Dany Kessel
Johanna Mollerstrom
Roel van Veldhuizen

Can simple advice eliminate the gender gap in willingness to compete?

Discussion Paper

SP II 2021–202

August 2021

WZB Berlin Social Science Center

Research Area
Markets and Choice

Research Unit
Market Behavior

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Dany Kessel, Johanna Mollerstrom, and Roel van Veldhuizen
Can simple advice eliminate the gender gap in willingness to compete?

Affiliation of the authors:

Dany Kessel
Södertörn University and IFN Stockholm

Johanna Mollerstrom
George Mason University, IFN Stockholm and DIW Berlin

Roel van Velhhuizen
Lund University and WZB Berlin

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

Can simple advice eliminate the gender gap in willingness to compete?

by Dany Kessel, Johanna Mollerstrom, and Roel van Veldhuizen*

As a recent literature has demonstrated, men and women differ in their willingness to sort into competitive environments. In particular, men are more willing than women to compete. We investigate whether it is possible to reduce the gender gap in willingness to compete through an information intervention that informs participants of the gap and advises them about the potential earnings implications. We find that this simple information intervention reduced the gender gap, both in a laboratory study at a German university and in a field study with Swedish high school students. Whereas some participants (primarily high performing women) benefited from the intervention, others lost out. We discuss the implications for efficiency and policy.

Keywords: Gender Differences, Competitiveness, Advice, Experiment

JEL classification: C91, D91, J16

---

*E-mail: dany.kessel@sh.se, jmollers@gmu.edu, rrvanveldhuizen@gmail.com. We are very grateful to the editor of this special issue, and to two anonymous referees for excellent comments and suggestions that have substantially improved the paper. We also thank Anat Bracha, 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, at the Behavioral Workshop at the Swedish Institute for Social Research, and in the session "Interventions to close the gender gap: What works, and what can backfire?" at ASSA 2020 for very valuable comments. The laboratory experiment reported here was pre-registered at https://www.socialscienceregistry.org/trials/3487.

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

1. Introduction

Why do men and women differ in their labor market outcomes? A recent literature in experimental economics and labor economics suggests that these differences may be driven in part by gender differences in psychological traits (c.f. Blau and Kahn, 2017). In particular, a series of laboratory experiments demonstrate that men are more willing than women to enter competitions (Niederle and Vesterlund 2007, Datta Gupta et al. 2013; see Croson and Gneezy, 2009; Niederle and Vesterlund, 2011, and Niederle, 2015 for overviews). The inherent competitiveness of activities like job applications and promotion contests may hence help to explain the observed gender gap in labor market outcomes. Indeed, differences in willingness to compete measured in laboratory settings have been found to predict, and partly explain, gender differences in important career choices outside the lab (see e.g. Buser et al, 2014; 2017; 2020).

Given that men and women differ in their willingness to compete, and that these differences appear to predict essential career choices, it is important to understand if 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 the seminal paradigm of Niederle and Vesterlund (2007), in which participants solve addition problems and can choose between receiving tournament or piece-rate pay. The main way in which our study differs from previous ones is that, prior to participants' final decision, we inform them of the evidence supporting the existence of a gender gap in willingness to compete. We also tell them that men and women in previous experiments 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 and 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; Gillen, Snowberg and Yariv, 2019). Therefore, we also test whether emphasizing one driver over another changes the effect of the information intervention on willingness to compete.

Our hypothesis, as 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. However, this was not a foregone conclusion. It could, for example, also be argued that the information intervention puts additional emphasis on the gender stereotype that women compete less than men, and thus could cause behavior to align even more with the stereotype (c.f. Eagly, 1987; Akerlof and Kranton, 2000; Rudman and Glick, 2001). If so, then presenting participants with information that triggers the stereotype could lead men to compete even more and women to compete even less. This could, in theory, increase the gender gap in willingness to compete.

We report three main results. First, our simple advice intervention decreased the gender gap in willingness to compete in both experiments, in line with our pre-registered hypothesis. Second, our intervention had a mixed impact on advisees' expected earnings. While some advisees, primarily high-performing women, benefited from advice, others, primarily low-performing women, were hurt by it. Overall, however, our intervention did increase the frequency with which high performers entered, and won, the competitions – a feature which can be helpful from the perspective of an employer if the competition is used to identify the best performers. Third, our laboratory data demonstrate that the advice effect did not differ significantly depending on whether the advice emphasized the role of risk preferences,

[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.

overconfidence or competitiveness, respectively. Overall, the results imply that simple advice can be an effective way to reduce the gender difference in willingness to compete and increase 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 include switching to within-gender competitions (Datta Gupta et al., 2013) or team-based competitions (Dargnies, 2012), instituting affirmative action policies like gender quotas (Balafoutas and Sutter, 2012; Niederle et al, 2013), introducing mentorship or sponsorship programs (Baldiga and Coffman, 2018; Porter and Serra, 2020), using priming (Balafoutas et al., 2018) 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; Apicella et al., 2020; Demiral and Mollerstrom, 2020). Other studies closely related to our work provide relative performance feedback or highly personalized tournament entry advice (Wozniak et al., 2014; Ertac and Szentes, 2011, Brandts et al., 2015, Berlin and Dargnies, 2016).[2]

While most of these previously studied interventions have proven at least somewhat effective at reducing the gender gap in willingness to compete, they are quite drastic and potentially expensive. As a result, they may not be practical or even feasible to implement. On the other hand, the policy intervention that we study, i.e., informing participants of the gender gap in willingness to compete, is simpler and potentially less costly – at least as long as data that can inform the advice is readily available, as is the case here. Therefore, it could be

[2] A parallel literature has studied determinants of gender differences in performance. For example, Schram, Brandts and Gërxhani (2019) and Shurchkov (2012) study the role of status-ranking and time pressure respectively, whereas Iriberri and Rey-Biel (2017) show that gender differences in performance only emerge when the task is perceived as favoring men and participants are primed with the presence of a rival.

relatively easily implemented in a variety of contexts – for example via career advisors in schools and universities, at career fairs, or through information campaigns using role models.[3]

Our work also relates to studies that seek to understand, and quantify, experimenter demand effects (e.g., Zizzo, 2010; de Quidt et al., 2018). In general, advice consists of both an "informational" and an "encouragement" component (see e.g., Sniezek and Buckley, 1995; Bonaccio and Dalal, 2006). Whereas the first is about making sure the advisee is *aware* of the consequences of different actions, the latter is aimed at *encouraging* the agent to pick a particular action. In other words, advice serves not only to convey information, but also to signal what the advisor regards as the appropriate behavior. In that sense, the encouragement component of advice is similar to an experimenter demand effect. For example, Zizzo (2010) defines the experimenter demand effect as occurring when participants change behavior "due to cues about what constitutes appropriate behavior" (p. 75).

In our setting, the adviser and the experimenter are the same person, and to the extent that the advice given in the study has not only an informational but also an encouragement effect, our results can be thought of partly as the result of a demand effect. With that said, our treatment effects in the field study are comparable or larger than what, for example, de Quidt et al. (2018) document even when they attempt to maximize the demand effect by explicitly requesting participants in one treatment to choose one option and participants in another choose the other option.[4] In addition, we find some evidence that advice shifts participants' beliefs

[3] 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), Guillen and Hakimov (2018), Gneezy et al. (2020) and Mollerstrom and Sunstein (2020).
[4] In their treatments that attempt to maximize the experimenter demand effect, De Quidt et al (2018) find an experimenter demand effect of 0.6 standard deviations (SD). The corresponding effect for the more subtle ("weak") experimenter demand effect treatments is 0.13 SD. Noting that the standard deviation of a binary variable equals $\sqrt{p(1-p)}$, it can be shown that these effect sizes correspond to effect sizes of 25-29 percentage points for the strong and 6-7 percentage points for weak experimenter demand effect, depending on the gender and the experiment. The effects we observe are 29-30 percentage points for both genders in the field and 0pp for men and 10pp for women in the lab.

about the expected benefits of entering the tournament, suggesting that the informational component played a role as well. As a result, we argue that it is unlikely that our results come about only because of the encouragement component of the advice, and that the informational component is also important in making the advice effective.

