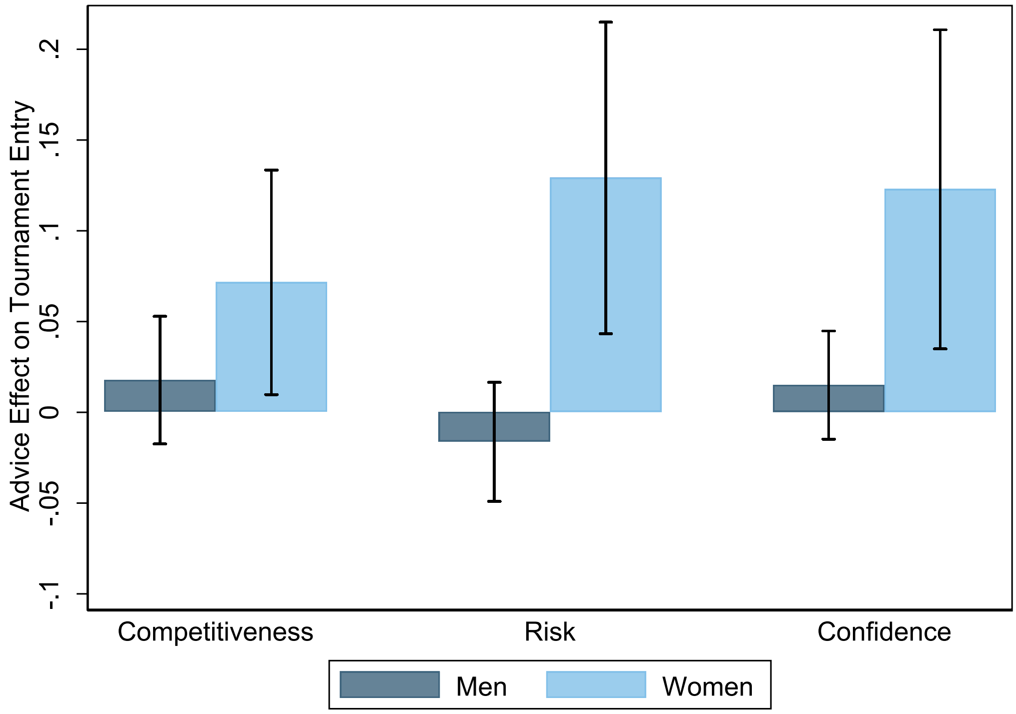
Notes: The data are pooled for the three treatments (there were no differences between the treatments pre-advice). Columns "Men" and "Women" show averages across the three treatments (standard deviations in parentheses). The column "Difference" shows the gender difference (standard errors in parentheses). "Should compete" is the fraction of participants who would have maximized their payment by competing based on their part 2 score. The next two rows report the fraction of participants who could have maximized their payment by choosing tournament (piece rate) but chose the piece rate (tournament) instead. Number of Risky Choices (Part 4) is the number of times the risky option was chosen in part 4. Belief is the reported rank in part 5. Investment is the amount invested in the part 6 investment task in ECU. +p<0.10, *p<0.05, **p<0.01, ***p<0.001; significance levels are from t-tests with robust standard errors for the gender difference.

We now move on to examining the effect of advice on tournament entry. Figure 1 plots the distribution of the advice effect separately by treatment and gender. For women, advice increases willingness to compete in all three treatments (p=0.024, p=0.004 and p=0.007 in treatments Competitiveness, Risk and Confidence respectively, t-tests). In fact, all women who responded to the advice did so in the intended direction, by shifting from piece rate to tournament. For men, there is essentially no effect in any of the three treatments (p=0.322, p=0.322 and p=0.321 in treatments Competitiveness, Risk and Confidence respectively, t-tests). Indeed, across the three treatments advice increases female willingness to compete by 10.6 percentage points (p<0.001, t-test), whereas it has no effect on men's average willingness to compete (p=0.565, t-test).

Table 2 presents the results of the pre-registered test of the treatment effect, which controls for gender differences in performance.[6] Column 1 shows that, when we control for gender differences in performance, advice significantly (p<0.001) reduced the gender gap in willingness to compete by 9.8 percentage points over all treatments combined. Hence, simple advice successfully reduced the gender gap in willingness to compete. Relative to a baseline gap of 25.9 percentage points (while controlling for performance), simple advice eliminated 38 percent of the gender gap in willingness to compete. Columns 2-4 present similar analysis done separately for each treatment. The point estimate for simple advice is largest in the Risk treatment, followed by the Confidence and Competitiveness treatments; in the Competitiveness treatment the effect of advice is no longer significant. In other words, simple advice significantly reduced the gender gap in willingness to compete when the advice was for women to be more risk taking and more confident, but not when it advised them to be more competitive. However, the advice effect does not differ significantly across treatments (see Online Appendix E) and we hence conclude that small variations in the wording of the advice do not seem to have a major effect on its effectiveness.

[6] Because there was a marginally significant performance difference between men and women in part 2, we investigate treatment effects using regressions that control for performance in part 2, as specified in the pre-analysis plan. However, all of the results in this section are robust to not controlling for performance and using t-tests instead, see Online Appendix E.

Figure 1: Effect of Advice in the Laboratory Study, by Gender and Treatment



Notes: The figure displays the advice effect on tournament entry by treatment and gender. The error bars represent 95% confidence intervals.

Table 2: Effect of Advice on Tournament Entry in the Laboratory Study

	(1) Advice Effect (All)	(2) Advice Effect ("Competitiveness")	(3) Advice Effect ("Risk")	(4) Advice Effect ("Confidence")
Female	0.098***	0.045	0.145**	0.110*
	(0.024)	(0.035)	(0.046)	(0.047)
Constant	0.007	0.024	-0.016	0.012
	(0.010)	(0.019)	(0.017)	(0.016)
Score Controls	yes	yes	yes	yes
Observations	374	127	123	124
R-squared	0.046	0.026	0.076	0.052

Notes: Robust standard errors are in parentheses. The dependent variable ("Advice Effect") measures the response to receiving advice. It is equal to 1 for participants who switched from piece rate to tournament after hearing the advice, -1 for those who switched in the opposite direction, and 0 for those whose decisions were not affected by advice. All regressions control for score in part 2 (demeaned). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Overall, we see evidence that simple advice makes women more likely to compete, in particular when this advice encourages women to be less risk averse or more self-confident. By contrast, men appear to be unaffected by advice. Taken together, this also implies that simple advice reduces the gender gap in willingness to compete.

Given that women, on average, competed too little prior to receiving advice, it seems plausible that advice would also increase the fraction of women who choose the payoff-maximizing (i.e., efficient) payment scheme. Table 3 displays the effect of advice on this measure of efficiency for the pooled sample and split by gender and by those who should compete (high ability) and those who should not (low ability). We compare the fraction of participants who chose the payment that maximized their earnings before and after advice. Across all treatments, simple advice made 12 participants shift to their payoff-maximizing options, whereas 11 participants shifted in the opposite direction. As a result, the number of participants choosing the payoff maximizing option did not improve because of advice (p=0.835, column 1). Columns 7 and 8 show that this is because advice induces both high-performing and low-performing women to compete more, leading to a net zero effect on efficiency overall. In Appendix E, we show that similar results obtain when we look at an alternative measure of earnings (expected foregone earnings) instead.[7]

One alternative definition of efficiency is the share of high performers that decide to compete. This definition may be of particular interest to firms, since it will be easier to fill high-level positions with good candidates if more high-performing candidates actually choose to apply for a job. As reported in column 6 of Table 3, there is a significant increase in the willingness of high-performing women to compete (10 percentage points). In the fourth column we can see that there is no effect on willingness to compete among high-performing men. Taken together, this implies that advice makes high performers 5.4 percentage points more likely to compete (p=0.001). This is an increase of 10.7 percent from a baseline of 50.2 percentage points.

[7] Our results differ from Brandts et al. (2015), who find that highly personalized advice does reduce foregone earnings due to wrong entry decisions. This is likely due to the fact that advice in their study was performance-based, such that high and low performers received personalized advice that encouraged them to select the expected earnings maximizing payment scheme.

Overall, our information intervention made both high and low performing women more likely to compete. This implies that efficiency, when defined as the fraction of participants who choose the payoff maximizing option, did not increase, though it is worth noting that it did not decrease either. On the other hand, our intervention did unambiguously increase the probability that high performers compete, and lower the gender gap in willingness to compete. Hence, our information intervention seems promising if the policy goal is to reduce the gender gap and increase tournament entry among high performers without lowering expected earnings overall.

Table 3: The Effect of Advice on Efficiency in the Laboratory Study

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max
Advice	0.003	0.005	0.000	0.010	0.000	0.100**	-0.112**
	(0.013)	(0.009)	(0.024)	(0.010)	(0.018)	(0.030)	(0.034)
Constant	0.610***	0.665***	0.556***	0.676***	0.650***	0.320***	0.820***
	(0.025)	(0.035)	(0.036)	(0.046)	(0.054)	(0.047)	(0.041)
Gender	All	Men	Women	Men	Men	Women	Women
Performance	All	All	All	High	Low	High	Low
Observations	374	185	189	105	80	100	89

The dependent variable is whether a participant chose the expected payoff-maximizing payment scheme given their part 2 performance and the part 2 performance of all other participants in this study. Robust standard errors in parenthesis. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

3. Field Study

We also conducted a field study with high school students in Sweden, allowing us to test whether the results from the laboratory generalize to a different setting. Specifically, the field study was conducted between 2011 and 2014 as part of a social science class at Nacka High School in a suburb of Stockholm, Sweden. There were three separate lecture groups in each cohort. Each of these twelve lecture groups took part in our experiment, for a total of 268 participants (151 women and 117 men). The average age was 17, and the number of students attending class on the day of the experiment varied between 18 and 29. Classrooms were set up in an exam setting to prevent the participants from looking at each other's answers. The experiment was conducted using pen and paper, instructions were read publicly, and correct answers were given on the white board at the end of the experiment. The teachers were present during the entire experiment. In total each session lasted around 30 minutes.

3.1 Design

The main part of the experiment consisted of three parts of five minutes each where the participants solved the same types of math problems as in the laboratory study. As in the laboratory, the first part consisted of a piece-rate compensation scheme that awarded participants 5 SEK per correct answer.[8] In the second part, participant were paid according to a tournament payment scheme; the main difference with our laboratory study is that we used a four-person tournament, where participants were paid 20 SEK per correct answer if they performed better than three peers from a comparison sample.[9] Ahead of the third part, participants could then choose between these two compensation schemes, as in the laboratory study.

In contrast to our laboratory study, the field study used a between-subject design. Specifically, we randomly assigned each of the twelve lecture groups into either receiving or not receiving the following message just before making their third part decision (translated from Swedish):

[8] SEK/USD ≈ 0.15 at the time when the experiment where conducted.

[9] We told participants that these peers were from the same population (high school students in the social science track) but did not give them any additional information. To generate a suitable peer group, we conducted a pilot session with a different set of students at the same school prior to the main experiment and used the scores from this session to construct the peer groups. The performance distribution in the pilot session mirrors the performance distribution among the participants in the experiment.

"Research has found that men are more prone to compete than women. Further, men, on average, tend to compete too much given that they want to maximize their expected pay-off while women, on average, do not compete as much as they should."

The message corresponds closely to the Competitiveness treatment in the laboratory study. The main difference is that participants who received this message saw it *prior* to making their initial tournament entry decision. Hence, the main comparison in this study is a between-subject comparison between the tournament entry decisions made by treated and untreated individuals.

Similar to the laboratory experiment, participants then went through two additional parts where confidence and risk preferences where elicited. In part 4, the participants were asked to guess their performance quartile in the second part relative to the comparison sample. A correct guess was rewarded with 100 SEK. In part 5 the participants were asked if they preferred a guaranteed amount of 10/30/50/70/90 SEK or a 50% chance of winning 100 SEK. Participants made a choice for each of the five guaranteed amounts; one of these choices was then randomly implemented and paid out. After the end of the experiment, we asked participants for their legal gender and two participants in each group were randomly selected to receive compensation.

3.2 Results

Descriptive statistics and the willingness to compete for the control group are reported in Table 4. The first three rows show that women, in all three parts, outperform men. The difference is however only statistically significant in the first part. This pattern departs somewhat from the laboratory experiment where, if anything, men performed slightly better. It is however comparable to Tungodden (2019)'s field experiment with Norwegian high school children (see also Dreber et al., 2014). The fourth row reports the fraction of participants that should compete based on their performance in round 2. More women than men should compete, but the difference is not statistically significant (the reason that the fraction that should compete is lower compared to our laboratory study is due to the fact that in the laboratory the participants compete against one other person while in the field study, they compete against three).

Table 4: Pre-Advice Results for the Field Study

	Men	Women	Difference
(1) Score, part 1	5.239	6.258	-1.019**
	(3.218)	(2.803)	(0.368)
(2) Score, part 2	6.205	6.344	0.139
	(3.223)	(2.985)	(0.381)
(3) Score, part 3	7.274	7.589	0.316
	(4.070)	(3.305)	(0.451)
(4) Should compete	0.308	0.391	0.083
	(0.464)	(0.490)	(0.059)
(5) Compete, control group	0.559	0.244	0.316***
	(0.501)	(0.432)	(0.080)
(6) Should compete but does not, control group	0.051	0.218	-0.167**
	(0.222)	(0.416)	(0.060)
(7) Should not compete but does, control group	0.271	0.051	0.220***
	(0.448)	(0.222)	(0.058)
(8) Beliefs about rank	2.525	2.244	0.282+
	(0.935)	(0.983)	(0.166)
(9) Number of risky choices	2.390	2.013	0.377*
	(0.929)	(0.781)	(0.146)
Observations	117/59	151/78	268/137

Notes: Columns "Men" and "Women" show averages for the respective genders (standard deviations in parentheses). Column "Difference" shows the gender difference (standard errors in parentheses). "Should compete" is the fraction of participants who would have maximized their payment by competing based on their part 2 score. The next two rows report the fraction of participants in the control group who could have maximized their payment by choosing tournament (piece rate) but chose the piece rate (tournament) instead. Belief is the reported rank (with 1 being the lowest and 4 being the highest). Number of Risky Choices is the number of times the risky option was chosen in the risk aversion elicitation. +p<0.10, *p<0.05, **p<0.01, ***p<0.001; significance levels are from t-tests with robust standard errors for the gender difference.

Rows 5-7 reports the comparative behavior and outcomes in the control group as a baseline for the treatment effects. In row five we can see that about 56% of men compete while only 24% of women do. The difference is highly significant and overall the results are very similar to those in the laboratory study. Just as in the laboratory study, we can calculate the share of participants choosing the expected payoff maximizing option. This is reported in rows 6 and 7. Among men, 85% of participants who fail to choose the expected payoff maximizing option are doing so by competing when they should not, while for women the corresponding number is only 19%. These differences are highly statistically significant.

Finally, in rows 8 and 9 we report elicited confidence and risk preferences for the control group. We can see that men rank their performance higher than women do, even though they, if anything, perform worse. We can also see that men are more risk-taking than women. Again, this is in line with findings in the previous literature and the findings in out laboratory study.

Table 5 report the results of the effect of advice on the participants' choice to compete. Even though we do not see any significant differences between men and women in performance in part 2 we include the demeaned score in part 2 as a control to be consistent with how we analyze the data from the laboratory study.[10] As a first step, column 1 shows that our treatment has no overall effect on willingness to compete. However, columns 2 and 3 demonstrate that the aggregate results obscure that the treatment has a large effect when looking at men and women separately. Men become 30 percentage points *less* likely to compete while women become 29 percentage points *more* likely to compete. The initial gender difference of 32 percentage points is hence almost entirely reversed. Thus, though the effects go in the same

[10] The results are robust to excluding this control. They are also robust to using randomization inference on the classroom (treatment) level. We refer to Online Appendix E for these robustness checks.

direction, participants appear to react more strongly to advice than in the laboratory (in particular women).[11]

Table 5: Effect of Advice on Tournament Entry in the Field Study

	(1)	(2)	(3)
	Compete	Compete	Compete
Advice	0.033	-0.298***	0.289***
	(0.051)	(0.072)	(0.064)
Constant	0.379***	0.564***	0.239***
	(0.038)	(0.058)	(0.044)
Score Controls	Yes	Yes	Yes
Subgroup	All	Men	Women
Observations	268	117	151
R-squared	0.277	0.374	0.360

Notes: Robust standard errors are in parentheses. The dependent variable is the tournament entry decision in part 3 (1-tournament, 0-piece rate). "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). All specifications control for score in part 2 (demeaned). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table 6 reports the treatment effects on efficiency as defined by the share of participants making payoff-maximizing decisions. The first column shows that in total the share of pay-off maximizing choices increases by 11.6 percentage points. This corresponds to a reduction of pay-off minimizing choices of around 40%. This result differs from the effects documented in the laboratory study where advice did not have any effect on this measure of efficiency. Column 2 and 3 report the effect on men and women separately. For men, the probability of choosing the expected payoff maximizing payment scheme significantly increases by 20 percentage points, while for women the effect of advice is not significant.

[11] Performance is not affected by the advice; average performance is nearly identical in the treatment and control group (difference of 0.062 points; p=0.89). In Online Appendix E we also show that the results from Table 5 are robust to using permutation tests instead.