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 paper. We provide this information here, as there are some differences between the two studies, – for example including 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.

The remainder of the paper proceeds as follows. Section 2 describes the design, analysis and results of the laboratory study. Section 3 does the same for the field study. Section 4 discusses the efficiency implications of our intervention, and related ethical aspects. Section 5 concludes.

2. Laboratory Study

We conducted the laboratory study at the experimental economics laboratory of the Technical University Berlin in Germany in November 2018. We pre-registered this study in the AEA registry. The pre-analysis plan is reprinted in Online Appendix A, and our analysis follows this plan. For each session, we invited 28 participants (14 men and 14 women). Following standard protocols for our laboratory, we invited only participants who had participated in a maximum of 10 previous experiments, and had at most one instance of failing to show up to a session they had registered for. -In total, 374 people participated: 185 men and 189 women. The 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 randomly at the individual level.

2.1 Design

Participants were randomly assigned to a computer upon entering the laboratory. They were 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 would always be 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.[5] 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 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. Otherwise, they were paid nothing (in case of a tie, participants had a 50 percent chance of being paid the 10 ECUs per correct problem). They were not informed about their performance, or whether they won or lost, until the very end of the experiment. Ahead of the third part of the experiment, participants were instructed to choose whether they wanted to apply the piece-rate or the tournament pay to their performance in Part 3. If they chose

[5] EUR/USD ≈ 0.89 at the time when the experiment was conducted.

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.[6]

Immediately following their initial tournament entry decision in Part 3, we presented participants with evidence supporting the gender gap in willingness to compete (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 this information intervention was to induce women to compete more often and men to compete less.

We ran three different treatments that framed the information in three different ways. In the treatment "Competitiveness," we informed participants that previous research had found that men compete too much and women compete too little. That is, 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 greater risks in similar experiments (and the reverse for men). 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). In that study, the average woman would have increased her earnings by being more willing to compete, while the opposite was true for men. Tournaments are riskier than the piece rate, and confidence positively predicts tournament entry. Thus, these results also imply that women (men) could have increased their earnings by taking on more (less) risk or being more (less) confident in the experiment. Subsequent studies, including two experiments run at the same

[6] Note that the fact that participants are compared with their competitors' past performance also implies that participants do not need to contemplate whether other participants are reacting to the advice in a way that could impact their own tournament entry decision (e.g., if the advice leads to more high quality competitors). Also note that our implementation differs from Niederle and Vesterlund (2007) in two ways. First, participants competed against only one competitor rather than three. Second, they had four minutes for the task in each round, instead of five. Similar changes implemented in previous work (e.g., Niederle et al 2013; Buser et al. 2014 or Buser et al. 2021) did not substantively affect the gender gap in willingness to compete.

laboratory (Van Veldhuizen, 2018 and Buser et al. 2021), have tended to find similar results in terms of both the gender gap in tournament entry[7] and the payoff implications of this gap.[8] While we originally based our advice on the results of Niederle and Vesterlund (2007), it was therefore also consistent with the findings of a number of subsequent studies.

The specific text that we used when giving the advice had an explicit informational component, but was only implicitly encouraging, given that we did not instruct participants exactly what they should do. We chose this structure as it is close to how advice is often formulated outside the laboratory: providing information suggesting that some action may maximize one criterion or another, but without explicitly mandating people to take that particular action. Translated from German it reads:

> *"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*

[7] Most studies building on Niederle and Vesterlund (2007) find a gender gap in tournament entry, including two the two previous experiments run at the same laboratory (Van Veldhuizen, 2018 and Buser et al 2021). At the same time, non-significant gender gaps have been observed among pre-school children (see e.g. Bartling et al., 2012 and Samek, 2013), among people in certain geographical locations (c.f. Gneezy et al., 2009 who study a matrilineal society in India) or in settings where stereotypically female tasks are employed (c.f. Shurchkov, 2012 who use a verbal task instead of a math task, as do Boschini et al., 2019). See Croson and Gneezy (2009), Niederle and Vesterlund (2011) and Niederle (2015) for surveys of this literature.

[8] Earnings implications of existing gender gaps in tournament entry are harder to track because many studies do not report them. But when looking at four studies that do (Balafoutas and Sutter, 2012; Dargnies, 2012, Niederle et al., 2013 and Buser et al. 2014), it appears that men could indeed have increased their earnings by reducing tournament entry (by 26pp, 18pp, 24pp and 11pp respectively), whereas it is less clear that women compete too little (the corresponding fractions are 2pp, 10pp, -2pp and 12pp respectively). In the two aforementioned studies run at the same lab, we find that men, on average, competed 11 percentage points too much whereas women competed 11 percentage points too little.

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 as to whether to apply the piece-rate or the tournament pay in Part 3. We 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. 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 ensured that the stakes in Part 4 were similar to the stakes in Part 3 for each individual in the experiment. While participants were informed about the size of the fixed payment and lottery options, they did not know that these depended on their own previous performance. The only difference between the 20 items in the choice list was the probability of winning the lottery, which varied from 5 percent to 100 percent 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. Taken together, the decisions in Part 4 are a measure of risk preferences, directly tailored to tournament entry decisions.[9]

In Part 5 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., 2021; 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, which they were asked to allocate between two options. The safe option returned the money invested at a rate of 1:1. The risky investment had a 50 percent chance of returning the money at a rate of 2.5:1, but otherwise would not return any money. The more risk-averse the participant, the less she should invest. Note that none of the measures of risk aversion and confidence collected in Parts 4-6 should be included as controls in regressions investigating the advice effect as they may be influenced by the advice itself and hence also by the treatment.[10]

---

[9] Part 4 is based on Van Veldhuizen (2018). 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.

[10] Part 4-6 were included in order to enable us to investigate the extent to which our sample is comparable to other studies with regards to confidence and risk aversion, and so that we could see if our treatments impact risk aversion and confidence. Results from, and further discussion of, the data from parts 4-6 can be found in Appendix C.

After Part 6, participants reviewed an overview of their earnings from the experiment. They also answered a questionnaire that elicited their age, gender and field of study. The questionnaire also asked about participants' risk preferences and self-confidence on Likert-scales, and included several questions about participants' reasons for making a choice in Part 3. It also asked participants what reasons they thought played a role in their fellow participants' decisions in Part 3. We then paid the participants individually and dismissed them from the laboratory.

2.2 Results

Table 1 presents choices and outcomes from the laboratory experiment, by gender. In line with previous research, we see that men were 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 scored slightly higher on the task in all three parts, though the difference is only significant at the 10 percent level in Part 2. Also in line with previous research, we find that men were more confident (they report lower ranks) and took significantly more risks in both the price list task and the investment game. Note that, following the pre-analysis plan, all the p-values in this section, and in the paper in general, are based on two-sided tests of proportions (binary variables), two-sided t-tests (non-binary variables) and linear regressions. However, all of our results are robust to using (ordered) probit or logit instead of linear regressions, and using different non-parametric tests instead of the t-tests and tests of proportions. For the main analysis, we present the results of these robustness checks in Appendix E.