When dividing the sample further based on whether a participant would maximizes their payment with tournament or piece rate pay, we can see that there are both winners and losers from the treatment. The winners are high-performing women and low-performing men. These groups increase their likelihood of making the pay-off maximizing choice by 45.7 and 33.4 percentage points, respectively. The losers are low-performing women, whose likelihood of choosing the payoff-maximizing payment scheme decreases by 20.8 percentage points. There are no statistically significant effects on high-performing men.[12]

Table 6: The Effects of Advice on Efficiency in the Field Study

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max
Treated	0.116*	0.201**	0.050	-0.100	0.339***	0.457***	-0.217**
	(0.051)	(0.075)	(0.070)	(0.138)	(0.089)	(0.103)	(0.080)
Constant	0.708***	0.678***	0.731***	0.850***	0.590***	0.469***	0.913***
	(0.039)	(0.061)	(0.051)	(0.082)	(0.080)	(0.090)	(0.042)
Gender	All	Men	Women	Men	Men	Women	Women
Performance	All	All	All	High	Low	High	Low
Observations	268	117	151	36	81	59	92
R-squared	0.019	0.059	0.003	0.016	0.160	0.238	0.075

Notes: Robust standard errors in parentheses. The dependent variable is whether participants chose the payoff-maximizing payment scheme given their part 2 performance. "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Similar to Section 2.2, we can also consider efficiency from the employer's perspective by examining the frequency of high performers that choose to compete. Table 6 shows that advice makes high-performing significantly more likely to compete (46 percentage points), and makes high-performing men less likely to compete (10 percentage points), though the latter effect is not significant. Taking into account the larger number of women in our sample, these results together imply that high performers in our sample were, on average, 25 percentage

[12] These results are qualitatively similar if we use the expected foregone earnings as our dependent variable instead, see Online Appendix E. The results also hold when using permutation tests instead as shown in Online Appendix E.

points more likely to compete after receiving advice (p=0.006, t-test). This is an increase of 40 percent from a baseline of 62 percentage points.

Overall, the results of the field study are similar to the lab study in that simple advice reduces the gender gap in willingness to compete and makes high performers more likely to choose tournament pay. However, the size of both effects is larger in the field study than in the lab. In addition, our information intervention makes participants more likely to choose the payoff maximizing payment scheme in the field, whereas it did not have an effect in the lab.

4. Discussion and Conclusions

We investigate the extent to which an information intervention that brings awareness to the existence of a gender gap in willingness to compete can reduce this same gender gap. The results from our laboratory study in Germany and our field study in Sweden both confirm that this is indeed the case. Further, comparing the effect size of our intervention shows that it does as good, or better, a job of reducing (and in the field even reversing) the gender gap as other, more complex, costly and elaborate interventions studied in previous work. For example, only the gender quota in Niederle, Segal and Vesterlund (2013) have a larger effect than the advice effect in the field study, and even our laboratory results are larger than the effect of repeating competitions with a non-female winner (Balafoutas & Sutter, 2012) or receiving highly personalized advice (Brandts, Groener and Rott, 2015), see Table A2 for more details. In addition, our intervention positively affects the ex-post efficiency of outcomes, in the sense that it makes high performers more likely to enter (and win) tournaments, while not reducing (in the lab) or even increasing (in the field) the share of participants who choose the payoff-maximizing payment scheme.

Even though the advice led to a reduction in the traditional gender gap in willingness to compete in both our laboratory and our field study, there were also some differences worth discussing. Most notably, the size of the treatment effect was larger in the field study, and whereas advice only affected women in the laboratory study, it affected both genders in the field study. One potential explanation for the latter result is that (in contrast to Niederle and Vesterlund, 2007 and the field study) men did not compete too much in the laboratory study. This, in turn, could imply that there was less room for the advice to have an effect on men in the laboratory. Other reasons why the advice effect was larger in the field include sample differences (age and nationality in particular) and differences in the perceived authority of the advice giver (it may be that an experimenter is perceived as more of an authority by high school students than by university students who are used to doing laboratory experiments).

However, differences in design between the two studies may also have had an impact on the effect size, and on who was affected by the advice. In the laboratory study, the advice was given after participants had already made a preliminary decision, whereas in the field study the advice was provided before any decision about payment scheme was made. It is possible that the exact timing of advice matters, and we regard this as an intriguing topic for future research. We also hope that future work will investigate other aspects of the effects of advice – it would for example be interesting to investigate how long-lasting the effects of advice are, and if it makes a difference if the advice is given repeatedly rather than only once.

Our results are promising from the perspective of a policy maker. In many countries, increasing the representation of women in high-level, prestigious and traditionally male professions remains an important policy goal. Willingness to compete has been linked to the underrepresentation of women in such positions, and our results therefore suggest that simple advice may be a cost-effective way to decrease female underrepresentation.

Our results may also have implications beyond competitiveness. Though part of the reason why advice was effective in our study may have been that it was quite specific and directly relevant to the choice participants were making, it may be possible to give similarly specific advice and directly related advice in related contexts as well. For example, similar to being underrepresented in competitions, women are also underrepresented in STEM-fields. We can therefore imagine an information intervention similar to the one in our study where students are reminded that it is possible for women to increase their expected earnings by sorting into STEM majors in college; testing this intervention would be another intriguing topic for future research.

References

Apicella, C L, E E Demiral, and J Mollerstrom. 2017. "No Gender Difference in Willingness to Compete When Competing against Self." American Economic Review (Papers and Proceedings) 107, no. 5 (2017): 136-40.

Akerlof, G. A., & Kranton, R.E. 2000. "Economics and Identity." *Quarterly Journal of Economics*, 115 (3): 715–53.

Bönte, W, V Procher, and D Urbig. 2017. "Gender Differences in Selection into Self-competition." *Applied Economics Letters* 25: 1-5.

Balafoutas, L and M Sutter. 2012." Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory." Science, 335(6068): 579-582.

Blau, F D and L M Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, 55(3): 789-865.

Brandts, J., Groenert, V., & Rott, C. 2015. The Impact of Advice on Women's and Men's Selection into Competition. *Management Science*, 61(5), 1018–1035.

Braun, S., Dwenger, N., Kübler, D., & Westkamp, A. 2014. Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior*, 85, 232–251.

Buser, T, M Niederle, and H Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics,* 129: 1409-1447.

Buser, T, N Peter, and S C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review (Papers and Proceedings),* 107(5): 125-130.

Buser, T., Ranehill, E and R. van Veldhuizen. 2018. "Gender Differences in Willingness to Compete: The Role of Public Observability" Working paper.

Carpenter, J P., R S. Frank, and E Huet-Vaughn. 2018. "Gender Differences in Interpersonal and Intrapersonal Competitive Behavior." *Journal of Behavioral and Experimental Economics*, 77: 170-176.

Castleman, B. L., & Page, L. C. 2015. Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115, 144–160.

Çelen, B., Kariv, S., & Schotter, A. 2010. An Experimental Test of Advice and Social Learning. *Management Science*, 56(10), 1687–1701.

Charness, G., and U. Gneezy. 2012. "Strong Evidence for Gender Differences in Risk Taking." *Journal of Economic Behavior &* Organization, 83(1): 50-58.

Chaudhuri, A., Schotter, A., & Sopher, B. 2009. Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Effort Games with Private, Almost Common and Common Knowledge of Advice. *The Economic Journal*, 119(534), 91–122.

Croson, R, and U. Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448–74.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. 2013. Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry*, 51(1), 816–835.

Dargnies, M.-P. 2012. Men Too Sometimes Shy Away from Competition: The Case of Team Competition. *Management Science*, 58(11), 1982–2000.

Del Carpio, L. and Guadalupe, M. 2018. More Women in Tech? Evidence from a Field Experiment Addressing Social Identity. *CEPR Discussion Paper No*. DP13234.

De Quidt, J., Haushofer, J. and Roth, C. 2018. Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11): 3266-3302.

Dreber, A., von Essen, E., & Ranehill, E. 2014. Gender and Competition in Adolescence: Task Matters. *Experimental Economics*, 17(1): 154-172.

Eagly, A. H. 1987. *Sex Differences in Social Behavior: A Social-Role Interpretation*. Hillsdale, NJ: Erlbaum.

Ertac, S., and B. Szentes. 2011. The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence. Working Paper.

Fischbacher, U. 2007. “z-Tree: Zurich Toolbox for Ready-Made economic Experiments.” *Experimental Economics*, 10, 171–178.

Gillen, B., Snowberg, E., and L. Yariv. 2019. “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study.” *Journal of Political Economy*, 127(4), 1826–1863.

Gneezy, U., and J. Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics*, 112(2), 631-645.

Greiner, B. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1(1), 114-125.

Guillen, P., and R. Hakimov. 2018. The effectiveness of top-down advice in strategy-proof mechanisms: A field experiment. *European Economic Review*, 101, 505–511.

Heβ, S. 2017. “Randomization Inference with Stata: A Guide and Software.” *The Stata Journal*, 17(3): 630–651.

Hoxby, C. and S. Turner. 2013. “Expanding College Opportunities for. High-Achieving, Low Income Students.” SIEPR Discussion Paper No. 12- 014.

Karni, E. 2009. "A Mechanism for Eliciting Probabilities." *Econometrica*, *77*(2), 603–606.

Klinowski, D. 2017 “Gender and the Willingness to Compete Against Own Past Performance.” Working Paper.

Mobius, M., Niederle, M., Niehaus, P., & T. Rosenblat. 2014. “Managing Self-Confidence: Theory and Experimental Evidence.” NBER Working Paper No. 17014

Niederle, M., and L. Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122 (3): 1067–1101.

Niederle, M, and L Vesterlund. 2011. “Gender and Competition.” *Annual Review Economics.* 3, no. 1: 601-630.

Niederle, M, C Segal, and L Vesterlund. 2013. “How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness” *Management Science,* 1: 1-16.

Niederle, M. 2015. “Gender”, In *The Handbook of Experimental Economics: Volume 2.* Edited by John H. Kagel, Alvin E. Roth, Princeton: Princeton University Press, 481-546, 2015.

Rudman, L. A., & Glick, P. 2001. “Prescriptive Gender Stereotypes and Backlash Toward Agentic Women.” *Journal of Social Issues*, 57 (4): 743–62.

Schotter, A. and Sopher, B. 2003. Social learning and coordination conventions in intergenerational games: an experimental study, *Journal of Political Economy*, 111(3) (June), pp. 498–529.

Schotter, A. and Sopher, B. 2006. Advice, trust and trustworthiness in an experimental intergenerational game, *Experimental Economics*, 9(2), pp. 123–45.

Schotter, A. and Sopher, B. 2007. Advice and behavior in intergenerational ultimatum games: an experimental approach, *Games and Economic Behavior*, 58(2), pp. 365–93.

Tungodden, J. 2019. When Parents Decide: Gender Differences in Competitiveness. Working Paper

Van Veldhuizen, R. 2018. "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness?" Working paper.

Wozniak, D, W T. Harbaugh, and U Mayr. 2014. “The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices.” *Journal of Labor Economics* 32 (1): 161-198.

Young, A. 2019. “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results.” *The Quarterly Journal of Economics*, 132(2): 557–598.

Online Appendix A: Pre-Analysis Plan and Power Calculations

Pre-Analysis Plan

For "Can simple advice eliminate the gender gap in willingness to compete?" by X and Y

Preliminaries

1. We construct our main dependent variable (DV) by subtracting the choice made prior to receiving advice (1-tournament, 0-piece rate) from the choice made after receiving advice.
2. We run a two-sided t-test investigating if performance in round 2 differs significantly by gender across all three treatments.

Main Analysis

3. Two-sided t-test investigating if the DV differs significantly by gender across all three treatments.
    a. Replaced by a regression of the DV on gender and performance in round 2 if performance differs by gender in step 2 ($p<.10$).
4. Two-sided t-test investigating if the DV differs significantly by gender within each treatment.
    a. Replaced by a regression of the DV on gender and performance in round 2 if performance differs by gender in step 2 ($p<.10$).
5. Difference-in-difference tests to compare whether the gender gap in the DV differs across the three treatments. We will run three pairwise comparisons.
    a. We also control for performance if performance differs by gender in step 2 ($p<.10$).

Robustness Analysis

6. Robustness check to (not) controlling for performance
    a. Replace t-test by regression controlling for performance or vice versa, depending on which test was used for the main analysis.

Additional Analysis

7. An analysis of the mechanisms driving the treatment effects:
    a. We use a difference-in-difference test to check whether any gender difference in the (a) average willingness to take risks and (b) average belief differs significantly by condition. This tells us whether different forms of advice have different impacts on two mechanisms that are thought to explain gender differences in willingness to compete (risk preferences and beliefs).
8. A second analysis of the mechanisms driving the treatment effects based on Van Veldhuizen (2018):
    a. We use the following outcomes:
        i. Tournament entry post advice.
        ii. Lottery entry in a control decision that eliminates the role of competitiveness by replacing the tournament with a lottery.

    iii. Lottery entry in a control decision that eliminates competitiveness and also eliminates the role of gender differences in overconfidence.

b. For more details on how exactly these choices are constructed, and how they control for possible differences in performance, we refer the reader to Van Veldhuizen (2018), where (ii) and (iii) are referred to as treatments 'NoComp' and 'IntEffect' respectively.

c. We can then identify the role of each mechanism as follows:
    a. Any gender difference in (ii-i) is evidence that the gender gap in tournament entry is (partially) driven by competitiveness.
    b. Any gender difference in (ii-iii) is evidence that the gender gap in tournament entry is (partially) driven by overconfidence.
    c. Any residual gender gap in (iii) is evidence that the gender gap is driven by risk preferences.

d. To identify the importance of each mechanism, we use a two-sided t-test for all three mechanisms both for the pooled data and for each treatment individually.

e. We will also compare the importance of each mechanism across treatments using difference-in-difference tests.

9. An analysis investigating whether simple advice improves efficiency:
    a. We consider two related measures of efficiency:
        i. The fraction of participants who chose the option (tournament or piece rate) that maximized their expected payment.
        ii. The expected monetary loss suffered by participants who did not choose the payoff-maximizing payment.
    b. We compute these measures in the following way:
        i. We compute the empirical probability of winning given forced tournament (task 2) performance (p_{win}). This is equal to the empirical fraction of participants with lower performance, plus 0.5 times the fraction of participants with identical performance (given the tiebreaker rule). Participants with $p_{win}>0.5$ ($p_{win}<0.5$) maximize their earnings by choosing the tournament (piece rate).
        ii. The expected monetary loss is then equal to $(T*p_{win}-PR)*x$ for $p_{win}>0.5$ and $(PR-T*p_{win})*x$ for $p_{win}<0.5$, where T and PR are the payment per exercise in the tournament and piece rate respectively, and 'x' equals performance on the task (in the task 2 forced tournament).
    c. We compare efficiency post-advice to efficiency pre-advice, both pooled across treatments and genders and separately for each treatment and gender.
    d. We also compare post-advice efficiency across treatments both pooled across genders and for each gender separately.

Reference

Van Veldhuizen, R., 2018. "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness?" Working paper.

Power Calculations

Our key hypothesis is that providing advice decreases the gender difference in willingness to compete. Since we use a within-subject design, we compute power $P(p < 0.05|H_1)$ using simulations. For this purpose, we start by taking the fraction of men and women that choose the tournament in an earlier study run at the same laboratory (Van Veldhuizen, 2018) as the benchmark for willingness to compete (men=58.6%, women=27.1%). We then specify several different alternative hypotheses H_1 that decrease male and increase female willingness to compete by a similar amount (e.g., men=48.6% and women=37.1% for a 20pp decrease in the gender gap). For each H1, we then simulate 1000 samples of 180 men and 180 women using a multinomial distribution. For this purpose, we assume that some women (e.g, 27.1% for the example of a 20pp decrease in the gender gap) compete in both cases, some women (10% in the example) compete only after receiving advice and the remainder (62.9% in the example) choose piece rate in both cases. The distribution for men is analogous. We also allow for some noise by assuming that some participants may change their decision after receiving advice based on reasons that are orthogonal to the treatment (e.g., by mistake). In Van Veldhuizen (2018), 21% of participants switched their decision across two similar tournament entry decisions in a way that left the gender gap unaffected. Since these two decisions were similar but not identical, we treat 21% as an upper bound for the expected noise in the sample, and compute power for a sample with 11%, 16% or 21% of switchers in each condition.

Table A1: Power Calculations

	Switchers: 21%		Switchers: 16%		Switchers: 11%	
	Power		**Power**		**Power**	
Effect Size (H_1)	Joint	Individual	Joint	Individual	Joint	Individual
10pp	0.496	0.201	0.624	0.233	0.723	0.304
15pp	0.838	0.414	0.886	0.472	0.956	0.575
20pp	0.974	0.597	0.989	0.675	0.998	0.755
25pp	0.998	0.781	0.999	0.825	1	0.892

As expected, power depends on the assumed effect size. To identify an appropriate effect size, Table A2 presents the effect sizes as well as sample sizes in a number of previous studies that also investigate whether the gender gap in willingness to compete decreases across treatments. With our target sample size, our joint test has a power of 0.838 to detect an effect that is 54% as large as the average effect in Table A2 if we conservatively assume that 21% of individuals will switch for reasons unrelated to advice. For the individual comparisons, our power is lower but still around 0.80 for proposed effect sizes that are close to the ones reported in the literature. Also as expected, power increases when we assume that fewer than 21% of individuals switch for reasons unrelated to advice. Finally, we also wish to note that our choice of a within-subject design was partially driven by power. For example, we calculated that we would need a total sample of 500 people to get a power of 0.844 for an effect size of 25pp. With a within-subject design, we are able to achieve similar power for a much smaller effect size (15pp) with a smaller sample (360 people).