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 payoff of tournament and piece rate pay for each participant based on their probability

Table 1: 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) Compete, after advice	0.541	0.360	0.181***
	(0.500)	(0.481)	(0.051)
(7) Should compete but does not, before advice	0.184	0.360	-0.176***
	(0.388)	(0.481)	(0.045)
(8) Should not compete but does, before advice	0.151	0.085	0.067*
	(0.359)	(0.279)	(0.033)
(9) Should compete but does not, after advice	0.178	0.307	-0.128**
	(0.384)	(0.462)	(0.044)
(10) Should not compete but does, after advice	0.151	0.138	0.014
	(0.359)	(0.345)	(0.036)
(11) Number of risky choices, Part 4	7.076	5.905	1.171**
	(3.316)	(3.564)	(0.356)
(12) Beliefs about rank, Part 5	8.000	10.444	-2.444***
	(3.907)	(3.854)	(0.401)
(13) 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. Rows 7 and 8 report the fraction of participants who could have maximized their expected payment by choosing tournament (piece rate), but chose the piece rate (tournament) instead before advice. Rows 9 and 10 present the same information after receiving advice. 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. Standard errors and significance levels in column 3 are based on t-tests with robust standard errors (rows 1-3 and 11-13) and tests of proportions (rows 4-10) respectively. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

of winning against a single opponent randomly drawn from the sample of all our participants. We can then define the fraction of participants that “should compete” as the fraction of participants who would have maximized their expected payment by choosing tournament pay. For men, the share that should compete is similar to the share that did actually compete, whereas for women it is much higher (compare Row 4 and 5 of Table 1). Indeed, 36 percent of women would have benefited from changing their decision to competing after receiving advice (Row 7). By contrast, only 9 percent of women would have benefited from changing in the opposite direction (Row 8). For men, the corresponding percentages are 18 percent and 15 percent respectively, suggesting that men in our study would not, on average, have benefited from changing their decisions about whether to compete. In other words, the women in our study did indeed "compete too little"; however, in contrast to Niederle and Vesterlund (2007), the men in our study did not "compete too much," even prior to receiving advice. This highlights the importance of basing advice on rich, previously gathered data, and illustrates how it can be beneficial to target the advice to specific populations, if data is available that make that possible.[11]

We now continue to examine the effect of advice on tournament entry. Figure 1 plots the distribution of the advice effect separately by treatment and gender; Table 2 presents the results of the pre-registered test of the treatment effect, which controls for gender differences in performance.[12] For women, advice increased willingness to compete in all three treatments (p=0.024, p=0.003 and p=0.006 in treatments Competitiveness, Risk and Confidence respectively, row 3 in Table 2). In fact, all women who responded to the advice did so in the direction suggested by the advice, by shifting from piece rate to tournament. For men, there is essentially no effect in any of the three treatments (p=0.223, p=0.356 and p=0.441 in treatments

[11] Note, however, that men did "compete too much" across two experiments run at the same laboratory, as discussed in footnote 8, with a total sample of more than 1300 participants. Potential reasons for why men did compete too much in these previous experiments but not in the laboratory experiment reported here include design features such as the number of competitors and the language that the experiment was run in.

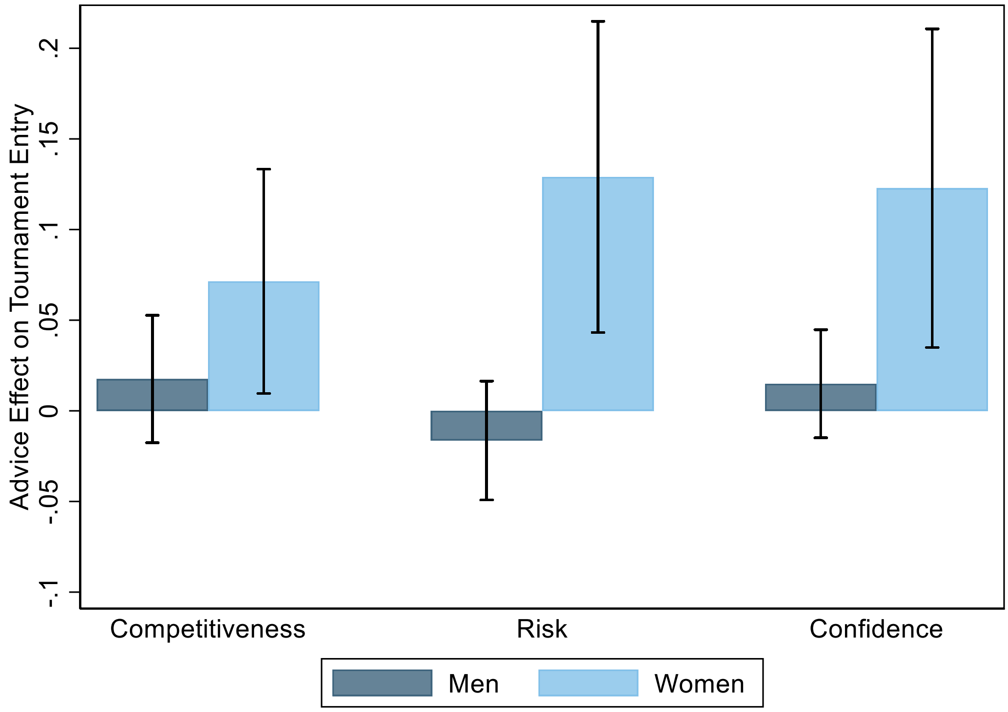
[12] Given that the performance difference between men and women was significant at the 10% level in Part 2, we investigate treatment effects using linear regressions that control for performance in Part 2, as specified in the pre-analysis plan. However, all of the results in this section are robust to using a two-sided t-test, Wilcoxon test, Fischer exact test or ordered probit, see Online Appendix E.

Competitiveness, Risk and Confidence respectively, row 2 in Table 2). Indeed, across the three treatments, advice increased female willingness to compete by 10.6 percentage points (10.5 percentage points when controlling for performance) relative to a baseline of 25.4% ($p<0.001$). By contrast, it had no effect on men's average willingness to compete ($p=0.494$, the baseline was 53.5%).

Given that advice increased tournament entry for women and did not have a significant effect on men, we might expect advice to also reduce the gender gap in willingness to compete. Column 1 of Table 2 shows that this was indeed the case when we pool the data from all three treatments: the gender gap fell by 9.8 percentage points when controlling for Part 2 performance ($p<0.001$). Hence, simple advice may successfully reduce (albeit in this setting not eliminate, c.f. row 6 in Table 1) 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 of Table 2 present similar analyses 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 the smallest in magnitude and 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 significantly so when it advised them to be more competitive. However, the advice effect does not differ significantly across treatments.[13] We therefore conclude that small variations in the wording of the advice do not seem to have a major effect on its effectiveness. In Appendix C, we examine the role of confidence, competitiveness and risk attitudes in explaining the (lack of) differences between the three advice treatments.

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

[13] The p-value for the difference between the female coefficients in the respective treatments is 0.607 for risk and confidence, 0.278 for confidence and competitiveness, and 0.106 for risk and competitiveness (difference-in-difference tests).



Notes: The figure displays the advice effect on tournament entry by treatment and gender. The error bars represent 95 percent 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")
(1) Female	0.098***	0.045	0.145**	0.110*
[diff-in-diff]	(0.024)	(0.035)	(0.046)	(0.047)
(2) Constant	0.007	0.024	-0.016	0.012
[effect on men]	(0.010)	(0.019)	(0.017)	(0.016)
(3) Female+Constant	0.105***	0.069*	0.129**	0.122**
[effect on women]	(0.022)	(0.030)	(0.043)	(0.044)
Score Controls	yes	yes	Yes	yes
Observations	374	127	123	124
R-squared	0.046	0.026	0.076	0.052

Notes: Linear regression estimates; 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. The independent variable ("Female") is equal to 1 for female and equal to 0 for male participants. All regressions control for score in Part 2 (demeaned). The third row tests whether the sum of the first two coefficients is significantly different from zero. The first row corresponds to the difference-in-difference test, the second and third row are the advice effect on men and women respectively. +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. This is particularly true when the advice encourages women to be less risk-averse or more self-confident. By contrast, men seem to be unaffected by advice. Taken together, this also implies that simple advice reduces the gender gap in willingness to compete.

3. Field Study

We also conducted a field study with high school students in Sweden, allowing us to test whether simple advice has a similar effect on behavior in a different setting and with a different population. Specifically, the field study was conducted between 2011 and 2014 in twelve social science classes at Nacka High School in a suburb of Stockholm, Sweden.[14] Each of the twelve classes 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 by the experimenter, and correct answers were given on the whiteboard at the end of the experiment. The experiment took place during class time with the teacher present. The students were not told beforehand about the experiment, so they were unable to sort into the experiment. In addition to the experimenter and the teacher, an assistant who helped distribute worksheets and collect answers was present. In total, each session lasted around 30 minutes.

3.1 Design

The main body of the experiment consisted of three parts. Each part was five minutes long and required participants to solve 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

[14] The reason for conducting the experiment over several years was to ensure a sufficiently large sample size as, in a given year, the number of social science classes at Nacka High School is limited.

awarded participants 5 SEK per correct answer.[15] In the second part, participants were paid according to a tournament payment scheme. The main way in which this experiment differed from the laboratory study is that the field study 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.[16] Before 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 to either receive or not receive the following message just before making their third part decision (translated from Swedish):

"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 were elicited. In Part 4, the participants were asked

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

[16] 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.

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 percent 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 monetary compensation according to the scheme outlined above. No additional personal information was collected about the participants to ensure anonymity.

3.2 Results

Descriptive statistics and the willingness to compete for both the control group and the treatment group are reported in Table 3. As in the previous section, all p-values in this section are based on two-sided tests of proportions (binary variables), t-tests (non-binary variables), or linear regressions. The first three rows show that women outperformed men in all three parts. However, the difference is only statistically significant in the first part. This pattern departs somewhat from the laboratory experiment where, if anything, men performed slightly better. It is nonetheless comparable to the field experiment of Tungodden (2019) with Norwegian high school children (see also Dreber et al., 2014). The fourth row reports the fraction of participants that should have competed based on their performance in Round 2. More women than men should have competed, but the difference is not statistically significant. The reason the fraction of participants who should have competed is lower as compared to our laboratory study is that in the laboratory the participants competed against one other person, while in the field study, they competed against three.

Table 3: 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) Compete, treatment group	0.259	0.534	-0.276**
	(0.442)	(0.502)	(0.084)
(7) Should compete but does not, control group	0.051	0.218	-0.167**
	(0.222)	(0.416)	(0.055)
(8) Should not compete but does, control group	0.271	0.051	0.220***
	(0.448)	(0.222)	(0.063)
(9) Should compete but does not, treatment group	0.069	0.027	0.042
	(0.256)	(0.164)	(0.037)
(10) Should not compete but does, treatment group	0.052	0.192	-0.14*
	(0.223)	(0.396)	(0.058)
(11) Beliefs about rank, control group	2.525	2.244	0.282+
	(0.935)	(0.983)	(0.166)
(12) Number of risky choices, control group	2.390	2.013	0.377*
	(0.929)	(0.781)	(0.146)
Observations (total)	117	151	268
Observations (control)	59	78	137

Notes: Columns "Men" and "Women" show averages for the respective genders (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. Row 7 and 8 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. Rows 9 and 10 present the same information for the treatment group. Belief is the reported rank (with 1 being the lowest performance quartile and 4 being the highest). Number of Risky Choices is the number of times the risky option was chosen in the risk aversion elicitation task. Standard errors and significance levels in column 3 are based on t-tests with robust standard errors (rows 1-3 and 11-12) and tests of proportions (rows 4-10) respectively. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Rows 5, 7 and 8 report the comparative behavior and outcomes in the control group as a baseline for the treatment effects. In Row 5, we can see that about 56 percent of men competed, while only 24 percent of women did. The difference is highly significant, and overall the results are very similar to those in the laboratory study. In Rows 7 and 8 we report the share of participants who fail to choose the expected payoff-maximizing option. Among the 32.2

percent of men who failed to choose the expected payment maximizing option, 84 percent did so by competing when they should not have done so. For women, the corresponding percentage is only 19 percent. This difference is statistically significant (p<0.0001, test of proportions).

Finally, in Rows 11 and 12, 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 our laboratory study.

Table 4 reports the main analysis of the effect of advice on the participants' decision to compete. Even though we do not see any significant differences between men's and women's performance in Part 2, we include the demeaned score in Part 2 as a control to remain consistent with how we analyze the data from the laboratory study.[17] As a first step, Column 1 shows that our treatment had no overall effect on willingness to compete when pooling across genders. However, Columns 2 and 3 demonstrate that the aggregate results obscure that the treatment had a large effect when looking at men and women separately. Men became 30 percentage points *less* likely to compete, while women became 29 percentage points *more* likely to compete; the difference between these two effects is significant (p<0.0001, difference-in-difference test). That is, the initial gender difference of 32 percentage points was not only eliminated but almost entirely reversed (c.f. also row 6 in Table 3). Thus, even though the effects go in the same direction, participants appear to have reacted more strongly to advice in the classroom, than in the laboratory.[18]

Table 4: Effect of Advice on Tournament Entry in the Field Study

	(1)	(2)	(3)

[17] The results are robust to excluding this control. They are also robust to using probit, t-tests, tests of proportions and several other tests, as well as to using randomization inference on the classroom (treatment) level. We refer to Online Appendix E for these robustness checks.
[18] Performance is not affected by the advice; average performance is nearly identical in the treatment and control group in part 3 (difference of 0.062 points; p=0.890).

	Compete	Compete	Compete
Advice	0.033	-0.298***	0.289***
	(0.051)	(0.073)	(0.064)
Constant	0.384***	0.568***	0.244***
	(0.038)	(0.057)	(0.044)
Score Controls	Yes	Yes	Yes
Subgroup	All	Men	Women
Observations	268	117	151
R-squared	0.277	0.374	0.360

Notes: Linear regression estimates; 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.

Overall, the results of the field study are similar to the laboratory study, in that simple advice reduced the gender gap in willingness to compete. However, the size of this effect is larger in the field study than in the laboratory.

4. The Efficiency Implications of Simple Advice

How did our intervention impact the efficiency of overall outcomes? Previous work about willingness to compete has approached the question of efficiency in two ways. The first approach attempts to capture efficiency by considering the expected outcomes of individual participants. In our pre-analysis plan, we included the two measures proposed by Niederle and Vesterlund (2007): whether a participant chose the expected value maximizing payment scheme and, if not, their expected foregone earnings from choosing the expected payoff-inferior option. The second class of measures captures the perspective of the employer or society: these measures include the average quality of tournament entrants as well as the gender gap in tournament entry per se (see e.g., Balafoutas and Sutter, 2012; Niederle, Segal and Vesterlund, 2013).

We start by considering the individual-based measures of efficiency. Given that advice made women more likely to compete and that women, on average, competed too little prior to receiving advice, it seems plausible that advice would increase the fraction of women who chose the payoff-maximizing payment scheme. Table 5 presents the effect of advice on this measure of efficiency in the laboratory (panel 1) and the field (panel 2). The results are grouped by gender and by those who should compete (high ability) and those who should not (low ability).[19]

In the laboratory, 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, while 11 participants shifted in the opposite direction. Hence, the advice was not universally beneficial to participants, and it did not significantly increase the number of participants choosing the payoff-maximizing option (p=0.835, column 1). Columns 6 and 7 show that this is because advice induces both high-performing and low-performing women to compete more, while the advice had no effect on men (c.f. columns 4 and 5). This leads to a net zero effect on efficiency, when defined in this way (c.f. column 1 in Panel 1). In Appendix E, we show that similar results are obtained when we look at the other pre-registered measure of efficiency (expected foregone earnings) instead, or use probit instead of a linear probability model.[20]

Table 5: The Effect of Advice on Efficiency

	(1) Payoff Max	(2) Payoff Max	(3) Payoff Max	(4) Payoff Max	(5) Payoff Max	(6) Payoff Max	(7) Payoff Max
				Panel 1: Laboratory Study			
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)

[19] High ability and low ability participants are participants who maximize their expected payment by choosing tournament and piece rate pay respectively based on their performance in part 2. Tournament pay is payoff-maximizing for participants with a win chance of at least 50% in the lab and at least 25% in the field experiment respectively.

[20] Our results hence differ from Brandts et al. (2015), who find that highly personalized advice does reduce the earnings loss 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 payment scheme that would maximize their expected earnings.