Table A2: Sample Size and Effect Size in Previous Work (Based on Table A2 in Buser et al., 2018)

Study	Sample Size		Fraction Competes		Test Used		Effect Size		
	Control	Treatment	Control	Treatment	Type	B/W	DiD	Men	Women
ADM2017Lab	50M,50W	52M,52W	0.58M,0.38W	0.55M,0.42W	DiD	Between	0.07	-0.03	0.04
ADM2017Online	129M,129W	112M,112W	0.40M,0.28W	0.31M,0.36W	DiD	Between	0.17	-0.09	0.08
BGR2015	56M,56W	56M,56W	0.59M,0.30W	0.59M,0.38W	DiD	Between	0.08	0.00	0.08
BS2012Quota	36M,36W	36M,36W	0.64M,0.31W	0.62M,0.53W	Women	Between	0.24	-0.02	0.22
BS2012REP	36M,36W	36M,36W	0.64M,0.31W	0.67M,0.39W	Women	Between	0.05	0.03	0.08
BS2012PT1	36M,36W	36M,36W	0.64M,0.31W	0.55M,0.58W	Women	Between	0.36	-0.09	0.27
BS2012PT2	36M,36W	36M,36W	0.64M,0.31W	0.50M,0.69W	Women	Between	0.53	-0.14	0.39
D2012	78M,74W	78M,74W	0.85M,0.51W	0.59M,0.62W	All	Within	0.37	-0.26	0.11
DER2014	108M,109W	108M,109W	0.36M,0.17W	0.33M,0.28W	M+W	Within	0.14	-0.03	0.11
GLL2009	34M,40W	52M,28W	0.50M,0.26W	0.39M,0.54W	DiD	Between	0.39	-0.11	0.28
HP2011	32M,32W	64M,62W	0.81M,0.28W	0.67M,0.45W	DiD	Between	0.31	-0.14	0.17
NSV2013	84M,84W	84M,84W	0.74M,0.31W	0.45M,0.83W	All	Within	0.81	-0.29	0.52
S2012Math	36M,36W	36M,36W	0.44M,0.19W	0.44M,0.36W	Women	Within	0.17	0.00	0.17
S2012Verbal	45M,45W	45M,45W	0.39M,0.30W	0.43M,0.57W	Women	Within	0.21	0.06	0.27
Average:	57M,57W	59M,57W	0.59M,0.30W	0.51M,0.50W			0.28	-0.08	0.20

Note: The table reports the results of fourteen separate experiments from nine published papers. For each experiment, the first two columns (‘Sample Size’) display the number of men (M) and women (W) in the control and treatment condition respectively. The next two columns (‘Fraction Competes’) display the fraction of men and women in the respective conditions who choose to compete. ‘Test Used’ displays both the ‘Type’ of test reported in the paper, either for a change in the gender gap (difference-in-difference, DID) or for a change in the fraction of men or women who compete, and whether the study used a between-subject or within-subject design (‘B/W’). The three ‘Effect Size’ columns report the gender gap, the fraction of men and the fraction of women competing in the treatment condition minus the corresponding fraction in the control condition. ‘ADM2017’ is Acipella, Dreber and Mollerstrom (2017), ‘BS2012’ is Balafoutas and Sutter (2012), ‘BGR2015’ is Brandts, Groener and Rott (2015), ‘D2012’ is Dargnies (2012), ‘DER2014’ is Dreber, von Essen and Ranehill (2014), ‘GLL2009’ is Gneezy, Leonard and List (2009), ‘HP2011’ is Healy and Pate (2011), ‘NSV2013’ is Niederle, Segal and Vesterlund (2013), and ‘S2012’ is Shurchkov (2012). The control condition in all experiments is a variation of the Niederle and Vesterlund (2007) design. In the treatment conditions, participants compete against their own past performance (ADM2017), participants are advised whether or not to compete (BGR2015), there is a minimum number of female winners (BS2012, Quota; NSV2013), tournaments are repeated if there are too few female winners (BS2012, REP), or female performance is artificially boosted by one or two units (BS2012, PT1 and PT2 respectively). In D2012, treated participants choose whether to enter as a team, in DER2014 treated participants face a verbal task, in GLL2009 the treated group consists of a different population (Khasi) than the control (Maasai), in HP2011 treated participants choose whether to compete based on the combined performance of themselves and a team mate and in S2012 treated participants are under time pressure. ADM2017 includes both a laboratory and online (MTurk) experiment. BS2012 includes four different affirmative action conditions that are all compared to the same control condition. S2012 includes separate sessions with verbal and mathematical tasks. ‘Average’ is the unweighted average of the respective column across all reported experiments.

Online Appendix B: Instructions

B.1 Instructions for the Laboratory Experiment[13]

Thank you for participating in today's experiment. During the experiment it is not allowed to use electronic devices or to communicate with other participants. Please do not try to exit the experimental program, and do not talk to the other participants. If you have a question, please raise your hand. We will then come to you and answer your question in silence. Please do not ask your questions out loud. If the question is relevant for all participants, we will repeat it loudly and answer it. If you violate these rules, we will disqualify you from further participation in the experiment and payment.

Today's experiment has six different parts. In total, we expect the experiment to take less than an hour. In addition to the 7 Euro payment that you receive for your participation, you will be paid an additional amount of money that you accumulate from decision tasks in the experiment. The exact amount you receive will be determined during the experiment, and will depend on your decisions, and the decisions of others.

From parts 1, 2, 3 and 4 of the experiment, one will be randomly chosen to matter for your payment. That means that your earnings from that randomly chosen part will be paid out to you. What you earn in part 5 and in part 6 of the experiment will always be paid out to you.

We will give you further information about each part of the experiment once we get there.

All monetary amounts you will see in this experiment will be expressed in experimental currency units (**ECU**). At the end of the experiment, your earnings in ECU will be exchanged into Euro at a rate of:

10 ECU = 1 Euro

If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you.

This is the start of the first part of the experiment.

In this part, you will earn money for you performance in a task. You will do the same task in part 2 and part 3 of the experiment. However, the way you are paid for your performance will vary between these three parts. We will explain how you will be paid at the beginning of each part.

The task consists of calculating the sum of five randomly chosen two-digit numbers.

Example: 24+56+97+71+45=?

[13] These are our original English instructions; the German translations that were used in running the experiment are available upon request.

In each of the first three parts of the experiment you have 5 minutes to complete as many sums as you can. You cannot use a calculator or phone to determine the sums, however, you are welcome to write the numbers down and make use of the provided scratch paper.

If the current part, Part 1, is selected for payment, you will be paid as follows: you will receive 5 ECU for each correctly solved sum.

Please raise your hand if you have any questions.

This is the start of Part 2. The task is exactly the same as before but the way you are paid for your performance is different.

In this part you will compete against another participant. This participant is chosen randomly among all other participants who are in the lab with you today.

If this part is selected for payment, you will be paid as follows: You receive 10 ECU for each correctly solved sum <u>if you solve more sums than your opponent</u>. You receive no payment if you perform worse than your opponent. In case of a tie, the computer will randomly determine whether you win or lose.

Please raise your hand if you have any questions.

This is the start of Part 3.

The task is the same as in the previous parts. However you will now get to choose which of the two previous payment schemes you prefer to apply to your performance in the third part. You can choose between *Piece Rate* as in Part 1 and *Competition* as in Part 2.

If Part 3 is the one randomly selected for your payment, then your payment is determined as follows:

- If you choose the Piece Rate, you receive 5 ECU per sum that you solve correctly (as in Part 1).

- If you choose Competition, your performance will be compared to a randomly selected opponent (as in Part 2). This opponent will be randomly selected among all other participants present here today. If you solve more sums than your opponent solved in Round 2, you receive 10 ECU per correct answer. If you perform worse than your opponent, you receive nothing. In case of a tie, the computer will randomly determine whether you win or lose.

Please raise your hand if you have any questions.

Which compensation scheme do you choose for Part 3?

Before you do the task in Part 3, we would like to share the following information with you.

Competitiveness Treatment:
In a large number of previous experiments, similar to this one, it has been documented that women are, on average, too reluctant to compete. That means that in those experiments women, on average, would have earned more money if they had been more willing to compete.

On average men, however, are in these previous experiments found to be too eager to compete. This means that men, on average, would have earned more money if they had been less willing to compete.

Risk Treatment:
In a large number of previous experiments, similar to this one, it has been documented that women are, on average, too reluctant to take on risks. That means that in those experiments women, on average, would have earned more money if they had been willing to take on more risk.

On average men, however, are in these previous experiments found to be too eager to take on risks. That means that men, on average, would have earned more money if they had been less eager to take on risk.

Confidence Treatment:
In a large number of previous experiments, similar to this one, it has been documented that women are, on average, not confident enough. This means that in those experiments women, on average, would have earned more money if they had displayed more confidence.

On average men, however, are in these previous experiments found to be too confident. This means that men, on average, would have earned more money if they had displayed less confidence.

As a reminder, you previously chose to apply the piece-rate pay (the competition pay) to your performance in Part 3.

We would now like you to confirm, or revise, this choice. Immediately thereafter, Part 3 will start (i.e. you will not get the chance to revise your choice again).

Please make your final choice about which payment scheme you would like to apply to your performance in Part 3.

Please click OK when you're ready to continue

The three parts of the experiment where you were asked to do the math task are now finished, and we are moving on to the fourth part of the experiment.

In this part you will make 20 decisions. For each of these, you will be choosing between a certain payment (the option to the left) and a lottery (the option to the right). The certain payment is identical for every decision problem: you will receive 5X ECU for certain. For the lottery you will receive either 10X or 0 ECU. The probability with which you will receive 10X ECU will differ for every decision problem.

Your earnings in this part are determined as follows. First, one of your 20 decisions will be chosen at random (they all have equal probability of being chosen). Second, in case you chose the certain payment (the first option) in the selected question, you will receive 5X ECU. In case you chose the lottery the second option), it will be randomly determined, with the probability stated in the selected question, if you receive 10X or 0 ECU.

Please choose the alternative you prefer in each of the 20 decisions below and please raise your hand if you have any questions.

We have now gotten to part 5 of the experiment.

Here we would like you to guess how well you performed in the math task in **Part 2**, compared to other people who did the same experiment and the exact same task here in the laboratory earlier this year. (In Part 2, you and the other participants all competed against a random opponent.)

We ask you to guess your rank among 20 randomly chosen participants in Part 2. You can receive a bonus of 20 ECU in this part, and the closer your guess is to the truth, the higher the probability that you receive the bonus. If you would like to get information about exactly how this probability is calculated, please raise your hand and an experimenter will come to you.[14]

What do you believe that your rank in Part 2 was compared to these 20 others? Please choose a value between 1 (you believe you were the best) and 21 (you believe you were the worst).

We have now come to the sixth, and final, part of the experiment, in which you will play an investment game.

You will receive detailed instructions on the next screen.

The decision we ask you to make is an investment decision. This decision is also for real money; the result of your decision will be added to your account and paid to you at the end of the experiment.

You start the investment task with a balance of 10 ECU. You choose how much of this amount (from 0 ECU to 10) you wish to allocate to an investment.

The ECU that you choose *not to invest* will be saved in your account and cannot be lost. You will receive these ECU at the end of the experiment.

The value of the ECU you choose *to invest* depends on the success or failure of the investment. The success or failure of the investment will be determined by a computerized random draw, similar to a coin flip. There are two possible outcomes:

- With 50% probability the investment fails and you lose the amount invested.

[14] The handout can be found at the end of this section.

- With 50% probability the investment succeeds and you receive 2.5 times the amount invested.

So, for any amount X that you invest, you will keep 10 - X, regardless of what happens with the investment. If the investment fails, which happens with 50% probability, your earnings from this decision will be 10-X, since you lose the amount that you invested. If the investment succeeds, which also happens with 50% probability, your earnings from the decision will be 10 - X + 2.5*X = 10 + 1.5*X

When you are ready, please enter the amount of EUC you wish to invest on your screen. If you have any questions in the meantime, please raise your hand and an experimenter will come to you.

<Payment Overview>

Thank you for participating in this study! While we prepare your payment, please answer the questions on this, and the following, screens.

How important do you think the following personal characteristics are for whether or not a person decides to compete in round 3 of the math task? (Scale 0-10 from Definitely not important to Definitely very important)

That the person is willing to take on risk
That the person is confident that s/he is better than others at the task
That the person enjoys competing

Please explain your reasoning, and also let us know if there are other factors that you think are important for whether a person chooses to compete or not.

What was your initial decision regarding which payment scheme to choose for round 3?

Piece Rate
Competition

How did you make your initial decision about whether or not to compete in Round 3?

What information did you get before confirming/revising the choice ahead of round 3?

Do you feel that the fact mentioned in this information is true for you?

Yes
No
Unsure

In general, did you find the information helpful and relevant before making the choice ahead of Round 3? Please explain!

Would you describe yourself as a person who is unwilling to take risk, or as someone who is willing to take risk?

Would you describe yourself as a person who is not very confident or as someone who is very confident?

Finally, we ask you to please provide us with some demographic information.

Gender

Age in years

The total money you earned is X

You have a receipt on your desk. Please fill it out, and enter your total earnings in the receipt, <u>rounding up to the nearest 50 cents</u>.

After you have filled out your receipt, please raise your hand. An experimenter will then come by to check the amount. After that you can go to the room next door to collect your payment. Do not forget to take the receipt and the wooden coin with your computer number with you.

B.2 Handout for the Belief Elicitation Task

Payment for the accuracy of your guess is determined in one of two ways:

(1) Performance Pay: You receive 2 Euros if your Round 2 performance exceeds the performance of one random opponent chosen among the 20 participants from the previous session.

(2) Lottery Pay: Receive 2 Euros based on a lottery. There are 21 possible lotteries, all equally likely. Lottery 1 pays out 2 Euros with 100% chance, lottery 2 pays out with 95% chance, and so on. Lottery 21 pays out 2 Euros with 0% chance.

At the end of the experiment, the computer will randomly draw one of the 21 lotteries. The experiment will then automatically implement either the lottery or performance pay depending on the rank you report. Specifically, if your reported rank is smaller than the lottery's number, you will be paid through performance pay. If your reported rank is greater or equal than the lottery's number, you will be paid by the lottery instead. This ensures that accurately reporting your rank maximizes your expected payoff.

Let us illustrate how this works with an example. Suppose that the computer randomly draws lottery 10 to be relevant for payment. This means that individuals who report rank 9 or better will be paid through performance pay. Individuals who report rank 10 or worse will receive lottery pay.

To see why it is optimal to accurately report your rank, note that lottery 10 has a 55% chance of paying out the 2 Euros. For individuals ranked 9th or better, performance pay will result in at least a 60% chance of obtaining the 2 Euros. To see this, note that the 9th ranked individual is worse than individuals ranked 1-8, but better than those ranked 9-21. Since the latter group consists of 12 individuals, the rank 9 individual therefore has a 12/20=60% chance of obtaining 2 Euros in performance pay. By contrast, individuals ranked 11 or worse have at most a 50% chance of receiving payment through performance pay, and hence maximize their payment through lottery pay. This also implies that accurately reporting your rank maximizes the probability that you receive the 2 Euro payment.

B.3 Instructions for the Field Experiment[15]

This is a social science experiment. The experiment has five different parts. In each part there is a possibility to earn money. How much you will earn depends on your choices and your performance. The different parts will be explained as we get to them. At the end of the experiment two of you will be randomly chosen to be compensated in accordance with your choices and performance in one random part of the experiment. It is important that you do not talk to each other or interact in any other way. It is forbidden to use mobile phones or similar devices during the experiment. Do not leave your place at any time during the experiment. Violating any of these rules will disqualify you for compensation.

Any questions?

Let's start with part one of the experiment. In this part your task is to solve as many simple math problems as you can in five minutes. The problems consist of adding five two-digit numbers. Should you get chosen to be compensated for this part you will earn 5 SEK per correct answer.

Any questions?

Sheets with the problems are now distributed, please to not turn them before we say that you can start.

Go ahead and start.

Time is now up. You now have one minute to record your answers before we collect them and distribute the correct answers for your reference.

Let's move on to the second part of the experiment. Your task will again be to solve as many simple math problems as you can in five minutes. Should you get chosen for compensation for this part you will earn 20 SEK per correct answer given that you are performed higher than three randomly selected highs-schools students in the social science track that have previously partaken in this experiment.

Any questions?

Sheets with the problems are now distributed, please to not turn them before we say that you can start.

Go ahead and start.

---

[15] These are the instructions translated to English; the instructions where originally written in Swedish.

Time is now up. You now have one minute to record your answers before we collect them and distribute the correct answers for your reference.

Let's move on to the third part of the experiment. Your task will again be to solve as many simple math problems as you can in five minutes. How you will be compensated should you and this part be chosen for compensation is up to you. You can choose between 5 SEK per correct answer as in part one or 20 SEK if you outperform three randomly selected highs-schools students in the social science track that have previously partaken in this experiment as in part two. Please record your choice before turning the worksheet.

If treated:

Before you make your decision, we do want to give you some additional information. Research has found that men are more prone to compete than women. Further, men, on average, tend to compete too much given that they want to maximize their expected pay-off while women, on average, do not compete as much as they should.

Any questions?

Sheets with the problems are now distributed, please to not turn them before we say that you can start.

Go ahead and start.

Time is now up. You now have one minute to record your answers before we collect them and distribute the correct answers for your reference.

Now it's time for the fourth part of the experiment. We will now ask you to rank your performance in comparison to your peers on a scale from 1-4 where 1 is among the top 25 percent, 2 is not in the top 25 percent but in the top 50 percent, 3 is not in the top 50 percent but not in the bottom 25 percent and 4 is in the bottom 25 percent. Should you and this part be chosen for compensation you will receive 100 SEK if your guess is correct and 0 SEK otherwise.

Any questions?

A sheet is now being distributed where you can record your answer. Raise your hand when you are done, and we will come and collect it.

Finally, it's time for the fifth part of the experiment. In this part of the experiment we will give you the choice between the guaranteed amount of 10/30/50/70/90 SEK or a 50% of winning

100 SEK. Make one choice in each of the five scenarios. If you and this part is chosen for compensation one of the scenarios will be randomly chosen for implementation.

A sheet is now being distributed where you can record your answers. Raise your hand when you are done, and we will come and collect it.

Lastly, before we randomly select who will be receiving compensation, we would like to record your legal gender as it is recorded in your passport. A sheet is now being distrusted where you can record this information. Raise your hand when you have done so, and we will come and collect it.

Online Appendix C: Explaining the Gender Gap in Willingness to Compete

What causes the gender gap in willingness to compete? Van Veldhuizen (2018) develops a method that makes it possible to decompose the gender gap into three components: competitiveness, risk preferences and confidence. The method relies on comparing the raw gender gap in willingness to compete to two novel measures that share a payoff structure with the baseline tournament entry choice but iteratively remove the role of competitiveness and confidence. If, for example, competitiveness is an important driver of the gender gap in willingness to compete, we should expect the gender gap in the control measures to be significantly smaller than in the baseline tournament entry decision. Note that since this approach relies on simple difference-in-difference tests, it is not susceptible to the measurement error critique (Gillen, Snowberg and Yariv, 2018).

The first measure, NoComp, removes the role of competitiveness. Consider a participant with a performance of 10 who expects to have a 75% chance of winning a tournament (reported as a rank of 6 in part 5). When deciding between tournament and piece rate pay in part 3, this participant is therefore implicitly deciding between obtaining some amount (50 ECUs) for sure (piece rate), or obtaining twice that amount (100 ECUs) with 75% chance (tournament). Yet these are also the exact payments faced by this participant in one of the 20 items in part 4 (the item where the lottery pays with probability .75). The key difference is that the relevant item in part 4 is no longer competitive, since neither the lottery nor the safe amount of money are a competition (in contrast to the tournament entry decision). In the example, we can therefore use 75%-lottery decision a non-competitive version of the tournament entry decision. We refer to this decision as our first measure, NoCompetitiveness (or NoComp). Our second measure, JustRisk, takes a similar approach except that it replaces the subjective probability of winning with the participant's actual probability of winning given the empirical distribution of performance across all sessions, which eliminates the role of confidence as well. A more detailed explanation of this method can be found in Van Veldhuizen (2018).

Figure A1: Gender Gap in Willingness to Compete, NoComp and JustRisk

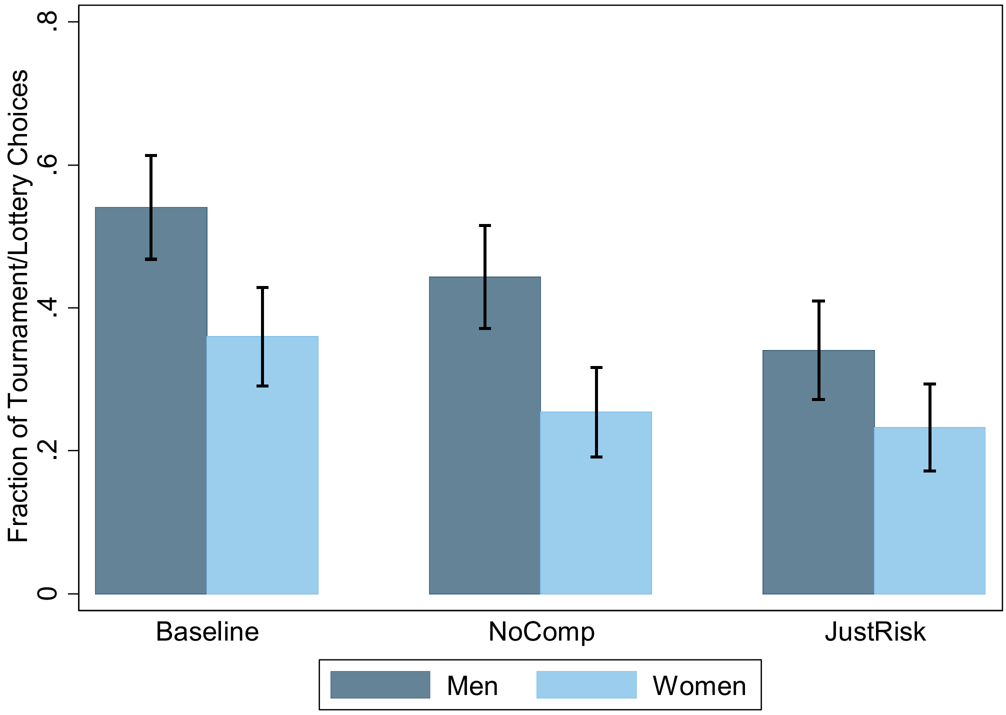


Figure A1 presents the results for the three choices, pooled across treatments. The gender gap is similar in the baseline tournament entry decision (part 3 after advice, 18.1pp) and the NoComp measure (18.9pp), whereas it falls somewhat in JustRisk (10.7pp). The point estimates imply that risk preferences explain about 59% (10.7/18.1, p=0.021, t-test) of the gender gap in willingness to compete. Confidence appears to explain 45% (8.2/18.1, p=0.116, t-test), whereas the point estimate for competitiveness is negative and not significant at -4% (-0.8/18.1, p=0.887). The limited role for competitiveness is consistent with Van Veldhuizen (2018), and may also explain why the Competitiveness treatment is less effective than the other treatments.

Online Appendix D

D.1 Description of the Belief Elicitation Task

The crossover mechanism that we used to elicit beliefs asks participants to choose between obtaining a given payment (20 ECUs) with a known probability (risky option), and obtaining 20 ECUs if their performance exceeds the performance of a random opponent from the comparison sample (uncertain option). Participants faced 21 different decisions, where the uncertain option was always the same and the probability of payment in the risky option varied from 100% to 0% in increments of five percentage points. One of the 21 decisions was randomly selected for payment at the end of the experiment.

Under expected utility theory, it is incentive compatible for participants to prefer the uncertain option if and only if the known probability of the risky payment exceeds their subjective probability of winning the tournament (i.e., the probability of receiving the uncertain payment; Karni, 2009). This in turn implies that it is incentive compatible for participants to switch from risky to uncertain payment once the known probability falls below the subjective probability of winning. The switch point then identifies the subjective probability of winning, i.e., how likely the participant thinks he or she is to win a tournament.

While the crossover mechanism is a theoretically attractive way to elicit the subjective probability of winning (in particular, since it does not require risk neutrality), it has been criticized for its complexity. We sought to mitigate this issue by presenting the mechanism to participants in a straightforward way by asking them to report their rank instead of a probability or switch point. This uses the fact that in two person-tournaments the probability of beating a random opponent from the comparison sample equals (21-rank)/20. We also omitted most of the details of the crossover mechanism from the on-screen instructions, instead putting them on a handout that we made available upon request.

The data suggest that participants understood the task well enough to pick up some expected patterns. For example, the elicited belief was highly correlated to performance in part 1 ($r=-.45$, $p<.0001$), part 2 ($r=-.46$, $p<.0001$) and part 3 ($r=-.44$, $p<.0001$). Elicited beliefs also significantly predicted tournament entry decisions ($r=-.40$, $p<.0001$), and differed significantly by gender (as we showed in Table 2).

D.2 Manipulation Checks for the Laboratory Experiment

While simple advice significantly reduced the gender gap in willingness to compete, the fraction of participants who changed their decision after receiving the advice was relatively small (5.6%). The relatively low number affected does not appear to have been driven by a failure to read the advice: 79% of participants were able to recall the message when asked to do so in the questionnaire, with the remainder typically either leaving the question blank, or recalling another element of the instructions. Instead, the questionnaire data suggest two potential reasons why relatively few people changed their choice after advice. First, even among the participants directly target by advice (women planning to choose piece rate and men planning to choose tournament), only 52% (47% of men, 55% of women) indicated that they felt the advice applied to them personally (the remainder were unsure or said it did not apply), and only 29% thought that the information was relevant or helpful. Second, even among those who thought the information was relevant or helpful, many participants did not actually change their decision after receiving the information. These participants referred either to general characteristics (e.g., "I am someone who is bad at math") or said they saw no reason to change their mind after having just carefully thought through their initial decision. In particular, the fact that participants had already carefully thought through their decision prior to receiving the advice may have reduced the scope for advice to have an effect.

Online Appendix E: Robustness Checks

In this Appendix we present results from robustness tests we mention in the paper.

E1. Alternative specifications for the main treatment effects

Table A3 and Table A4 reprints the results of Table 2 and Table 5 respectively while removing the controls for performance in part 2. The coefficient estimates and p-values are essentially unchanged.

Table A3: Effect of Advice on Tournament Entry in the Laboratory Study

	(1) Advice Effect (All)	(2) Advice Effect ("Competitiveness")	(3) Advice Effect ("Risk")	(4) Advice Effect ("Confidence")
Female	0.100***	0.054	0.145**	0.108*
	(0.024)	(0.036)	(0.046)	(0.046)
Constant	0.005	0.018	0.016	0.015
	(0.009)	(0.018)	(0.017)	(0.015)
Observations	374	127	123	124
R-squared	0.043	0.016	0.076	0.048

Notes: Robust standard errors are in parentheses. The dependent variable ("Advice Effect") measures the response to receiving advice. It is equal to 1 for participants who switched from piece rate to tournament after hearing the advice, -1 for those who switched in the opposite direction, and 0 for those whose decisions were not affected by advice. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table A4: Effect of Advice on Tournament Entry in the Field Study

	(1) Compete	(2) Compete	(3) Compete
Advice	0.033	-0.303***	0.287***
	(0.051)	(0.071)	(0.065)
Constant	0.380***	0.562***	0.245***
	(0.040)	(0.059)	(0.047)
Subgroup	All	Men	Women
Observations	268	117	151
R-squared	0.277	0.374	0.360

Notes: Robust standard errors are in parentheses. The dependent variable is the tournament entry decision in part 3 (1-tournament, 0-piece rate). "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

E2. Alternative measure of efficiency.

An alternative way to measure the efficiency effects of advice is to look at the earnings foregone by participants who failed to choose the payoff maximizing payment scheme. Intuitively, average individuals have very little to lose from choosing either option since their expected earnings will be similar for both options. By contrast, very high-performing individuals have a lot to lose from foregoing an almost certain tournament victory in favor of the piece rate. Table A5 shows the results for this alternative measure in the laboratory study. The results are very similar: high-performing women gain from receiving advice whereas low-performing women lose, leading to a net zero effect overall.

Table A6 prints the corresponding results for the field study. As with the lab, the results are qualitatively similar, especially when accounting for the fact that decisions not to compete when one should, due to the higher stakes involved, will mechanically generate a larger expected loss than the decision to compete when one should not. The scale difference between the results of the laboratory and the field studies are due to the relevant currency being EUR in the laboratory, and SEK (10 SEK ≈ 1 EUR) in the field.

Table A5: The Effect of Advice on Expected Foregone Earnings in the Lab

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max
Advice	-0.001	-0.007	0.005	-0.012	0.000	-0.130*	0.157**
	(0.020)	(0.009)	(0.038)	(0.012)	(0.015)	(0.054)	(0.048)
Constant	0.772***	0.703***	0.838***	0.908***	0.435***	1.381***	0.228***
	(0.080)	(0.113)	(0.112)	(0.190)	(0.071)	(0.191)	(0.054)
Gender	All	Men	Women	Men	Men	Women	Women
Performance	All	All	All	High	Low	High	Low
Observations	374	185	189	105	80	100	89

The dependent variable is the participant's expected foregone earnings. Robust standard errors in parenthesis. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table A6: The Effects of Advice on Expected Foregone Earnings in the Field

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max	Payoff Max
Treated	-4.184*	-3.599*	-4.599	2.738	-6.417***	-17.906*	4.249*
	(1.941)	(1.673)	(3.172)	(3.753)	(1.790)	(6.987)	(1.686)
Constant	7.571***	6.277***	8.550**	3.219	7.845***	18.092*	1.912*
	(1.809)	(1.300)	(3.027)	(2.130)	(1.597)	(6.986)	(0.925)
Gender	All	Men	Women	Men	Men	Women	Women
Performance	All	All	All	High	Low	High	Low
Observations	268	117	151	36	81	59	92
R-squared	0.019	0.059	0.003	0.016	0.160	0.238	0.075

Robust standard errors in parentheses. The dependent variable is the participant's expected foregone earnings given their part 2 performance. "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

E3. Permutation tests

In the field study treatment is on the class level rather than the individual level. There are 12 classes in total (six treated and six non-treated). This means that classical methods of inference might not be valid. We therefore use randomization inference to confirm that the results reported in the main article are valid. Below we describe this methodology and reprint all the main results estimated using this approach.

We collapse the data to the classroom (treatment) level and estimate treatment by the following equation:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

where T_i is an indicator if class i was treated or not, Y_i is the share of students in class (or subgroup in class) i that selected into competition or that chose the payoff maximizing payment scheme, ε_i is the error term. β measures the causal effect of treatment. The regressions will be weighted with the relevant frequencies (the number of students in the class/subgroup in the class). This is done to give β the same interpretation the main article: the treatment effect on the probability that an individual competes/makes the payoff-maximizing choice.

Given randomization of treatment, strict exogeneity holds and hence β will be an unbiased estimator of the treatment effect. The small sample ($N = 12$) does however imply that we cannot rely on the asymptotic properties that underlie most conventional econometric methods for drawing inference. Inference regarding the treatment effects will therefore be based on randomization inference. In a setting where treatment is random, randomization inference provides an exact test of sharp hypotheses no matter the sample size (Young, 2017).

The motivating idea behind randomization inference is that, given the sample of participants (in this case classes), the only stochastic element is the allocation of treatment. A null-hypothesis of no treatment effect on any of the participants ($Y_i(t_i) = Y_i(0) \ \forall \ i$) can then be tested by calculating all possible realizations of a test statistic and rejecting the null if the observed realization, of the test statistic in the actual experiment, is extreme enough.[16]

As calculating the outcome for each possible perturbation is tedious, we use simulations. We resample the treatment variable 1000 times for each outcome of interest and subgroup.[17] For each perturbation Regression 1 is estimated and the corresponding β is documented. The fraction of perturbations that yielded a more extreme value of β than the one observed in the actual experiment is computed. This fraction is then used as the randomized p-value.[18] This is implemented in Stata using

[16] See Young (2017) for a detailed description of the concept.
[17] 200 perturbations have been shown to be sufficient for the p-value to stabilize (Young, 2017).
[18] Note that it is possible for the β in a given perturbation to be the same as in the observed data. This usually happens when the treatment group in the perturbation is the same as in the actual sample or when both the

the "ritest" command described in Heβ (2017). As three classes where sampled into the treatment group in 2011 and three in 2014, we will resample using the year of treatment as a stratum.

Table A5 reprints the results of Table 5 using randomization inference and while removing the controls for performance in part 2 (we cannot control for this on the individual level in this analysis as the data is collapsed to classroom level). The coefficient estimates and are essentially unchanged and the results are still highly significant.

Table A7: Advice Effect on Willingness to Compete in the Field Study (Permutation Test)

	(1) Compete	(2) Compete	(3) Compete
Advice	0.033	-0.303***	0.287***
	(0.350)	(0.000)	(0.000)
Constant	0.380	0.562	0.245
Subgroup	All	Men	Women
Observations	12	12	12
R-squared	0.0886	0.8317	0.8617

Notes: Randomization p-values are in parentheses. The dependent variable is the tournament entry decision in part 3 (1-tournament, 0-piece rate). "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table A6 reprints the results of Table 5 using randomization inference and while removing the controls for performance in part 2 (we cannot control for this on the individual level in this analysis as the data is collapsed to classroom level). The coefficient estimates and are essentially unchanged and the results are still highly significant.

Table A8: Advice Effect on Efficiency in the Field Study (Permutation Test)

	(2) Payoff Max	(3) Payoff Max	(4) Payoff Max	(5) Payoff Max	(6) Payoff Max	(7) Payoff Max
Advice	0.199***	-0.113***	0.334***	0.051***	0.457**	-0.208***
	(0.000)	(0.000)	(0.000)	(0.00)	(0.000)	(0.000)
Constant	0.681	0.864***	0.589***	0.729***	0.467***	0.904***
Female	No	No	No	Yes	Yes	Yes
Performance	All	High	Low	All	High	Low
Observations	12	12	12	12	12	12

Notes: Randomization p-values are in parentheses. The dependent variable is whether participants chose the expected payoff-maximizing payment scheme given their part 2 performance and the part 2 performance of all other participants in this study. "Advice" is a binary variable specifying the treatment status (1-advice, 0-control treatment). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

outcome and the treatment variable are binary variables. We use strict inequalities and hence these cases will be counted towards acceptance so as not to over-reject.