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)
				Panel 2: Field Study			
Advice	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
N (Lab Study)	374	185	189	105	80	100	89
N (Field Study)	268	117	151	36	81	59	92

Notes: Linear regression estimates; robust standard errors are in parentheses. 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. "Advice" is a binary variable for decisions made after receiving advice (lab study) or for participants in the treated group (field study). For high (low) performers, the constant indicates the share of participants choosing (not) to compete, which for them was the payoff-maximizing choice. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

By contrast, in the field study our intervention did significantly increase the share of payoff-maximizing choices by 11.6 percentage points (panel 2 in Table 5, the coefficient of 0.116 indicates that the share of people who chose the payoff maximizing payoff scheme increases by 0.116, i.e. with 11.6 percentage points, when treated with the advice). This corresponds to a reduction of payoff minimizing choices of around 40 percent. Columns 2 and 3 show that this increase was primarily driven by men; the effect for women is not significant. The remaining columns show that the reason that men are more positively affected is that high-ability men do not listen to advice: while their tournament entry rate falls from 85 to 75 percentage points, this effect is small and not significant (column 4). By contrast, low-performing men overwhelmingly heed the advice, reducing tournament entry from 41% to only 7%, thereby increasing the rate of payoff maximizing choices from 59% to 93% (column 5). When it comes to women, high-ability participants increase their tournament entry rates from 47% to 93%. Low-ability women also increase tournament entry from 7% to 30%, and as this group is larger than the group of high-ability women, the net effect on efficiency is zero for women, as evidenced in column 3, Panel 2. Here, too, the results are similar if we look at expected foregone earnings instead, see Online Appendix E.

As the next step, we look at the share of high performers who decide to compete. This definition may be of particular interest to firms, as it will be easier to fill high-level positions with good candidates if more high-performing candidates actually choose to apply for a job. In the laboratory study, advice creates a significant increase in the willingness of high-performing women to compete (10 percentage points, column 6), and no effect on men (column 4). Taken together, this implies that advice makes high performers 5.4 percentage points more likely to compete (p=0.001, t-test). This is an increase of 10.7 percent from a baseline of 50.2 percentage points.

In the field study, advice made high-performing women significantly more likely to compete (46 percentage points; column 6), while not significantly making high-performing men less likely to compete (column 4). Taking into account the greater number of women in our sample, these results together imply that high performers in our sample were, on average, 25 percentage points more likely to compete after receiving advice (p=0.008, test of proportions). This is an increase of 40 percent from a baseline of 62 percentage points.

Overall, the aggregate efficiency implications differ depending on the measure and the study. In the field study, the advice intervention increased efficiency along all the measures we considered in this section. In the laboratory study, the rate at which high-performers competed also increased, which may be particularly relevant for firms. However, in contrast to the field study, the rate at which participants chose the payoff-maximizing option did not increase due to advice. Overall, our information intervention hence seems most promising if the policy goal is to reduce the gender gap and increase tournament entry among high performers without lowering expected earnings overall.

However, when considering the ethical aspects of implementing interventions like the one studied here, it is also important to note that it had very different effects on different segments of the population. In particular, whereas our intervention increased the expected earnings of high-performing women and (in the field) low-performing men, it decreased the expected earnings of low-performing women. These women follow the advice to enter the tournament even though they would have earned more by sticking to the piece rate payment.

One way to reduce or eliminate the negative impact on low-performing women would be to better tailor advice to each individual's ability. However, the precise individual-level information that this requires may be difficult and costly to acquire in practical applications, or may not be available at all. In addition, individual-level assessment appears to be particularly prone to gender stereotypes, as demonstrated by Carlana (2019), Bohren et al. (2018) and Sarsons et al. (2020). Hence, there is likely to be a trade-off between less precise but cheaper and more easily implemented general advice, and more precise but more expensive and potentially more gender stereotyped individual advice.

A related issue pertains to the desirability of competitive preferences in society as a whole. Previous work tends to emphasize the positive aspects of being competitive (similar to the related trait of 'grit', as discussed by Duckworth, 2016, and experimentally investigated by e.g., Alan et al., 2019). This line of reasoning emphasizes that more competitive individuals may be more willing to 'compete' for jobs and apply to more ambitious ones, which will benefit their expected lifetime earnings. Indeed, Buser et al. (2014, 2017, 2020) find that competitive individuals choose more ambitious college majors with a greater associated expected lifetime earnings profile. In a similar fashion, by increasing the overall tendency to apply for competitive jobs, encouraging competitive behavior might help alleviate search frictions on the labor market.

At the same time, it should be noted that being competitive might not be the best choice for everyone. In our experiments, for example, participants may well have avoided the tournament because they realized they were unlikely to win, disliked competitions or were risk averse. More generally, encouraging everyone to compete could have long-lasting negative impacts on some individuals. For example, while successfully majoring in a STEM field or being a professional athlete is associated with greater earnings, not everyone may be able to succeed in these careers, or be willing to try.

Taken together, our results suggest that a simple advice intervention can reduce the gender gap in tournament entry, increasing the expected earnings of high-performing women and (in the field) low-performing men but reducing the earnings of low-performing women. The overall desirability of these results therefore depends on the relative weight given to the outcomes of 'winners' and 'losers' as well as the extent to which it seems desirable to have a competitive society where men and women make similar decisions.

5. Discussion and Conclusions

We investigate whether it is possible to reduce the gender gap in willingness to compete by using an information intervention that informs participants of the existence of the gap and advises them about its potential earnings implications. We find that it is indeed possible, as confirmed by the results from our laboratory study in Germany and our field study in Sweden. Further, comparing the effect size of our intervention shows that it is at least as effective at reducing this gender gap as other, more complex, costly and elaborate interventions studied in previous work. For example, among the 14 experiments we survey in Table A2, only the gender quota in Niederle, Segal and Vesterlund (2013) had a larger effect than the effect of advice in our field study. Likewise, even the effect in our laboratory results (which is smaller in size than the effect in the field) is greater than the effect of re-doing competitions with a non-female winner (Balafoutas & Sutter, 2012) or receiving highly personalized advice (Brandts, Groener and Rott, 2015).

Our intervention also positively affects the ex-post efficiency of outcomes, in the sense that it makes high performers more likely to enter, and win, tournaments, while holding steady (in the laboratory) or even increasing (in the field) the number of participants who choose the payoff-maximizing payment scheme. To follow the advice is, however, not beneficial for everyone, and indeed advice negatively affects the expected earnings of low performing women both in the laboratory and in the classroom. Hence, while simple advice could be cheap and easy to implement (relative to other policies), the negative effect on a substantial part of the population is a drawback that also needs to be taken into account when pondering the implementation of similar interventions outside the laboratory.

We designed our advice intervention to emphasize the "informational" component of advice while never explicitly "encouraging" participants to choose a particular action. In line with this, additional analysis in Appendix C shows that the gender gap in confidence was significantly reduced after advice (in the field) or smallest in the treatments that were most effective (in the lab). This suggests that advice may have induced participants to update their beliefs in the direction suggested by the informational component. However, it should be noted that we did not design our experiments to distinguish between the "informational" and "encouragement" components of advice. Instead, we implemented advice in a way that we felt would be representative of how advice is implemented in real world applications. For example, in our experience career advisers often highlight the benefits and costs of different career paths, without explicitly giving a career recommendation; in some applications (such as financial advice) giving a direct recommendation might even be illegal. Nevertheless, further distinguishing between the two advice components would be an interesting avenue for future research.

The role played by the informational and encouragement components may also depend on the stake size. Though the incentives in our laboratory study were sizeable by the standards of most laboratory experiments, they pale in comparison to many key decisions we make in the real world, such as choosing what college to attend or what career to choose. Intuitively, it seems plausible that greater incentives could increase the value of information while reducing the (relative) value of conforming to the advisor's suggestions.[21] If this is true, the importance of the informational component may be greater in key decisions outside the laboratory than in our experiments, and the importance of the "encouragement" component might be smaller. This implies that the net effect of advice could be both larger and smaller than the effects reported in this study, depending on which of these two effects dominates.

Although our information intervention worked to reduce 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 treatment effect was greater in the field study, and only in the field study did advice increase the expected earnings of the average participant. Likewise, while 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 affect men in the laboratory. That women react stronger to the advice than men is also in line with the findings reported in Croson and Gneezy (2009), Ellingsen et al., (2013), and de Quidt et al. (2018) where it is documented that women are more sensitive to changes in contextual factors in experiments.

Other reasons for why the advice effect was greater in the field could include sample differences (age and nationality in particular) and differences in the perceived authority of the

[21] Empirical evidence is mixed, however. For example, de Quidt et al. (2018) find that an increase in incentives only reduces the strength of demand effects (similar to the encouragement component) if these demand effects were "strong".

advice-giver. It may, for example, be that an experimenter is perceived as more of an authority by high school students – especially when their teacher is also present – than by university students, who are used to doing laboratory experiments. This could make both the informational and the encouragement component of advice more effective in the field setting.

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 may impact the results 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, if stake size matters, and if it makes a difference whether the advice is given repeatedly or only once.

Our results are promising from the perspective of a policy-maker interested in increasing the representation of women in high-level, prestigious and traditionally male professions. In many countries, this remains an important policy goal. Willingness to compete has been linked to the underrepresentation of women in such positions. Our results therefore suggest that simple advice may be a relatively easy way to decrease female underrepresentation. When thinking about implementations outside the laboratory it is, however, important to also keep in mind that relatively large amounts of data on which to base the advice needs to be available. While the large pre-existing literature on gender and competitiveness implies that this condition is fulfilled in the setting of our experiments, this may not always be the case. Likewise, an awareness of the fact that there will also be losers from the intervention – in our case the main example are low-performing women – is necessary. For example, it might not be regarded as ethical to implement an intervention where the number of losers is substantial, especially if the losers are already part of a vulnerable group (e.g. low-performers), even if the average person benefits.

One advantage of our experiment is that it is relatively easy to determine whether an individual participant should or should not compete in terms of maximizing their expected earnings. Outside of controlled experimental settings, however, it may not always be easy to determine the ‘right’ advice, since this may depend on factors that are not perfectly observable to the advisors, including aptitude and ambition. Careful research is hence needed to determine the ‘right’ advice for different groups of people. In addition, contextual factors, such as the relative stakes of decisions in various contexts and how that potentially interacts with the effect of advice, need to be investigated. It is also worth noting that the effectiveness of our intervention may at least in part be explained by its close temporal proximity to the final decision. Whether this is indeed the case would be interesting to study in future work, and if true, it may imply that simple advice should ideally be administered close in time to the relevant decision.

Keeping these caveats in mind, a successful implementation of our policy would likely involve several steps. First, sufficient data would need to be obtained to give the “right” advice to a sufficiently large proportion of the target population. Second, the data would have to be fitted into an easily digestible format (e.g., a few short sentences, or an intuitive graph). Third, this information would have to be relayed to third parties that can reach the target population (such as teachers or school counselors for high school students). However, the exact implementation would likely depend on many factors that may be specific to a given context. Comparing different ways of implementing our intervention in practice could be an interesting and important topic for future research.

Our results may also have implications beyond competitiveness as it could be possible to give similar advice in related contexts as well. For example, just as women are underrepresented in competitions, they are also underrepresented in STEM-fields. One could 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 such an intervention would be another intriguing topic for future research.

References

Alan, S, T Boneva, S Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *Quarterly Journal of Economics*, 134(3): 1121-1162.

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) 187: 108878.

Apicella, C L, E E Demiral, and J Mollerstrom. 2020. "Compete with Others? No Thanks. With Myself? Yes, Please!" *Economics Letters* 187, 108878.

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, H Fornwagner, and M Sutter. 2018. Closing the gender gap in competitiveness through priming. *Nature Communications* 9**:** 4359.

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

Baldiga, N R, and Coffman, K B (2018). Laboratory Evidence on the Effects of Sponsorship on the Competitive Preferences of Men and Women. *Management Science*, 64(2), 888–901.

Bartling, B, E. Fehr, and D Schunk (2012). Health Effects on Children's Willingness to Compete. *Experimental Economics*, 15(1): 58-70.

Berlin, N, & Dargnies, M-P (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, 130, 320–336.

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

Bonaccio, S., & Dalal, R. S. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151.

Boschini, A, A Dreber, E von Essen, A Muren, and E Ranehill (2019) Gender, Risk Preferences and Willingness to Compete in a Random Sample of the Swedish Population. *Journal of Behavioral and Experimental Economics*, 83: 101467.

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, M Niederle, and H Oosterbeek. 2020. "Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures." *Tinbergen Institute Discussion Paper*, 2020-048.

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. 2021. "Gender Differences in Willingness to Compete: The Role of Public Observability." *Journal of Economics Psychology*, 93: 102366.

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.

Demiral, E E and Mollerstrom, J. 2020. Competitiveness and Employability. Working Paper, George Mason University, mimeo.

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

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

Duckworth, A. 2016. *Grit: The Power of Passion and Perseverance.* New York, NY: Scribner/Simon & Schuster.

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

Ellingsen, T., M. Johannesson, J. Mollerstrom and S. Munkhammar. 2013. Gender Differences in Social Framing Effects. *Economics Letters*, 118: 470-472.

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.

Gneezy, U, K L Leonard, J A List (2009). Gender Difference in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5): 1637-1664.

Gneezy, U., Saccardo, S., Serra-Garcia, M., & van Veldhuizen, R. (2020). Bribing the Self. *Games and Economic Behavior*, 120: 311-324.

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.

Iriberri, N., & Rey-Biel, P. 2017. Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*, *135*, 99–111.

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

Mollerstrom, J., and C. Sunstein. 2020. "How Special is Democracy? An Experimental Investigation of Recommendations in the Minimum Effort Game." *Economics Letters*, 190: 109066

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.

Porter, C., & Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied* (forthcoming).

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

Samek, A C (2013) Is there a Gender Gap in Preschoolers' Competitiveness? An Experiment in the US. *Journal of Economics Behavior and Organization*, 92: 22-31.

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

Schram, A., Brandts, J., & Gërxhani, K. 2019. Social-status ranking: a hidden channel to gender inequality under competition. *Experimental Economics*, *22*(2), 396–418.

Shurchkov, O. 2012. Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints. *Journal of the European Economic Association*, *10*(5), 1189–1213.

Sniezek, J. A., & Buckley, T. 1995. Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes*, 62(2), 159–174.

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.

Zizzo, D. J. 2010. Experimenter Demand Effects in Economic Experiments. *Experimental Economics*, 13(1), 75–98.

Online Appendix A: Pre-Analysis Plan and Power Calculations

Note: the pre-analysis and power calculations presented here are re-printed from our pre-analysis plan, which is available at https://www.socialscienceregistry.org/trials/3487.

A1: Pre-Analysis Plan for the Laboratory Study

For "Can simple advice eliminate the gender gap in willingness to compete?"

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.

A2: Power Calculations for the Laboratory Study

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 chose the tournament in an earlier study run at the same laboratory (Van Veldhuizen, 2018) as the benchmark for willingness to compete (men=58.6 percent, women=27.1 percent). 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 percent and women=37.1 percent 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 percent for the example of a 20pp decrease in the gender gap) compete in both cases; some women (10 percent in the example) compete only after receiving advice; and the remainder (62.9 percent 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 percent 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 percent as an upper bound for the expected noise in the sample, and compute power for a sample with 11 percent, 16 percent or 21 percent of switchers in each condition.

Table A1: Power Calculations

	Switchers: 21 percent		Switchers: 16 percent		Switchers: 11 percent	
	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 and sample sizes from 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 (which aggregates the data from the three treatments) has a power of 0.838 to detect an effect that is 54 percent as large as the average effect in Table A2 if we conservatively assume that 21 percent 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 percent 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., 2021)

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 14 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, i.e., the fraction of men and the fraction of women competing in the treatment condition, minus the corresponding fraction in the control condition. "ADM2017" is Apicella, 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 teammate; 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[22]

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=?

[22] These are our original English instructions; the German translations that were used in running the experiment can be found in a separate file.

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.[23] For

[23] Here “X” is equal to the participant’s actual performance in part of the experiment.

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.[24]

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:

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

- With 50 percent probability the investment fails and you lose the amount invested.
- With 50 percent 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 percent 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 percent 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 most people that you know?

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, rounding up to the nearest 50 cents.

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 percent chance, lottery 2 pays out with 95 percent chance, and so on. Lottery 21 pays out 2 Euros with 0 percent 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 percent chance of paying out the 2 Euros. For individuals ranked 9th or better, performance pay will result in at least a 60 percent 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 percent chance of obtaining 2 Euros in performance pay. By contrast, individuals ranked 11 or worse have at most a 50 percent 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[25]

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.

---

[25] 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 percent 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.

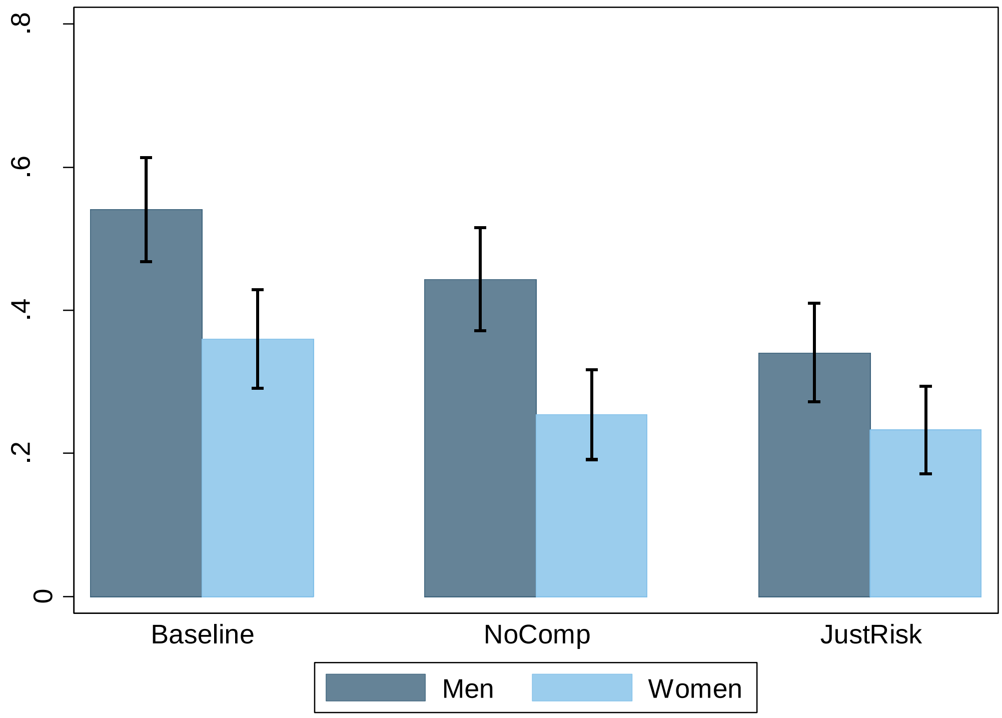
Lastly, before we randomly select who will be receiving compensation, we would like to record your legal gender as it is recorded in your passport. A sheet is now being distributed where you can record this information. Raise your hand when you have done so, and we will come and collect it.

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, 2019).

The first measure, NoComp, removes the role of competitiveness. Consider a participant with a performance of 10 who expects to have a 75 percent 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 percent 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 a 75 percent chance). The key difference is that the relevant item in Part 4 is no longer competitive, as 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 the 75 percent-lottery decision as 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, but 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). Note that we can only implement this approach on the data from the laboratory study.

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



Notes: The figure displays the fraction of participants choosing the tournament (baseline) or lottery (NoComp and JustRisk) for men and women respectively. The error bars represent 95 percent confidence intervals.

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 percent (10.7/18.1, p=0.021, t-test) of the gender gap in willingness to compete. Confidence appears to explain 45 percent (8.2/18.1, p=0.116, t-test), whereas the point estimate for competitiveness is negative and not significant at -4 percent (-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.

Another way to examine the role of risk attitudes and confidence is to look directly at our elicited measures for these variables. Table A3 presents evidence suggesting that, relative to the Competitiveness treatment, the gender gap in risk aversion (columns 1 and 3) and confidence (column 2) is smaller in the Confidence and Risk treatments, though the effect for risk attitudes is only significant using the part 6 investment measure. Keeping in mind that not all coefficients are significant at conventional levels, this analysis tentatively suggests that these treatments may have been more effective at reducing the gender gap in risk attitudes and confidence, and that this might explain their greater effectiveness (relative to the Competitiveness treatment). Columns 4 and 5 present a similar analysis for the field experiment, showing that the treatment significantly reduced the gender gap in confidence while having no effect on risk attitudes. Taken together, these results tentatively suggest that our intervention may have changed behavior, in part, by reducing the gender gap in self-confidence; the results for risk attitudes are less robust.[26]

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

	(1)	(2)	(3)	(4)	(5)
	Part 4 Risk	Part 5 Rank	Part 6 Investment	Field Confidence	Field Risk
Female	-1.265*	3.716***	-2.608***	-0.282+	-0.377*
	(0.555)	(0.660)	(0.520)	(0.165)	(0.150)
Treatment Risk	0.412	1.596*	-1.056*		
	(0.600)	(0.689)	(0.532)		
Female*Treatment Risk	0.436	-1.466	1.732*		
	(0.885)	(0.991)	(0.772)		
Treatment Confidence	-0.069	1.550*	-0.933+		
	(0.563)	(0.695)	(0.524)		
Female*Treatment Confidence	-0.174	-2.302*	1.453+		
	(0.813)	(0.938)	(0.758)		
Treatment				-0.060	0.007
				(0.171)	(0.163)
Female*Treatment				0.474*	0.076
				(0.241)	(0.203)
Constant	6.965***	6.912***	8.351***	2.525***	2.390***
	(0.379)	(0.488)	(0.344)	(0.122)	(0.121)
Study	Laboratory	Laboratory	Laboratory	Field	Field
Observations	374	374	374	268	268
R-squared	0.039	0.111	0.074	0.025	0.044

Notes: Linear regression estimates; robust standard errors are in parentheses. The dependent variables are the number of risky choices in Part 4 (column 1), the reported rank in Part 5 (column 2), the amount invested in Part 6 (column 3), the elicited beliefs in the field (column 4) and the elicited risk preferences in the field (column 5). The independent variables are a gender dummy (0-men, 1-women), three dummies specifying the treatments and the interaction effect between each treatment variable and gender. The first three columns contain data from the laboratory study, the last two report data from the field study. The Competitiveness treatment serves as the baseline treatment in columns (1)-(3). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

[26] Two additional caveats apply to the data from the laboratory experiment. First, the laboratory experiment does not have a baseline (untreated) measure of confidence and risk attitudes, forcing us to use the Competitiveness treatment as the baseline. Second, the point estimates in Table A3 tell us that most of the effect is driven by men. This is not consistent with the behavioral results reported in Table 2, which were driven exclusively by women.

Online Appendix D: Additional Procedures

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 percent to 0 percent 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 with 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 1).

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 (6.1 percent). The relatively low number affected does not appear to have been driven by a failure to read the advice: 79 percent 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 targeted by advice (women planning to choose piece rate and men planning to choose tournament), only 52 percent (47 percent of men, 55 percent 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 percent 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

Tables A4 and A5 reprint the main results from the two experiments (Tables 2 and 4) using an ordered probit model (laboratory experiment) and a probit model (field experiment) respectively. The p-values are similar to Tables 2 and 4; the main difference is that the difference-in-difference estimate in the laboratory experiment is larger in the Risk treatment in this specification. Tables A6 and A7 show that we also obtain similar p-values using t-tests, Wilcoxon tests and, when applicable, tests of proportions and Fisher Exact tests as well. Finally, Table A8 reprints the efficiency results from Table 5 using a probit model; once again the p-values are almost identical. Here, too, we obtain similar results using t-tests or tests of proportions; these results are available upon request.

Table A4: Effect of Advice in the Laboratory Study using Ordered Probit

	(1) Advice Effect (All)	(2) Advice Effect ("Competitiveness")	(3) Advice Effect ("Risk")	(4) Advice Effect ("Confidence")
Female	1.041***	0.543	4.683***	1.024*
	(0.290)	(0.480)	(0.204)	(0.465)
Score Controls	Yes	Yes	Yes	Yes
Observations	374	127	123	124
Pseudo R-squared	0.110	0.083	0.181	0.116

Notes: Ordered Probit Estimates; 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. Since ordered probit regressions do not include a conventional constant, only the difference-in-difference result is reported. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table A5: Effect of Advice in the Field Study using Probit

	(1) Compete	(2) Compete	(3) Compete
Advice	0.121	-1.112***	1.147***
	(0.174)	(0.290)	(0.266)
Constant	-0.366**	0.298	-1.002***
	(0.130)	(0.198)	(0.207)
Score controls	Yes	Yes	Yes
Subgroup	All	Men	Women
Observations	268	117	151
R-squared	0.342	0.347	0.299

Notes: Probit estimates; 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.

Table A6: The Effect of Advice in the Laboratory Study with Different Tests

	(1) Advice Effect (All)	(2) Advice Effect ("Competitiveness")	(3) Advice Effect ("Risk")	(4) Advice Effect ("Confidence")
		Men		
Table 2	0.494	0.223	0.356	0.441
T-test	0.565	0.322	0.321	0.321
Wilcoxon	0.564	0.317	0.317	0.317
		Women		
Table 2	<0.001	0.024	0.003	0.006
T-test	<0.001	0.024	0.004	0.007
Wilcoxon	<0.001	0.025	0.005	0.008
		Gender Difference		
Table 2	<0.001	0.198	0.002	0.022
T-test	0.001	0.157	0.002	0.015
Wilcoxon	0.001	0.156	0.002	0.015
Fischer Exact	<0.001	0.159	0.006	0.016
Ordered Probit	<0.001	0.258	<0.001	0.028
Observations	374	127	123	124

Notes: This table presents the p-values for the advice effect in the laboratory study using different tests. The first panel presents the p-values for men, the second for women and the third for the gender difference. Within each panel, the first row reprints the results of Table 2. Row 2 uses t-tests and row 3 uses Wilcoxon tests instead. For the gender difference in analysis we also include Fisher's Exact test and the results of the ordered probit analysis presented in Table A4.

Table A7: The Effect of Advice in the Field Study with Different Tests

	(1) Overall	(2) Men	(3) Women	(4) DiD
Table 4	0.520	<0.001	<0.001	<0.001
T-test	0.586	<0.001	<0.001	<0.001
Wilcoxon	0.586	0.001	<0.001	
Test of Proportions	0.585	<0.001	<0.001	
Fisher Exact	0.618	0.001	<0.001	
Probit	0.487	<0.001	<0.001	0.001
Subgroup	All	Men	Women	All
Observations	268	117	151	268

Notes: This table presents the p-values for the gender difference in the advice effect in the field study using different tests. The first column presents the pooled advice effect, columns 2 and 3 present the results for men and women respectively, and column 4 presents the difference-in-difference test. The first row reprints the results of Table 4. Row 2 uses t-tests, row 3 uses Wilcoxon tests, row 4 uses a test of proportions and row 5 uses a Fisher Exact test instead. The final row reprints the results of Table A5. For the difference-in-difference analysis, we use only tests that admit testing for a difference-in-difference; for the second row the relevant test is the gender*treatment interaction term in a linear regression that does not control for performance.

Table A8: The Effect of Advice on Efficiency using Probit

	(1) Payoff Max	(2) Payoff Max	(3) Payoff Max	(4) Payoff Max	(5) Payoff Max	(6) Payoff Max	(7) Payoff Max
				Panel 1: Laboratory Study			
Advice	0.007	0.015	-0.000	0.027	0.000	0.266***	-0.369***
	(0.033)	(0.026)	(0.060)	(0.027)	(0.048)	(0.080)	(0.110)
Constant	0.278***	0.426***	0.140	0.457***	0.385**	-0.468***	0.916***
	(0.066)	(0.096)	(0.092)	(0.128)	(0.145)	(0.131)	(0.156)
				Panel 2: Field Study			
Advice	0.385*	0.201**	0.160	-0.362	1.238***	1.525***	-0.848**
	(0.172)	(0.075)	(0.224)	(0.490)	(0.357)	(0.426)	(0.328)
Constant	0.548***	0.678***	0.615***	1.036**	0.227	-0.078	1.360***
	(0.113)	(0.061)	(0.153)	(0.347)	(0.204)	(0.224)	(0.264)
Gender	All	Men	Women	Men	Men	Women	Women
Performance	All	All	All	High	Low	High	Low
N (Lab Study)	374	185	189	105	80	100	89
N (Field Study)	268	117	151	36	81	59	92

Notes: Probit coefficient estimates; robust standard errors are in parentheses. 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. "Advice" is a binary variable for decisions made after receiving advice (lab study) or for participants in the treated group (field study). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

E2. An 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, as 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 A9 shows the results for this alternative measure. For the laboratory study (panel 1) the results are very similar to Table 5: high-performing women gain from receiving advice, whereas low-performing women lose, leading to a net zero effect overall.

For the field study (panel 2), the results are qualitatively similar to the results reported in the main text, 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 currencies being euros in the laboratory, and Swedish krona (10 SEK ≈ 1 EUR at the time of writing this paper) in the field.

Table A9: The Effect of Advice on Expected Foregone Earnings

	(1) Foregone Earnings	(2) Foregone Earnings	(3) Foregone Earnings	(4) Foregone Earnings	(5) Foregone Earnings	(6) Foregone Earnings	(7) Foregone Earnings
				Panel 1: Laboratory Study			
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)
				Panel 2: Field Study			
Advice	-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
N (Lab Study)	374	185	189	105	80	100	89
N (Field Study)	268	117	151	36	81	59	92

Notes: Linear regression results; robust standard errors are in parentheses. The dependent variable is the participant's expected foregone earnings. "Advice" is a binary variable for decisions made after receiving advice (lab study) or for participants in the treated group (field study. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

E3. Permutation tests

In the field study, treatment assignments are implemented on the classroom 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 the results reported in the main article. 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 analysis: the treatment effect on the probability that an individual competes/makes the payoff-maximizing choice.

Given randomization of treatment, strict exogeneity holds. 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. In a setting where treatment is random, randomization inference does however provide 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.[27]

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.[28] For each perturbation the regression model 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.[29] This is implemented in Stata using 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 A10 reprints the results of Table 4 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.

[27] See Young (2017) for a detailed description of the concept.
[28] 200 perturbations have been shown to be sufficient for the p-value to stabilize (Young, 2017).
[29] 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 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.

Table A10: 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***
	(0.0154)	(0.003)	(0.001)
Constant	0.380	0.562	0.245
Subgroup	All	Men	Women
Observations	12	12	12
R-squared	0.0886	0.8317	0.8617

Notes: Linear regression analysis; 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.

Discussion Papers of the Research Area Markets and Choice 2021

Research Unit: **Market Behavior**

Jeanne Hagenbach and Frédéric Koessler — SP II 2021-201
Selective memory of a psychological agent

Dany Kessel, Johanna Mollerstrom, and Roel van Veldhuizen — SP II 2021-202
Can simple advice eliminate the gender gap in willingness to compete?

Research Unit: **Economics of Change**

Maja Adena, Ruben Enikolopov, Maria Petrova, and Hans-Joachim Voth — SP II 2021-301
Bombs, broadcasts and resistance: Allied intervention and domestic opposition to the Nazi regime during World War II

Marta Serra-Garcia and Nora Szech — SP II 2021-302
Choice architecture and incentives increase COVID-19 vaccine intentions and test demand

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice