The Impact of Advice on Women's and Men's Selection into Competition

Jordi Brandts
Valeska Groenert
Christina Rott

September 2012

Barcelona GSE Working Paper Series

Working Paper nº 663

The impact of advice on women's and men's selection into competition[1]

Jordi Brandts[2]
Universitat Autònoma de Barcelona,
Institut d'Anàlisi Econòmica (CSIC),
and Barcelona GSE

Valeska Groenert[3]
Universitat Autònoma de Barcelona[4]

Christina Rott[5]
Universitat Autònoma de Barcelona

September 2012

ABSTRACT: We conduct a laboratory experiment to study how advice affects the gender gap in the entry into a real-effort tournament. Our experiment is motivated by the concerns raised by approaching the gender gap through affirmative action. Advice is given by subjects who have already had some experience with the participation decision. We show that advice improves the entry decision of subjects, in that forgone earnings due to wrong entry decisions go significantly down. This is mainly driven by significantly increased entry of strong performing women, who also become significantly more confident, and reduced entry of weak performing men.

KEYWORDS: experiments, advice, gender gap in competitiveness
JEL CLASSIFICATION NUMBERS: C91; J08; J16

---

[1]The authors acknowledge the financial support of the Antoni Serra Ramoneda Research Chair (UAB - Catalunya Caixa), the Spanish Ministerio de Economía y Competitividad (Grant: ECO2011-29847-C02-01) and the Generalitat de Catalunya (Grant: 2009 SGR 820) for their financial support. Valeska Groenert gratefully acknowledges support from the Spanish Ministry of Science and Innovation through grant "Consolidated Group-C" ECO2008-04756 and FEDER.

[2]jordi.brandts@uab.cat; Universitat Autònoma de Barcelona, Institut d'Anàlisi Econòmica (CSIC), Barcelona GSE; 08193 Bellaterra (Spain)

[3]valeska.groenert@uab.cat; MOVE (Markets, Organizations and Votes in Economics), UAB; 08193 Bellaterra (Spain); Tel: +34 93 581 4653.

[4]European Commission, DG for Economic and Financial Affairs, as of September 1, 2012.

[5]christina.rott@uab.cat; Dep. of Business Economics, UAB; 08193 Bellaterra (Spain)

> "[...] If you earned your degree in areas where we need more women – like computer science or engineering – reach back and persuade another student to study it, too. [...] Until a girl can imagine herself, can picture herself as a computer programmer, or a combatant commander, she won't become one. [...]."
>
> Barack Obama at the Barnard College Commencement on May 14 of 2012

1 Introduction

The underrepresentation of high performing women in top level jobs continues to be a major concern of politicians and society in general and has received much attention by researchers in economics and other fields. For some time research on this topic focused on explanations related to differences in human capital, discrimination, and child-rearing. More recently, the literature has explored an additional explanation: gender differences in entry into competition as studied in the seminal paper by Niederle and Vesterlund (2007; NV07 henceforth). Participants in their experiment perform a simple addition task, for which they are given the choice between a competitive compensation scheme (in which payoffs depend on relative performance) and a non-competitive one (in which payoffs depend only on own performance). They show that women opt for the competitive scheme far less than men do though there are no significant gender differences in performance. In other words, there exists a gender gap in entry into competition.

Subsequently, a number of papers have explored designs to reduce this gender gap. One set of papers considers the effects of affirmative action, such as quotas or other forms of positive discrimination (Niederle, Segal, and Vesterlund, 2010; Balafoutas and Sutter, 2012). It is argued that these measures are effective and do not substantially harm efficiency. However, affirmative action policies are highly controversial and it remains difficult to find majorities for them. Despite the results in the aforementioned studies, there remains a concern that they do not lead to efficient allocations, possibly promoting weak performing women at the cost of high performing men. Another concern is that they actually do not serve women because women who obtain a job under an affirmative action scheme may be stigmatized of being selected only because of this action.

The idea behind our experiment is inspired by the literature on naive advice and by mentoring programs that often exist in real-world settings.[6] The advice literature

[6]Two examples of mentoring programs are the one organized by the organization Women in

consistently shows that advice improves decisions. Our aim is to study whether receiving advice from a person who has experienced the same particular competitive situation and has some informational advantage can improve competitive choices and therefore constitute an alternative to affirmative action. In addition, our design allows us to study the effects of receiving advice from a person of the same gender.

We conduct two different advice treatments. In the first treatment, advisors and advisees are randomly matched irrespective of their gender and subjects do not learn the gender of their matched partner. In the second treatment, we match female advisors with female advisees and male advisors with male advisees and inform subjects about the gender of their matched partner. We conducted this second advice treatment because it reflects how in reality mentoring programs are often designed (where, e.g., senior women mentor younger women). Therefore we wanted to check whether there are any differential effects if subjects give/receive advice to/from the same gender.

For purposes of maximal comparability our basic experimental design closely follows the design in NV07. Participants face a real effort task, which consists of adding up series of five two-digit numbers. They have five minutes to solve as many problems as possible. Subjects complete this task once under a piece rate scheme and once under a winner-take-all tournament scheme (for the tournament they are matched in groups of four). Then, before performing the addition task a third time, they are given the choice between being paid under the piece rate or the tournament scheme. To help with this decision, subjects in our advice treatment receive a message from an advisor, who can either recommend choosing the "piece rate" or the "tournament." Advisors are subjects who had to make the same decision some time before and who received information about the performance levels in the first two rounds of their own group and of their advisee. We chose this design because it reflects in the most natural way how advice is given in the real-world: in most situations, both, previous experience with the same decision as well as information about the own generation's and the advisee's performance influence an advisor's recommendation.[7]

Our results show that advice indeed improves the selection into competition

Technology and the one by CSWEP - the Committee on the Status of Women in the Economics Profession (see Blau, Currie, Croson, and Ginther, 2010, for an analysis of the effects of the mentoring program run by CSWEP).

[7]Wozniak, Harbaugh, and Mayr (2011) and Ewers (2012) study the effects of information feedback of different types. We think that receiving information through another person's advice is a natural way to approach the issue.

by various measures. Without advice, women who enter the competition do not perform significantly better than those who do not. With advice those who enter are the ones with significantly stronger performance. For men, receiving advice also leads to an improvement in the selection into those who enter and those who do not, but the effect is not as strong as for the women. Following NV07, we also calculate the opportunity cost of taking the 'wrong' entry decision. There are two types of 'mistakes,' weak performing subjects (whose expected payoff would be higher under the piece rate) who enter the competition and strong performing subjects (whose expected payoff would be higher in the tournament) who do not. These forgone earnings from the two types of mistakes are significantly lower in our two treatments with advice than in the control group without advice. Examining who are the subjects that improve their entry decisions under advice, we find that it is in particular the strong performing women who enter significantly more (an increase from about 35% to over 80%). There is also a substantial reduction of entry among the weak performing men in the treatment without gender matching (from over 40% to about 20%). We are also able to shed some light on the channels through which advice affects entry decisions and find that it changes confidence levels and leads subjects to think longer before making a decision.

Analyzing how advice is given, we find that advice tends to be 'correct' in the sense that nearly all those who, from an expected payoff maximizing point of view, should enter receive advice to do so. Conversely, most subjects that should not enter receive advice not to do so. It is also true that women are more hesitant in giving the advice to enter. Compared to men, they require higher performance by their advisees in the previous tasks before recommending entry. This gender difference in advice giving disappears if we control for the advisors' own entry decisions. This leads us to conclude that (some of) the factors that explain the gender gap in entry also explain the gender differences in advice giving.

The rest of the paper is organized as follows. Section 2 provides an overview of the literature, Section 3 describes the experimental design and sets out our main research questions, Section 4 contains all results, and Section 5 concludes.

2 Literature Review

2.1 Competitiveness and gender differences

One of the aspects, the literature on gender differences in competitive behavior (see Croson and Gneezy, 2009, for a survey) has focused on is the tendency to select into competition. The gender gap in entry into competition found by NV07 seems to be quite robust as the results in a number of papers with similar designs show (see for example Dargnies, 2011; Balafoutas and Sutter, 2012; Cason, Masters, and Sheremeta, 2010; Niederle, Segal, and Vesterlund, 2010; Sutter and Rützler, 2010; Wozniak, Harbaugh, and Mayr, 2010). In addition, similar gender gaps were found under a variety of different designs (see for example Gupta, Poulsen, and Villeval, 2011; Gneezy, Leonard, and List, 2009; Dohmen and Falk, 2011) and in the field (Flory, Leibbrandt, and List, 2010).

One important reason for the gender gap in entry into competition seems to be that men are relatively more overconfident (see review of Croson and Gneezy, 2009), in particular in a tournament environment (Charness, Rustichini, and van de Ven, 2011; Reuben, Rey-Biel, Sapienza, and Zingales, 2012). Not only on the supply side of the labor market, beliefs about one's performance are important, but also in hiring decisions, women are discriminated because of biased beliefs about their abilities (Reuben, Sapienza, and Zingales, 2010). NV07 underline that, in addition, preferences for competing play an important role, but other studies do not find significant gender differences in the willingness to compete when controlling for other factors such as distributional preferences, risk attitudes and past performance (Balafoutas, Kerschbamer, and Sutter, 2011) or for confidence levels (Charness, Rustichini, and van de Ven, 2011).

A number of papers have looked at designs to mitigate or even overcome the gender gap in entry. Niederle, Segal, and Vesterlund (2010) show that affirmative action in form of a quota increases female participation in a tournament. Balafoutas and Sutter (2012) also confirm that quotas and other forms of positive discrimination (preferential treatment and repetition of competition if a man wins) encourage women to enter competitions. Both of the aforementioned studies argue that this is achieved with at most a modest efficiency loss and driven by the increased entry of high performance women. Wozniak, Harbaugh, and Mayr (2010) show that providing relative feedback also can eliminate the gender gap. Finally, Dargnies (2011) shows that the gender gap disappears if participants compete as part of a team. This is a consequence of men choosing to compete significantly less often when they have to compete in a team rather than alone.

2.2 Advice and mentoring

The main conclusions from the experimental economics literature on decision-making with naive advice (in the sense of uninformed word-of-mouth advice as compared to advice from experts) are summarized in the survey by Schotter (2003). The experimental design employed in the surveyed papers is usually one where participants are split into two non-overlapping 'generations.' Participants of the first generation experience the decision-making situation once and then become advisors to participants of the second generation, who face the same situation. Schotter (2003) highlights five results: (1) Subjects tend to follow advice; (2) advice tends to change behavior; (3) decisions with advice are closer to theoretic predictions; (4) if subjects can choose between receiving advice or the information on which the advice is based, they opt for the advice; (5) advice improves decisions because it forces advisees to think about the problem.[8]

In a similar vein, several field experiments study the effect of mentoring and coaching on performance, success, and behavior. In particular, Blau, Currie, Croson, and Ginther (2010) suggest that mentoring programs for female assistant professors lead to a significant increase in performance in terms of publications and grants obtainment.[9]

Finally, advice has been extensively studied in the organizational psychology literature. Bonaccio and Dalal (2006) provide a review of that literature. In the typical set-up, the decision-maker is asked for a tentative decision on the problem and the advisor makes a recommendation to the decision-maker, who then gets the opportunity to revise his original choice. The main results that pertain to our experiment are that (1) advice improves the accuracy of final decisions (in many set-ups there is a correct answer to the problem such as estimating the year of a specific event in US history, Gino, 2008), and (2) advice is discounted, that is it is not fully utilized. The following factors may influence advice utilization and the accuracy of final decisions: (a) Whether the decision-maker is asked to form an initial opinion (in particular, if people display a confirmation bias, this may make

[8]These conclusions are drawn from the results in Celen, Kariv, and Schotter (2010), Iyengar and Schotter (2008), Schotter and Sopher (2003, 2007), and Chaudhuri, Schotter, and Sopher (2009).

[9]See also Bettinger and Baker, 2011, and Rodríguez-Planas, 2010, for some more loosely related field experiments.

a difference); (b) whether the decision-maker is given the option to solicit advice; (c) the number of advisors; (d) the type of decision, e.g. whether the problem has a correct answer or it is rather a choice or a judgement problem; (e) the amount and type of interaction between advisor and decision-maker. Factors that reduce advice discounting are the advisor's expertise and amount of information she holds, quality of advice, making advice costly, increasing task complexity, and congruence of the goals of the advisor and the decision-maker.

3 Experimental design and research questions

We will first describe the basic experimental design with regard to the choice of participating in a competition and then turn to the design of the advice part of the experiment.

3.1 The basic set-up

For the purpose of maximal comparability, we keep the experimental design regarding the participation decision as close as possible to the one in NV07, which involves four performance tasks plus a self-evaluation task. The real effort task consists in participants having five minutes to add up sets of five two-digit numbers without using a calculator (see screenshot provided with the instructions in the appendix). In order for subjects to familiarize themselves with the two payment schemes, they first perform the task under the non-competitive payment scheme (task 1 - piece-rate), then under the competitive payment scheme (task 2 - tournament), and finally, before performing a third time, they have to choose between the competitive and the non-competitive payment scheme (task 3).[10] In addition, in task 4 subjects have to decide whether to apply the competitive or the non-competitive payment scheme to their (past) task 1 performance.[11] Finally, subjects have to estimate their relative performances in tasks 1 and 2.

Under the piece-rate payment scheme, subjects receive €0.5 for each correct sum. For the competitive payment scheme, subjects are matched in groups of four and only the person with the best performance receives payment in form of €2 for each correct sum. If a subject chooses the competitive payment scheme in task 3,

[10]Tasks 1 and 2 are also useful for us to assess a subject's ability level in the addition task.

[11]We will use the task 4 entry decision as a proxy for risk preferences. We will explain this more fully in the results section.

her task 3 performance is evaluated against the task 2 performance of her group members. Thus the subject 'wins' the tournament in task 2 and task 3 if she solves more problems correctly than each of her group members in task 2, with ties being broken randomly among the best performers.[12]

3.2 Advice

Upon arrival to the experiment, subjects are divided evenly into two different rooms (separated by a glass window). Subjects in one room have the role of the advisors and subjects in the other room the role of the advisees (but they do not learn about their roles until the advice stage begins). The advice stage for advisors follows after they have completed tasks 1-4 and the self-evaluation task, while the advice stage for advisees follows after they have completed tasks 1 and 2, that is immediately before they have to choose the payment scheme for task 3, see table 1. The advice consists of a recommendation as to the choice of the payment scheme for task 3 and (possibly) some reasons for the recommendation. Each advisee is randomly matched to exactly one advisor and each advisor has only one advisee. An advisor is paid 50% of her advisee's task 3 earnings.[13]

The exact sequencing of the advice-stage is as follows. The advisee sends information about his task 1 and 2 performances to his advisor. Upon receiving this information, the advisor sends a message, telling her advisee whether or not she recommends entering the competition. The advisor is then asked to give her advisee a reason for her recommendation. We provide three pre-formulated reasons for each recommendation ('tournament' or 'piece rate') from which the advisor can select as many as she wishes to.[14] After having received the advisor's recommendation and (possibly) some reasons for this recommendation, the advisee decides whether to enter the competition in task 3. The exact statement of each reason is provided in our discussion of how reasons were selected in subsection 4.4.

[12] As is argued in NV07, the fact that subjects in task 3 compete with the performance of subjects in task 2 ensures that a subject's entry decision is not influenced by beliefs about the other subject's entry decisions and therefore allows a cleaner analysis (though this feature of the experiment is admittedly not very realistic).

[13] On average €3.98 was earned for the advice, with payoffs ranging from €0 up to €27. Thus incentives to give 'good' advice were substantial.

[14] Ideally, we would have liked the interaction between advisor and advisee to be less structured, with them being able to exchange chats. However, since our experiments were conducted in Spanish we were concerned that the use of adjectives could reveal a subject's gender to their matched partner, which could have led to effects stemming from gender-pairing.

Table 1: Timeline of tasks and compensation schemes in the experiment

Generation 1 (advisors)	Generation 2 (advisees)
Task 1: 5-minute addition task - piece rate (€ 0.5)	
Task 2: 5-minute addition task - tournament (€ 2, winner takes it all)	
Task 3: 5-minute addition task - selection of compensation scheme	
Task 4: Selection of compensation scheme task 1	Task 1: 5-minute addition task - piece rate (€ 0.5)
Self evaluation task 1 and 2 (€ 1 per correct guess)	Task 2: 5-minute addition task - tournament (€ 2, winner takes it all)
Performance feedback own group	
Recieve advisee's performance info	**Send own performance info**
Give advice (50% of the advisee's task 3 earnings)	**Receive advice**
Choose up to three reasons (preference for competition, confidence, risk preferences)	**Receive up to three reasons (preference for competition, confidence, risk preferences)**
	Task 3: 5-minute addition task - selection of compensation scheme
	Task 4: Selection of compensation scheme task 1
	Self evaluation task 1 and 2 (€ 1 per correct guess)
	Performance feedback own group

Advisors in our experiment are no experts in the task, but they have experienced the situation once and hold some informational advantage. At the end of the self-evaluation task, each participant receives feedback on the tasks 1 and 2 performances of all her group members. Therefore, advisors have not only made one entry decision, but also have seen how people perform in the addition task in a small sample of four people. Note, however, that when the advisors receive the information about performance levels in their group they do not yet know about the advice stage. We chose this design because we felt that this was the most natural set-up. Usually, a person who has previously participated in a competition task will have some idea about performance levels in that task, but it is not so clear whether she will remember correctly. Finally, an advisor knows that her advisee has just completed tasks 1 and 2.

Similarly, an advisee knows that her advisor has just completed all tasks and has information about the task 1 and 2 performances of her own group. However, the advisee does not know if, or how, his advisor is compensated for giving advice. We thought that this would be the most natural set-up and that advisees would be most independent in their entry decision if they did not take into account the monetary consequences for the advisor.[15]

[15]We decided not to ask advisees for a tentative entry decision, as it is common in the organizational psychology literature, because we were concerned of a possible confirmation bias. Without

3.3 Gender and treatments

Each group of four participants who compete against each other is composed of two women and two men. It has been shown that the gender composition of (potential) competitors can affect the willingness to participate in a competition (see inter alia Gneezy, Niederle, and Rustichini, 2003; Booth and Nolen, 2012; Gupta, Poulsen, and Villeval, 2012). We use an equal gender composition because it is the most 'neutral' set-up and it is the one used in NV07. We made sure however, that participants were not aware that we controlled for the gender composition as there is evidence that making this information salient changes people's behavior (Iriberri and Rey-Biel, 2011).[16] Instead, we had each group of four sharing the same row in the computer laboratory and told participants that their competitors were seated in the same row as them.

The group of advisors, who complete tasks 1-4 as in the original NV07 setup, i.e. without receiving advice, also serves as our control group to evaluate the effect of receiving advice. We will refer to this group as generation 1. The advisees form generation 2 and are devided into two treatment groups depending on the way advisors and advisees are matched. In the first treatment matching is done irrespectively of gender and participants are not informed about the gender of their partner. We refer to the group of advisees matched in this way as generation 2noM (for 'no matching'). In the second treatment, advisees are randomly matched to an advisor of the same gender and everybody is informed about the gender of their partner at the beginning of the advice stage. We refer to the group of advisees matched in this way as generation 2sM (for 'same gender matching').

3.4 Tasks and payoffs

For their participation, subjects receive a show-up fee of €5 plus €4 for completing tasks 1-4. The group of advisees is paid an additional €2 because they have a waiting period of approximately 15 minutes at the beginning of the experiment. This waiting period was necessary to ensure that advisors and advisees reach the

a tentative entry decision we can not measure advice utilization in a direct, within-subject, fashion, so that instead, we measure it in a between-subject comparison, contrasting entry decisions in our control group with entry decisions of those who receive advice.

[16] In the psychology literature this effect was coined 'stereotype threat' (Steele, 1997). The idea is that if a task is stereotypically be thought of as one in which one gender is better than the other (though in fact this might not be the case), then somebody from the 'weak gender' might underperform simply because he or she is aware of this.

advice stage at roughly the same time.[17] In addition we choose one of tasks 1-4 at random and pay participants according to their performance in that task. Finally, we pay subjects for the self-evaluation task and the advisors for giving advice. On average our participants earned €18.50. The average duration of a session was 1 hour 25 minutes starting with reading aloud the general instructions and finishing after participants filled out a questionnaire. Table 1 provides a timeline and briefly summarizes each task and its compensation scheme.

The experiment was conducted between December 2011 and April 2012 at the Universitat Autònoma de Barcelona. Subjects were recruited from a pool of subjects via the online recruitment system ORSEE (Greiner 2004). A total of 336 subjects, 168 men and 168 women, participated in the experiment. Of the 336 subjects, 168 were assigned the role of an advisor (generation 1), 112 the role of an advisee, who did not receive information about the gender of his advisor (generation 2noM), and 56 the role of an advisee, who did receive information about the gender of his advisor (generation 2sM). The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

3.5 Research questions

With our experiment we would like to address the following questions.

Does advice affect the entry decision of men and women? This is our most basic question. We will examine whether the advisees' entry decisions differ systematically from the advisors' entry decisions, and have a closer look at whose entry decisions are affected (men's, women's, those of the strong performers, the weak performers etc.)? We would also like to know how men and women react to the specific advice they receive ("enter the competition" or "choose the piece rate"), and what happens to the gender gap in entry.

Does advice lead to efficiency gains? This is the central question we would like to answer. We would like to see how advice as a soft intervention fares in terms of efficiency gains as compared to affirmative action or feedback on relative performance. We think of efficiency mainly in terms of a good working self-selection process, in the sense that strong performers are encouraged to enter while weak performers enter less. In the results section, we will discuss in detail how we define

[17] Recall that before entering the advice stage, advisors have to complete tasks 1-4 and the self-evaluation while the advisees only have to complete tasks 1 and 2. During this 15 minutes waiting period, advisees are not yet informed about the content of the experiment.

efficiency.

Through which channel does advice alter the entry decisions? The main goal of our experiment is to see whether advice in a relatively natural setting leads to efficiency gains through improved entry decisions. We have therefore not designed our experiment to pinpoint through which channels advice affects decisions (and neither is our purpose to make a fine distinction of which part of the change in behavior is due to advice as such and which part is an effect of indirect information transmission).[18] Nonetheless, we will be able to shed some light on this question as well. First, all aspects that are factored in in an individual's entry decision are natural candidates for the channels of advice. We might see confidence in relative performance, preferences for competition, or risk attitudes be altered by advice. In addition, advice might help the person who receives advice to think through the problem, see for example Schotter (2003) and Bonaccio and Dalal (2006). We will examine whether this is true in our setting as well.

Do men and women give advice in different ways? As we argue below, we expect an advisor's recommendation to be mainly driven by her risk attitude. If this is true and women are more risk-averse than men, then women would recommend entry less often than men if the advisees' performance is the same.

4 Experimental results

This section presents the results of our experimental study. We first report a few preliminary results, comparing performances across gender and generations. Subsection 4.1 presents our main result, demonstrating how adivce leads to efficiency gains in the entry decisions and what drives these efficiency gains. In subsection 4.2, we examine what happens to the gender gap in entry and to its causes. In these first two subsections we focus on our main 'gender-blind' treatment (generation 2noM). We summarize our main findings from our treatment with gender matching (generation 2sM) in subsection 4.3. Finally, subsection 4.4 examines how advice is given.

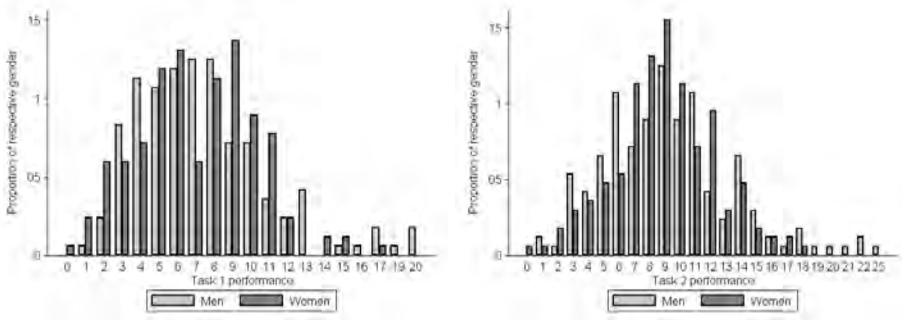
Throughout the paper, whenever we mention performance we mean the number of problems solved, and when we say a subject 'solved' a problem, we take it to mean that the subject solved the problem correctly. If not otherwise noted we use

[18] Wozniak, Harbaugh, and Mayr (2010) use a design that assesses the impact of providing information about relative performance.

all 168 observations for generation 1, 112 observations for generation 2noM, and 56 observations for generation 2sM.

To get a sense for the performance of men and women in the addition task, Figure 1 shows the probability density distributions in task 1 and task 2. On average in task 1, women solved 7.02 problems, while men solved 7.42, and in task 2 women solved 8.95 problems and men solved 9.29. These performance differences between women and men are not statistically significant.[19] We also compared performances in task 1 and 2 in various other ways: across generations 1, 2noM, and 2sM, between women and men for each generation seperately, and across generations 1, 2noM, and 2sM for each gender seperately. We found statistical differences in only one instance: Women in generations 1 and 2sM perform significantly better than women in generation 2noM in task 1, which in turn leads to a significantly better task 1 performance of all participants in generations 1 and 2sM.[20] However, differences in performance do not affect our results as we will condition our further analysis on performance anyway.

Figure 1: PDF of number of correctly solved problems by gender



As in NV07, performances in tasks 1 and 2 are highly correlated, with subjects performing significantly better in task 2 than in task 1. It is not clear to which extent this improvement can be attributed to a learning effect or to the change in incentives when moving from the piece rate to a tournament scheme and we will not attempt to separate the two potential causes (see NV07 for some discussion).

[19]For the sake of brevity, we will omit the details of the tests for our preliminary results, but the interested reader can find those and more descriptive statistics in the appendix.

[20]An indication that this difference is a random event is that the differences constitute less than one third of the standard deviations.

4.1 Efficiency gains in the entry decision

There are a number of ways to measure the efficiency of the entry decision. To organize thoughts, we suggest three different reference groups to evaluate efficiency changes: 1) The 'winner pool' consists of all participants who enter the tournament and win it. On the job market, these are the applicants who successfully apply for a high ranking job. 2) The 'applicant pool' consists of all participants who enter the tournament (who 'apply for the high ranking job'). 3) The 'labor pool' consists of all our participants - in the job market analogy, we think of them as the labor force. If one is not only concerned about the quality of those who obtain a high ranking job or who apply for one, but also wants those who have little or no chance to obtain the job to assess their chances correctly and not waste resources on an application for a high ranking job and accept a low ranking job, this is the right pool to look at. Efficient decisions in this group are reflected by 'correct' self-selection: weak performers refrain from entering the competition, while strong performers do enter. We will be mainly concerned with the efficiency of the entire labor pool.[21] We refer the interested reader to the appendix for an extended discussion of the efficiency gain analysis and of the reasons, why we concentrate on efficiency gains in the labor pool.

To assess the efficiency of the self-selection process in the labor pool, we consider several measures. First, we show that with advice the spread between the ability of those who enter the competition and those who stay away widens and that participants take their own performance better into account. Second, we assess whether a participant's entry decision maximizes her expected earnings and if not, calculate the forgone earnings. We then show that under advice these forgone earnings decrease significantly. Third, we provide evidence that actual task 3 earnings increase under advice.

To show that the spread in performance widens between those who enter the competiton and those who do not, we use four different measures of performance: performance in tasks 1, 2, and 3, and the change in performance from task 1 to task 2. The corresponding average performance values are presented in table 2 for women and men in generations 1 and 2noM. For each of the four subgroups, the average number of solved problems is calculated separately for those who chose the

[21] Efficiency gains in each of the three pools are related. In particular, an improved self selection process in the labor pool implies efficiency gains in the applicant pool. And 'normally' (that is without the artificial feature of competing against the task 2 performance), efficiency gains in the applicant pool should lead to efficiency gains in the winner pool.

piece rate and those who chose the tournament in task 3. As expected, the average performance of those who enter the tournament is better in almost all subgroups for each of the four performance measure.[22]

Table 2: Performance by choice of compensation scheme in task 3

	Generation 1 (without advice)			
	Women		Men	
Average performance	*Piece rate*	*Tournament*	*Piece rate*	*Tournament*
Task 1 (piece rate)	7.05 (3.0)	7.97 (3.2)	6.97 (3.4)	8.34 (4.2)
Task 2 (tournament)	8.82 (3.2)	9.24 (3.1)	**7.70 (4.0)**	**10.32 (4.5)**
Task 2 - Task 1	1.76 (2.3)	1.28 (2.3)	0.73 (3.1)	1.98 (2.9)
Task 3 (choice)	9.42 (3.0)	10.17 (3.3)	**8.22 (3.9)**	**11.21 (5.5)**
	Generation 2noM (with advice and no gender matching)			
	Women		Men	
Average performance	*Piece rate*	*Tournament*	*Piece rate*	*Tournament*
Task 1 (piece rate)	5.83 (3.1)	7.05 (3.7)	**5.09 (2.2)**	**8.18 (4.3)**
Task 2 (tournament)	**7.74 (3.1)**	**9.86 (4.4)**	**6.57 (2.6)**	**11.27 (4.4)**
Task 2 - Task 1	1.91 (2.0)	2.81 (2.2)	**1.48 (2.1)**	**3.09 (2.4)**
Task 3 (choice)	**7.94 (2.8)**	**10.14 (4.2)**	**7.00 (2.4)**	**11.73 (5.2)**

Average number of solved problems for each subgroup (standard deviation in parenthesis).
Sample in generation 1 (2noM) is 84 (56) women and 84 (56) men.
Bold piece rate-tournament value pairs are statistically significantly different.

To test for differences in the performance distributions, we use two-sided Mann-Whitney U tests. In the following, unless otherwise noted, reported test results are those from a two-sided Mann-Whitney U test.[23] In generation 1, our control group that does not receive advice, women who enter the tournament do not perform significantly better than women who choose the piece rate ($p > 0.164$ for each of the four performance measures). That means that the strong and weak performing women do not separate well into those who enter into the competition and those who stay away (which is in line with the findings in NV07). Men in generation 1 take their performance better into account. The task 2 and task 3 performance of those who enter the tournament is significantly better than the task 2 and task

[22]The improvement in performance from task 2 to task 1 of women in generation 1 is lower for those who enter the tournament than for those who choose the piece rate.

[23]Tests showed that the performance variables are not normally distributed, and we will therefore not report results of t-tests. For an easier reading, we will not explicitly describe the Mann-Whitney U test results with "differences in the distribution of the tested variable."

3 performance of those who choose the piece rate (p < 0.011 for the two performance measures), and marginally significantly better for the other two performance measures (p = 0.108 for task 1 and p = 0.102 for the change from task 1 to 2). In generation 2noM (with advice) women and men make their decisions more in line with their performance: In task 2 and task 3, women who enter the tournament now perform significantly better than women who choose the piece rate (p < 0.051 for both measures). In task 1 the gap widens and the change from task 1 to task 2 becomes larger for those who enter the tournament even though both gaps are not statistically significant (p > 0.127). For men, the gaps clearly widen and remain statistically significant, no matter which performance measure we look at (p < 0.014 for each of the four performance measures).

Table 3: Logit of entry decision in task 3 for generation 1 and 2noM

	Entry task3			
VARIABLES	(1)	(2)	(3)	(4)
Female	-0.897***	-0.939***	-0.569	-0.0869
	(0.325)	(0.265)	(0.382)	(0.440)
Task2 performance	0.109**	0.125***	-0.0423	-0.060
	(0.0491)	(0.0439)	(0.0612)	(0.0621)
Task2 - Task1 performance	0.0106	0.0221	0.147*	0.280***
	(0.0582)	(0.0591)	(0.0798)	(0.101)
Generation 2noM	-1.344*		-0.882	-1.204
	(0.804)		(0.961)	(0.983)
(G2noM)*Female	-0.00504		0.280	0.547
	(0.540)		(0.630)	(0.651)
(G2noM)*(Task2 performance)	0.168**		0.105	0.126
	(0.0831)		(0.0962)	(0.0975)
Advice "tournament"		1.327***		
		(0.474)		
Advice "piece-rate"		-0.367		
		(0.314)		
Guessed rank in task2			-0.756***	-0.744***
			(0.229)	(0.229)
Tournament entry in task4			2.547***	2.652***
			(0.426)	(0.436)
Preference for competition			0.212**	0.225**
			(0.0893)	(0.0906)
Female*(Task2 - Task1 performance)				-0.279**
				(0.127)
Observations	280	280	280	280
log-Likelihood	-172.2	-168.5	-135.9	-133.4

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Dependent variable takes value 1 for tournament and 0 for piece rate.
Constant is not reported. Sample is generation 1 and 2noM.

Table 3 presents results of a logit regression of the entry decision in task 3. In the first model, the entry decision is regressed on a dummy variable for gender (which takes the value 1 for a female participant and 0 for a male participant), the number of correct answers in task 2, the difference in performance between task 2 and task 1, a dummy variable for generation (which takes the value 1 for generation 2noM and 0 for generation 1), an interaction term between generation and gender, and an interaction term between generation and task 2 performance. We will discuss more aspects of these regressions later. At this point only note that in model one the positive coefficient estimate for the interaction term of generation and task 2 performance is highly significant, confirming that participants in generation 2noM are better in factoring in their own performance than participants in generation 1 when making their entry decision. It may also be noted that, as expected, the coefficient estimate for the dummy variable for gender is negative and highly significant, i.e. women are less likely to choose the competition than men keeping the other variables constant.

We now examine whether entry decisions maximize expected earnings in task 3. If a participant chooses the (for her) payoff inferior payment scheme, we refer to the difference between expected earnings under the payment scheme the participant did NOT choose and the one she chose, as the forgone earnings. We will show that these forgone earnings are significantly reduced if participants receive advice (i.e., comparing generations 1 and 2noM).

To calculate forgone earnings under each payment scheme, we need two ingredients: the number of problems a participant is expected to solve in task 3 and the corresponding probability of winning the tournament in task 3. For the number of problems a participant is expected to solve in task 3, we use the participant's task 2 performance.[24] Table 4 summarizes the results of the probability calculation. Since men and women perform quite similarly these differences are small.[25] We will come back to these probabilities later. The interested reader can find a detailed description of the calculation of the forgone earnings in the appendix.

Table 4: Probability of winning given a certain performance level in task 2

Task 2 performance	<6	6	7	8	9	**10**	11	12	13	>13
Men (in %)	<0.7	0.7	2.1	5.7	13.8	**26.3**	39.7	54.2	65.2	>65
Women (in %)	<0.9	0.9	2.6	6.1	13.6	**24.8**	38.1	51.6	60.1	>60
3 performance groups	Weak			Intermediate				Strong		

[24]There is not significant change in performance from task 2 to task 3, so indeed, task 2 seems to be a good predictor for task 3 performance.

[25]NV07 use a bootstrap method to calculate the probabilities of winning. We do not expect the two approaches to lead to significant differences.

Because the tournament-rate (€2) is four times the piece rate (€0.5), expected earnings are higher in the tournament if the probability of winning is larger than 25%. This is the case if a participant solves at least eleven problems in task 2. If a subject solves eleven or more problems in task 2 and does not enter the tournament, we count this as under-entry. The corresponding forgone earnings from under-entry are the difference between expected tournament earnings and the expected piece rate earnings. The probability of winning the competition is lower than 25% if a participant gives nine or less correct answers in task 2. If such a participant nonetheless enters the tournament, then we count this as over-entry. The forgone earnings from over-entry are the difference between the expected piece rate earnings and the expected tournament earnings in the tournament.

Table 5: Forgone earnings of over- and under-entry in task 3 (choice) tournament for generation 1 and generation 2noM

	Generation 1		Generation 2noM	
	Women	Men	Women	Men
Ex-ante calculation (based on task 2)				
Under-entry	15 of 23	6 of 26	5 of 15	2 of 19
Over-entry	16 of 48	22 of 50	10 of 36	12 of 33
Total forgone earnings	155.6	126.7	64.5	38.1
Average forgone earnings	**2.19**	1.67	**1.26**	0.73
Average forgone earnings (men and women)	**1.92**		**1.00**	

Under-entry: Number of those who do not enter out of those who should enter (>10 correct answers).

Over-entry: Number of those who do enter out of those who should not enter (<10 correct answers).

Average forgone earnings: Average over those who could have made a 'wrong' decision.

i.e. everybody who did not solve exactly ten problems in task 2.

Bold value pairs are statistically significantly different.

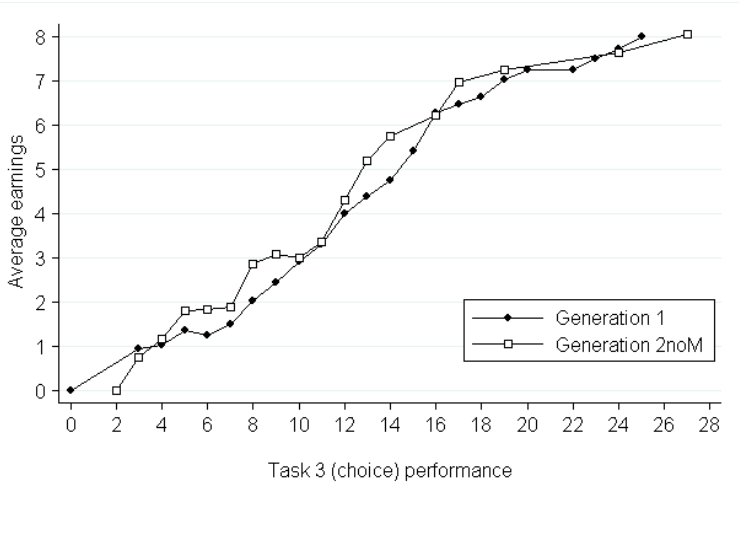
The forgone earnings due to under- and over-entry are summarized in table 5. Columns 1 and 2 report the numbers of women and men who 'under-enter' and 'over-enter' and the total/average forgone earnings. The average is taken over all those who potentially could make a 'wrong decision,' meaning everybody who did not solve exactly ten problems. Columns 3 and 4 report the same values for generation 2noM. The forgone earnings are higher for women than for men in G1 and G2noM because the 'typical mistake' of a woman (under-entry) is costlier than the 'typical mistake' of a man (over-entry).[26] Note that the level of forgone

[26]Under-entry is costlier because it is a mistake of strong performers and, with many correct

earnings is not interesting by itself since it depends on how much is paid for the experiment. We are only interested in relative comparisons across groups.

Pooling men and women, advice reduces average forgone earnings from €1.92 to €1.00, a significant reduction ($p = 0.006$, one-sided Mann-Whitney U test). Breaking down the analysis by gender shows that the larger part of the reduction in forgone earnings is due to the improvement of the entry decisions of women. For them, forgone earnings are significantly lower if they receive advice ($p = 0.012$, one-sided Mann-Whitney U test). For men, average forgone earnings are also reduced if they receive advice, but the reduction is not significant. However, from an ex-post point of view (i.e., if we take actual task 3 performance as our basis for the forgone earnings calculation), forgone earnings are signficantly lower for men as well ($p = 0.062$, one-sided Mann-Whitney U test).

Figure 3: Average earnings if the number of correctly solved problems in task 3 is at most a given level



Our last measure to assess efficiency are actual task 3 earnings. One difficulty with the analysis is that actual earnings depend on performance levels and we have only few observations at each performance level. More over, actual earnings are to a certain extent subject to randomness and do not necessarily reflect the "correctness" of the entry decisions. For example a strong performer who made the right decision and entered the tournament (and performed well in it) might end up with zero payoff instead of €2 per correct answer because another subject in her group

problems, the expected earnings from winning the competition are much higher than the earnings from the piece rate. Over-entry is less costly because it is a mistake of weak performers and the forgone earnings are low (the piece rate times a low number of correct answer).

performed even better in task 2. Since we have only few observations for most performance levels and the spread of payoffs is particularly large among those who enter the tournament, this random element is not averaged out. We deal with this problem by comparing the task 3 average earnings of those who solve at most a certain number of problems in task 3, as shown in figure 3. For instance, those who solve fourteen or less problems earn on average €4.7 in generation 1 and €5.8 in generation 2noM. As is evident from the figure, average earnings in generation 2noM exceed average earnings in generation 1 for almost all performance levels. So indeed, advice does not only in theory improve entry decision, but also has a positive affect on actual earnings.

In summary, we can provide a positive answer to our second research question "Does advice lead to efficiency gains?". Advice improves the self-selection process as evidenced through a widened performance spread between those who enter and those who do not, lower forgone earnings due to wrong entry decisions, and higher actual earnings.[27] Moreover, the general insights from the advice literature we discussed at the end of the literature section, lead us to conjecture that in practice efficiency gains could be increased by the right design choices.

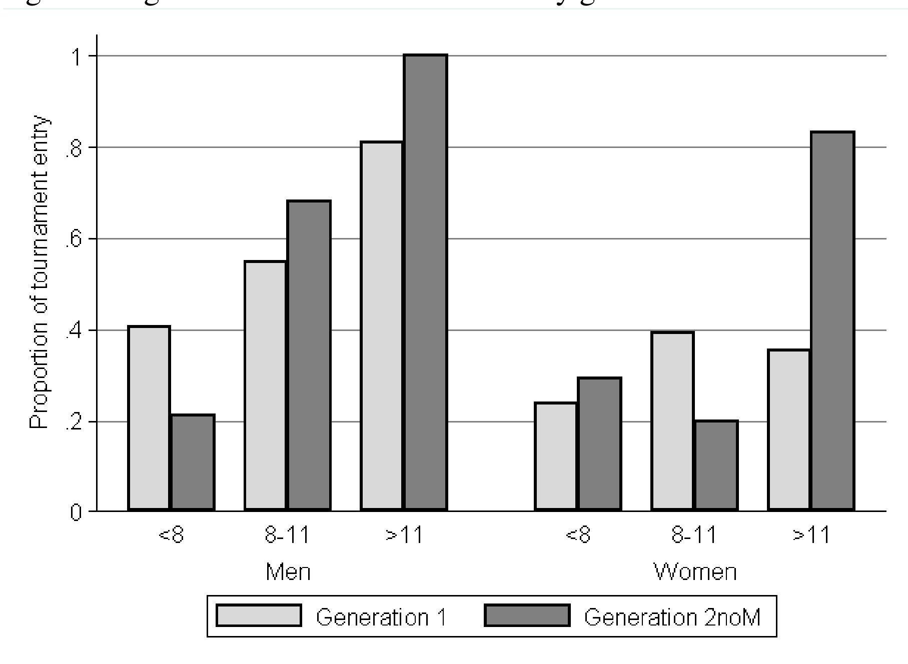
4.1.1 Whose entry decisions are affected?

To examine more closely how advice improves the entry decision we look at the entry decision of women and men conditional on their performance levels. We use task 2 (tournament) performance to create the performance groups because it is arguably more informative about the expected task 3 performance than task 1 performance (In fact, there are no significant differences between performances in tasks 2 and 3.). We split participants into three groups: (1) 'weak' performers, (2) 'intermediate' performers, and (3) 'strong' performers. Weak performers are those who solve between zero and seven problems in task 2 (33% of all 336 participants). Intermediate performers solve between eight and eleven problems (44%) and strong performers solve between twelve and twenty five problems (23%).

[27] Of course, it may be the case that it is optimal for an individual with a strong performance to choose the piece rate, because she is very risk-averse or does not like competitions. In that sense, we cannot assess the optimality of an individual's entry decision. However, using a question from a questionnaire at the end of the experiment, where participants state how much they like to compete on a scale from 1 to 7, we find that if anything advice improves the extent to which participants take into account their own preferences. In generation 1, the average answer to the question is 4.6 and 5.2 for those who chose, respectively, the piece rate and the tournament. In generation 2noM, the average answer is 4.2 and 5.5, that is the gap widens.

There are two rationales for the split we chose. One are the probabilities of winning as summarized in table 3. With at most seven solved problems, a subject's probability of winning the tournament is less than 2.7%. Arguably, even a very risk-loving person who in addition enjoys participating in a competition would not enter the competition, knowing the low chances of winning. Intermediate performers have a probability of winning between 5.7% and 39.7%. This is the group for which it is difficult to judge whether or not they should enter since we do not know their preferences. Strong performers have a probability of winning of at least 51.6% and, except for extreme preferences, for most of them it will be optimal to enter the competition. Another rationale for this split is the likelihood for obtaining a certain rank in the competition. Weak performers have a probability of ranking 3rd or 4th in their groups of at least 78%. Strong performers have a probability of ranking 1st of at least 51% while for intermediate performers the probability of the different ranks is more evenly distributed. [28]

Figure 5: Proportion of men and women who enter the competition for a given range of correct answers in task 2 by generation



Ideally, one would want advice to increase entry rates among strong performers and decrease them among weak performers (while for the intermediate performers,

[28]As a robustness check, we have checked for alternative splits, changing the intermediate group to either 7-11 (second and third quartile), 8-10, 9-10, 9-11, or 10 (always keeping 10 because this is the break-even performance level where expected earnings from the competition and the piece-rate are roughly the same), and this has not changed the qualitative nature of our results.

it is not as clear whether one decision is superior to the other). Figure 5 shows the proportion of men and women who choose the tournament for each of the three performance groups and for generations 1 and 2noM. The direction of change is the expected one for both male and female strong performers and for male weak performers. Only female weak performers unexpectedly enter slightly more often under advice (though the change is not significant).

Clearly, the effect of advice on the entry decision is strongest for the group of high ability (i.e. strongly performing) women. In this group, only 35% of women who do not receive advice enter the tournament (6 out of 17) as opposed to 83% of women who receive advice (10 out of 12). This increase in the entry rate is statistically significant (p = 0.013, one-sided Fisher's exact test). This is particularly important because the group of strong performing participants is the one which should ideally enter the tournament and 'be available for high-ranking jobs'. There are no clearly significant effects in the other subgroups, but we would like to point out that entry by weak performing men is reduced from 41% (13 out of 32) to 21% (4 out of 19) (p = 0.129, one-sided Fisher's exact test). We believe that the effect would be significant in a larger experiment. Also the entry rate of strong performing men if anything increases (from 81% to 100%) (p = 0.146, one-sided Fisher's exact test). In addition, and though it is not our focus in this subsection, note that there is a marginally significant effect (p = 0.108, one-sided Fisher's exact test) for women in the intermediate group: they enter *less* when they receive advice (4 out of 20 instead of 18 out of 46, a reduction from nearly 40% to 20%). We will come back to this when we discuss the gender gap in section 4.2.

4.1.2 Reactions to advice received

For our analysis of how participants react to the advice, we will compare entry rates of those who received a certain piece of advice, for instance "piece rate", with entry rates of those who did not receive any advice (i.e., generation 1). A problem with this comparison is that the performance of those who receive the advice "piece rate" ("tournament") is on average worse (better) than in generation 1 (see section 5.4 on how advice is given). To account for these performance differences, we compare entry rates of those in generation 2noM who received a certain piece of advice (e.g. "piece rate") with the expected entry rate of a reference group in generation 1 whose performance is the same as in the subgroup of generation 2noM who received a certain piece of advice (e.g. "piece rate").[29]

Looking at the pooled data in table 6 (we will get to the rest of the table in section 4.2 on the gender gap), we see that participants react in the expected way on the advice they receive, i.e. they enter on average less if they receive the advice "piece rate" (a reduction from 39.7% to 33.3%) and they enter on average more if they receive the advice "tournament" (an increase from 47.9% to 79.4%). The reaction to the advice "tournament" is considerably stronger than the reaction to the advice "piece rate" (an increase of 66% versus a decrease of 15.1%). This is consistent with findings in the literature that show that individuals react more to positive feedback than negative one (see for example Möbius, Niehaus, Niederle, and Rosenblatt, 2011).

Table 6: (Expected) Entry rates for generation 1 and 2noM in %

Advice received	Piece rate		Tournament	
	Gen 1*	Gen 2noM	Gen 1(**)	Gen 2noM
All				
Pooled women and men	39.7	33.3	47.9	79.4
Women	35.3	26.3	37.9	62.5
Men	45.6	40.5	61.9	94.1
Intermediate performers				
Pooled women and men	45.0	35.5	45.5	71.4
Women	39.0	14.3	36.8	33.3
Men	53.7	52.9	57.2	100.0

*(**) Expected entry rate in a subgroup of G1 with the same task 2 performance distribution as in the corresponding group of G2noM that received the advice "piece rate" ("tournament").

The results from the logit regression in model two of table 3 confirm this finding as well. There, we replace the generation dummy and the interaction terms by two variables: The dummy variable "advice 'tournament'" takes the value 1 if a participant receives the advice 'tournament' and 0 otherwise. The dummy variable "advice 'piece rate'" takes the value 1 if a participant receives the advice "piece

[29]For illustration, let us stick with the advice "piece rate". We calculate for each task 2 performance level that we observe among those who received the advice "piece rate" in generation 2noM, the corresponding entry rate in generation 1. If there is no observation in generation 1, we eliminate the corresponding observation in generation 2noM as well. By doing so, we lost at most two observations in each subgroup. We then calculate what would be the expected entry rate in a subgroup of generation 1 that has the same size and exact same performance distribution as our subgroup of generation 2noM.

rate" and 0 otherwise. The positive and significant coefficient of the variable "advice ‘tournament’" shows that a participant is more likely to enter the tournament if she receives the advice "tournament" compared to receiving no advice (and controlling for performance). For the variable "advice ‘piece rate’," we see that the coefficient is negative and not significant. If anything, a participant is less likely to select into competition if she gets the advice "piece rate" compared to receiving no advice. The absolute value of the coefficient estimate is smaller (and the corresponding p-value is larger) for "advice ‘tournament’" than for "advice ‘piece rate’," which means that participants react more to the advice "tournament" than to the advice "piece rate".

Summarizing, the last two subsections we can answer our first research question "Does advice affect the entry decision of men and women?" affirmative. Going a bit deeper, we have seen that they react more strongly to the advice to enter the tournament than to the advice to choose the piece rate. In addition, we saw that the strongest effects are on the entry decisions of female strong performers and of male weak performers.

4.1.3 Through which channels does advice work?

In this subsection, we first discuss the determinants of tournament entry (and the gender gap), partially replicating and confirming results in NV07 and Niederle and Vesterlund (2011). We then turn to the channels of advice, where we find that changes in self-assessment move in the same way as changes in entry rates, suggesting that this to be one of the main channels.

In the logit regression model three in table 3, we add three possible explanations for the tournament entry decision: (1) Guessed rank for task 2 (where 1 stands for the best rank and 4 for the worst rank), which we use as an inverse measure for confidence, (2) the entry decision in task 4, which we use as a proxy for risk attitudes, and (3) a variable from the questionnaire at the end of the experiment, which asks for the preference for competition on a scale from 1 to 7.[30] All three added variables have significant coefficients and are therefore important determinants of the entry decision, with the signs going in the expected direction. The likelihood of

[30]The rationale for using tournament entry in task 4 as a proxy for risk attitude is that, at least in model three, we control for all variables that could potentially affect task 4 entry except for risk attitudes (task 1 performance is indirectly controlled for by including task 2 performance and the change in performance from task 1 to task 2). Therefore, the differential impact of task 4 entry as explanatory variable should be (mainly) driven by risk attitudes.

entering the tournament goes up with smaller self-assessed rank, lower degrees of risk-aversion, and increased preferences for competition. Note also that in model three, the coefficent for the performance change from task 1 to task 2 is positive and significant, i.e. the larger the performance difference the more likely that a participant enters the tournament.

Having determined which factors matter for the entry decision, we would ideally like to shed some light on the question which of these factors are changed through advice. As we argued in Section 4, we do not expect advice in our experiment to have a significant impact on risk preferences or preferences for competition.[31]

Instead, one channel through which advice in our experiment changes entry decisions is by changing confidence levels. Recall that, in the self evaluation after tasks 1-4, we elicit beliefs about the own rank in tasks 1 and 2. Participants have to guess their possible ranks (ranging from 1-best to 4-worst) in these two tasks and receive €1 per correct guess. We then contrast a participant's actual guess with what would have been his or her *optimal* guess.[32] A participant's optimal guessed rank is the one he or she is most likely to obtain given his or her performance.[33] In the following we say participants are overconfident (underconfident) if their guessed rank is lower (higher) than their optimal guessed rank.

We will only discuss the effect of advice on confidence levels in task 2 since this is the task that is most relevant for the entry decision.[34] There is one interesting aspect, which underscores our previously made point that participants factor in the

[31]It could very well be the case that real-world mentoring programs are apt to influence preferences, but the anonymous one-time interaction in our experiment will hardly do so. Indeed, using self-reported preferences for competition, which we ask for in a questionnaire at the end of the experiment, we do not find any evidence that advice changes these preferences.

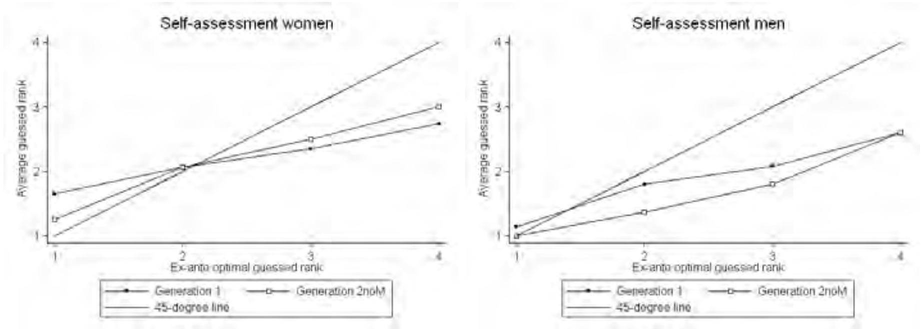
[32]We think this contrast leads to a better measure of confidence than using actual ranks. If we used actual ranks we would, for example, count a person as overconfident whose optimal guess (due to a strong performance) is rank 1, but whose actual rank is 4 because she happens to be matched to other participants with strong performances.

[33]Using the actual distribution of performances among our participants, we calculated the probability of obtaining each of the possible four ranks for each performance level, assuming that the participant is randomly matched in a group of two men and two women. For example, the chances of a female participant with twelve correct answers in task 2 to obtain ranks 1, 2, 3, and 4, respectively, are 51.6%, 38.3%, 9.4%, and 0.8%, respectively. Her optimal guessed rank is thus rank 1.

[34]The general features of confidence levels we find are consistent with findings in the previous literature. In particular, pooled over all generations, participants overestimate their rank in task 2 ($p < 0.002$ for all three generations separately, two-sided Wilcoxon signed-rank test) and also when testing separately for men and women ($p < 0.056$ for men and women in generation 1, 2noM and 2sM separately; women in generation 2sM are the only exception: $p = 0.228$, two-sided Wilcoxon signed-rank test). Men are more overconfident about their task 2 performance than women in generations 1 and 2noM ($p < 0.014$), but not in generation 2sM ($p = 0.486$).

performance change from task 1 to task 2 when they make their entry decision. Comparing confidence levels of tasks 1 and 2, we find that in all three generations, women and men are more optimistic about their rank in task 2 than in task 1 ($p < 0.067$ for each of the six tests, two-sided Wilcoxon signed-rank test). It seems that either subjects do not take sufficiently into account that all participants tend to improve their performance in task 2 or that they feel that they are particularly effective when performing in a tournament.

Figure 6: Guessed rank for task 2 (tournament) performance for women (left panel) and men (right panel)



The self assessment for task 2 is presented separately for generations 1 and 2noM in figure 6, mapping the optimal guessed rank (horizontal axis) against the guessed rank (vertical axis). The graph on the left shows the self evaluation of women and the one on the right shows the self evaluation of men. Each optimal guessed rank corresponds to a performance group: The optimal guess for those with, respectively, 0-6, 7-8, 9-11, and at least 12 correct answers in task 2 is respectively, rank 4, 3, 2, and 1. Note that the group with optimal guessed rank 1 corresponds to the group of strong performers in our previous analysis. To facilitate the reading of the graph, the dotted 45-degree line also represents the optimal guess. Values to the left of that line indicate underconfidence and values to the right indicate overconfidence.[35]

Judging from these graphs, overall, advice does not seem to have a clearly 'correcting' effect in the sense that those who receive advice do not guess more correctly on average. However, strong performing participants in generation 2noM significantly improve their self evaluation ($p = 0.038$).[36] The effect is stronger for

[35]Note that by design those whose optimal guessed rank is 1 cannot be overconfident, while those whose optimal guessed rank is 4 cannot be underconfident.

[36]For ranks two to four, the effect of advice on the self assessment is not significant for women and men separately ($p > 0.354$ for any of the five tests), except for men with the optimal rank two who become even more overconfident ($p = 0.068$).

women than for men (p = 0.038 for women versus p = 0.176 for men). This result mirrors the result of the effect of advice on entry decisions of strong performers. Both female and male strong performers enter the competition more under advice, but only for women the effect is strong and significant. This suggests that for strong performers, particularly women, an important effect of advice is that it makes them more confident.

Table 7: Logit of guessed rank for task 2 for generation 1 and 2noM

	Guessed rank task2		
VARIABLES	(1)	(2)	(3)
Female	0.934***	1.232***	0.693**
	(0.317)	(0.255)	(0.345)
Task2 performance	-0.335***	-0.369***	-0.334***
	(0.0532)	(0.0477)	(0.0537)
Task2 - Task1 performance	-0.224***	-0.235***	-0.316***
	(0.0591)	(0.0592)	(0.0792)
Generation 2noM	0.936		1.089
	(0.742)		(0.742)
(G2noM)*Female	0.667		0.544
	(0.510)		(0.518)
(G2noM)*(Task2 performance)	-0.177**		-0.186**
	(0.0803)		(0.0804)
Advice "tournament"		-1.134**	
		(0.448)	
Advice "piece-rate"		0.260	
		(0.285)	
Female*(Task2 - Task1 performance)			0.188*
			(0.104)
Observations	280	280	280
log-Likelihood	-257.3	-256.2	-255.6

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Ordered logit of guessed rank for the own task 2 performance, constants are not reported.
Dependent variable takes values between 1 (best in group) and 4 (worst in group).

The logit regression in table 7 provides further evidence that one reason why advice has an impact on entry decisions is that it affects confidence levels. In the first model, the task 2 guessed rank is regressed on a dummy for gender, task 2 performance, the performance change from task 1 to task 2, a dummy for generation, an interaction term of generation and gender, and an interaction term of generation and task 2 performance. We find that women are significantly less confident about their performance than men and that advice leads to a more accurate self-assessment (the coefficient of the interaction term of generation and task 2 performance is negative and significant, recall that rank 1 stands for best performer). We confirm in regression model two that the more accurate self-assessment is an important channel through which advice influences the entry decision. As in model two in table 3 (regression of the entry decision), we include instead of the generation dummy and the interaction terms the two dummy variables "advice 'tournament'" and "advice 'piece rate'" in model two in table 7. Participants become more confident if they get the advice "tournament" compared to receiving no advice (and controlling for performance) and become less confident if they get the advice "piece rate" compared to receiving no advice. The coefficient estimate (and the corresponding p-value) for the advice "tournament" is larger (smaller) than for the advice "piece rate", which means that participants' confidence levels change more with the advice "tournament" than with the advice "piece rate," which parallels our result of the larger change in entry rates through the advice "tournament." We will discuss model three in the next subsection on the gender gap.

As is pointed out in the advice literature (see section 2.2), advice tends to improve decisions because it forces subjects to think more carefully. Looking again at the time participants need to arrive at the entry decision, we find some evidence for this effect. Participants in generation 2noM need significantly longer to arrive at a decision (on average 21 seconds) than participants in generation 1 (14.8s). The difference is highly significant (p = 0.000). Thus, subjects who receive advice do indeed seem to think about their entry decision more carefully. This is true even though advisees in our experiment have, in contrast to the advisors, the entire advice stage to think about their entry decision, before arriving at the screen that asks for their entry decision. However, they also have more information to read on that screen since it also informs them about the recommendation of their advisor.

In summary we have obtained at least some insights as to our third research question "Through which channel does advice alter the entry decisions?." We find positive evidence that advice changes entry decisions by changing confidence levels and that receiving advice leads advisees to think more about their decisions. There may be other channels that play a role, which however, we are not able to address with our design.

4.2 Effect on the gender gap in tournament entry

Pooling all performance groups, the overall gender gap is not changed through advice. In generation 1, 47 of 84 men enter the competition in task 3, while only 29 of the 84 women do so. This difference is statistically significant (p = 0.008, Fisher's exact test). In generation 2noM, 33 of the 56 men and 21 of the 56 women choose the tournament, still a significant difference (p = 0.037, Fisher's exact test). This is also confirmed in the regression models one and two in table 3, where the coefficient estimate for the female-dummy is negative and highly significant. In both of these generations, the gender gap, defined as the percentage point gender difference in entry rates, is 21.4 percentage points.

Across performance groups (see figure 5), there are interesting differences regarding the gender gap and how advice affects it.[37] We first note that in generation 1 men enter more often than women in all three groups, but the effect is significant only among the strong performers (p = 0.007, two-sided Fisher's exact test). So the overall gender gap is mostly driven by the differences in entry behavior among strong performers. In contrast, the gender gap in generation 2noM is now driven by the differences in entry behavior among intermediate performers, where 68% (17 out of 25) men and 20% (4 out of 20) women enter (p = 0.002, two-sided Fisher's exact test). The gender gap is reduced and becomes insignificant among participants with high performance levels (100% of men and 83% of women enter, p = 0.478, two-sided Fisher's exact test) and among low performing participants, women enter even slightly more than men. The gender differences in entry changes in the intermediate group are mirrored by gender differences in changes of confidence levels. Whereas women in generation 2noM with the optimal guessed rank two and three become less overconfident compared to generation 1 (the average guessed rank of women increases), men in generation 2noM become even more overconfident than men in generation 1, see figure 6.

Breaking down the analysis of the reaction to advice by gender, we see that the emerging gender gap in the intermediate group is due to differences between men and women in their reactions to the advice they receive.[38] The data summarized in table 6 shows that overall there are no clear differences in the way women and men react to the advice they receive, though women seem to react slightly stronger to both types of advice. Upon receiving the advice "piece rate", they reduce entry by 25.5% while men reduce entry by 11.2%. Upon receiving the advice "tournament",

[37] Again, all results regarding the gender gap that follow are robust to alternative splits, changing the intermediate group to either 7-11 (second and third quartile), 8-10, 9-10, 9-11, or 10.

[38] The two advice types are equally spread between men and women in all three performance groups: Among weak performers in generation 2noM, 5.3% of men and 8.3% of women receive the recommendation to enter the competition. For intermediate performers the corresponding rates are 32% (men) and 30% (women) and for strong performers 83.3% (men) and 75% (women). None of these differences are statistically significant.

women increase entry by 64.9% while men increase entry by 52.0%. However, if we further break down the analysis to performance groups, we find an interesting gender difference in reaction among the intermediate performers (but not for the weak or strong performers). Women in this group react strongly to the advice "piece rate" and show almost no reaction to the advice "tournament", while for men it is the other way around. In the intermediate group, women who receive the advice "piece rate" reduce entry by 63.3% compared to the reference value in generation 1 (from 39% to 14.3%). If they receive the advice "tournament" we see a slight *reduction* of 9.5% (from 36.8% to 33.3%). In contrast, for men in this group who receive the advice "piece rate" we see an almost unchanged entry rate - a slight reduction of 1.5% (from 53.7% to 52.9%). If these men receive the advice "tournament", they increase their entry rate by 74.8% (from 57.2% to 100%). Testing directly for gender differences in reaction to advice within performance groups confirms the impression from these raw numbers. There is no difference within the weak or the strong group ($p > 0.474$ for the two performance groups separately and for advice "piece rate" and "tournament", Fisher's exact test), but there is one for the intermediate group ($p = 0.015$ for the advice "tournament" and $p = 0.057$ for the advice "piece rate", two-sided Fisher's exact test).[39]

Interestingly, the finding that among intermediate performers women are more reluctant in following the advice to enter the competition while men are more reluctant in following the advice to choose the piece rate, is also mirrored in the time participants need for their entry decisions upon receiving advice. The longer a participant needs to choose a compensation scheme after having received advice, the more likely the advice has produced a conflict between the recommendation and the individual's own idea of whether to enter the competition. Thus we interpret the time a participant needs as a proxy for his or her initial decision (which we do not elicit explicitly for reasons discussed earlier). Men (women) in generation 2noM need on average 20.9 (21) seconds to reach a decision. Men who receive the advice "piece rate" need on average longer for their decision (23.4s) than men who receive the advice "tournament" (16.1s) and also longer than women who receive the advice "piece rate" (19.9s). For women, we find the reverse: Upon receiving

[39]We also find gender differences in the reaction to the reasons provided to support the advice given. If the advice "tournament" is supported by emphazising potentially higher earnings/encouraging to trust in one's ability, men enter more often the competition than women ($p < 0.036$, two-sided Fisher's exact test). If the advice "piece rate" is supported by emphazising secure earnings/discouraging to trust in one's ability, women enter less often than men ($p < 0.111$, two-sided Fisher's exact test).

the advice "tournament" they need longer to make up their mind (23.6s) than if they receive the advice "piece rate" and also longer than men who receive the advice "tournament". The difference in response times between men who receive the advice "piece rate" and men who receive the advice "tournament" is significant ($p = 0.043$), but the other time differences are not significant.[40] The time differences become partly larger if we restrict the analysis on the intermediate performers.

An interesting thought experiment is to see what happens to the gender gap if only strong performing women received advice. We saw that 83.3% of this group enter the tournament if they receive advice (most of them receive the advice to enter the tournament). If we assume that 83.3% of strong performing women in generation 1 enter the tournament, we obtain an overall entry rate for women of a bit over 44% (compared to 34.5% if no woman receives advice and compared to 56% of men who enter without receiving advice). In other words, this would reduce the gender gap by close to 10 percentage points, almost cutting it in half.

We use regression model four (and the comparison to model one) in table 3 to identify an additional explanation for the gender gap in entry. In this model we add an interaction term of the gender dummy and the performance change from task 1 to task 2. This term captures gender differences in the perception of the performance improvement or gender differences in associating the performance improvements with the payment schemes. We obtain a significant negative coefficient estimate, i.e. for the same performance change men seem to be more likely to choose the tournament than women keeping the other variables constant. To our knowledge, previous studies have not considered this as an explanation for the gender gap. By contrasting model four with model one, we find again that the three other factors, namely gender differences in confidence levels, risk-attitude, and preferences for competition, seem to explain part of the gender gap in entry in generations 1 and 2noM (this confirms results in NV07).[41] In regression model three of the guessed rank in table 7, we also add an interaction term for gender and the performance change from task 1 to task 2 to model one, which gives us gender differences in

[40]This is because variances of response times are generally quite high.

[41]We cannot easily conclude that the four factors are the only explanations for the gender gap because they are correlated with the dummy variable for gender. Task 2 performance and the performance change from task 1 to task 2 are correlated, too. To see whether the importance of the factors that explain the gender gap change with advice, we add four interaction terms between the generation dummy and all four reasons for the gender gap to model four. There seem to be no significant differences with and without advice as to the importance of any of the four factors confidence, risk attitudes, preference for competition and gender gap in the perception of the performance improvement: None of the four interaction terms is significant.

the perception of the performance change. The positive and significant coefficient confirms that a performance improvement from task 2 to task 1 lifts the confidence of men more than the confidence of women.[42] We can conclude that advice (especially the advice "tournament" for strong performing participants) improves the self evaluation.

4.3 Results for the treatment with same-gender matching

We use this section to briefly summarize the results for our treatment, in which advisors and advisees were matched according to gender (women with women and men with men) and were informed about the gender of their matched partner. The number of advisees in this treatment is relatively low (28 women and 28 men), but recall that our motivation for this treatment was the fact that real-world mentoring programs for women often have more senior women be their mentors. We wanted to verify whether our result that strong performing women significantly increase their participation in the tournament still holds if they receive advice from women.

Indeed, the entry decisions of women of generation 2sM seem to be similarly affected as the ones from women in generation 2noM. In particular, the entry rate of strongly performing women increases to 87.50%, a significantly higher rate than in generation 1. But interestingly, men who receive advice from other men hardly improve in the quality of their entry decision as compared to men who do not get advice. As a result, the improvement in the entry decision is stronger among women than men, by each of the three improvement measures from section 4.1, and the reduced forgone earnings are to an even larger extent than in generation 2noM driven by the improved entry of women.

A difference to the gender-blind treatment is that men react more to the advice "piece rate" and women react more to the advice "tournament". However, overall there are no significant gender differences in the reaction to advice among participants in generation 2sM. The average response times of men and women in generation 2sM exhibit the same pattern as the ones in generation 2noM, suggesting that men have an initial 'taste' for the tournament and women for the piece rate.[43]

[42]The gender difference in the perception of the performance improvement does not differ across generations. Adding to regression model three an interaction term between generation, gender, and performance change, we find that this interaction term is insignificant.

[43]Men who receive the advice "piece rate" need on average longer for their entry decision (23.6 seconds) than men who receive the advice "tournament" (17.7s). For women, we find the reverse: women who receive the advice "piece rate" (14.8s) need on average less time than women who receive the advice "tournament" (22.1s).

As in generation 2noM, a change in confidence levels seems to be one important channel through which advice influences the entry decision: Strong performing women and men become significantly more confident; the effect is stronger for women than for men. Furthermore, the changes in the entry decisions of participants in generation 2sM compared to participants in generation 1 are consistent with the changes in self-assessment. We also find once again some indication for participants who receive advice to think longer about the entry decision. Participants in generations 2sM need significantly longer to arrive at a decision (on average 19.3s) than participants in generation 1 (14.8s). If we perform the time analysis for women and men separately, we make an interesting observation: women who receive advice from women do not take significantly more time to think (the average increases slightly, but not significantly from 16s to 17.7s).

In contrast to our gender-blind treatment, the overall gender gap disappears in generation 2sM: 15 of 28 men and 13 of 28 women select into the competition (p = 0.790, two-sided Fisher's exact test). As in generation 2noM, the clearest gender difference in entry behavior is now among the intermediate performers.[44]

4.4 Provision of advice

In this subsection, we turn our focus to the advisors. The first question we address is whether advice given is 'good', i.e. in line with the information advisors receive from their advisees. This is important in order to guarantee that advice works in the right direction and this is what we would expect after having observed the improvement in the entry decision of participants who receive advice. Afterwards, we check whether women and men give different advice. We will report tests and regressions pooling over all of generation 1, i.e. pooling advisors who are matched with an advisee of the same gender (generation 1sM) and advisors who are not (generation 1noM).[45]

4.4.1 Correctness of advice

We have shown that advice improves the selection into competition, i.e. overall participants take their own performance better into account if they receive advice.

[44]In this group, 53.8% of men (7 out of 13) enter and 23.1% of women (3 out of 13) (p = 0.113, one-sided Fisher's exact test).

[45]We performed all of our analysis for each group of adivsors seperately, but unless otherwise noted, there were no differences in advice giving.

This suggests that advice is 'good' in the sense that advisors recommend the well performing participants to enter the tournament and discourage participants who solve few problems - which is what we find. Recall that advisees inform their advisor on the number of problems they solved in tasks 1 and 2. Additionally, before giving advice, advisors receive information on the performance of all members of their own group in tasks 1 and 2. As one would expect, both types of information are important for the advice.

Advice is overall in line with the performance information provided by the advisees. From the advisors in generation 1, only 12.23% (12 out of 98) of those who get the information that their advisee's number of correct answers is nine or less in task 2 recommend to enter the tournament in task 3, whereas 73.2% (41 out of 56) of those who get the information that their advisee gave at least 11 correct answers, recommend entry. The average task 2 performance information is significantly larger if advisors recommend entering the tournament ($p = 0.000$).[46]

As expected, the information about the performance in the own group plays an important role. If the advisee's reported task 2 performance is smaller than the number of correct answers of the best performing subject in the advisor's own group, 78.46% of the advisors (102 out of 130) suggest to choose the piece rate. If the advisee's reported task 2 performance is better than the highest performance level in the own group, 92.31% of the advisors (24 out of 26) recommend entering the tournament.[47] In none of the just discussed constellations of reported performance and own group performance, there is a gender gap in the choice of advice.

4.4.2 Gender differences in advice giving

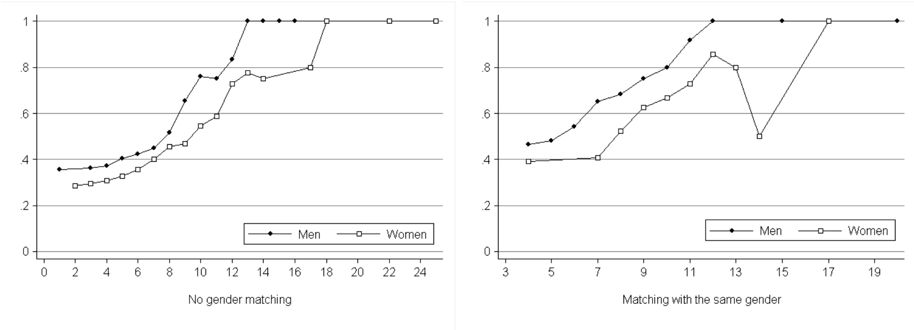
At a first glance, women and men do not give different advice ($p > 0.421$, Fisher's exact test). There are no gender gaps if we use the information on the advisee's performance and split the observations into two subgroups according to whether the advisee would maximize expected earnings under the piece rate (nine or less correct answers in task 2) or in the tournament (eleven or more correct answers in

[46]The difference between task 2 and task 1 performance is larger in generation 2noM than in generation 1noM ($p = 0.114$). This improvement in performance from task 1 to task 2 is perceived positively by the advisors in generation 1noM: Advisors who recommend to enter the tournament observe a significantly larger change from task 1 to task 2 performance of the advisee than advisors who suggest to select the piece rate ($p = 0.030$).

[47]For the same number of correct answers, the advice is quite balanced between piece rate and tournament (8 out of 12 advisors recommend the tournament).

task 2) ($p > 0.227$ for both subgroups, Fisher's exact test).[48]

Figure 8: Proportion of women and men who advise to enter the competition if the information on the number of correctly solved problems in task 2 (tournament) is at least a given level



If we analyze the advice given by women and men conditioning on the performance information, we find that women are throughout the entire performance range more hesitant to advise tournament entry. Figure 8 shows the proportion of women and men who give the advice to choose the tournament among those who receive a self-reported task 2 performance from their advisee of at least a given number. The left panel in figure 8 presents the proportions of women and men in generation 1noM (no gender matching) and the right panel shows the proportion of women and men in generation 1sM (matching with the same gender). For any minimum performance information level and in both treatments, women are more hesitant to advice selection into competition than men except for very high performance levels. Whereas men recommend (men) uniformly to choose the tournament if the information on task 2 is at least thirteen (twelve) correct answers, women advise (women) only to do so if the information is eighteen (seventeen) solved problems or more.

The logit regression model one in table 8 shows that women are more hesitant than men to recommend tournament entry when we control for the advisee's self-reported task 2 performance, the maximum of the task 2 performance in an advisor's own group and whether advisors are matched with the same gender (value 1) or not (value 0). Furthermore, the lower the own group maximum and the higher

[48]We can split the performance information into the three groups 'weak', 'intermediate' and 'strong'. For none of the three ranges, there are gender differences in advice giving ($p > 0.241$, Fisher's exact test). All results regarding the gender gap in advice giving are robust to alternative splits, changing the intermediate group to either 7-11, 8-10, 9-10, 9-11, or 10.

the advisee's performance information are the more likely is the recommendation "tournament".

Table 8: Logit of advice giving for generation 1noM and 1sM

	Advice "tournament"	
VARIABLES	(1)	(2)
Female	-0.755*	-0.630
	(0.431)	(0.444)
Information task2 performance (advisee)	0.572***	0.586***
	(0.0948)	(0.0977)
Maximum task2 performance in own group	-0.171**	-0.151**
	(0.0692)	(0.0757)
Own entry task3		0.837*
		(0.457)
Own task2 performance (advisor)		-0.0510
		(0.0595)
Matching with same gender	0.455	0.503
	(0.441)	(0.447)
Observations	168	168
log-Likelihood	-69.17	-67.32

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Dependent variable takes value 1 for advice "tournament" and 0 for advice "piece rate".

Constant is not reported. The sample is generation 1noM and 1sM.

There are two further indicators that women are more averse to recommend entry into competition than men. One is the time that women and men need to reach a decision for their advice and the other are the reasons they select for the advice. In generation 1noM, women (19.3 seconds) are faster than men (25.1s) among advisors who recommend the piece rate (p = 0.071, one-sided Mann-Whitney U test). The difference is not significant when it comes to the advice "tournament": Women need 27.3 seconds to decide on the advice and men 29.6 seconds (p = 0.494, one-sided Mann-Whitney U test). For generation 1sM, the time analysis is similar.[49]

[49] Interestingly, there is hardly any difference in the decision time betweeen men (24.8 seconds) and women (23s) in generation 1sM when the advice is "piece rate". However, women are slightly faster (21.8s) compared to choosing the piece rate when they advice women to enter the tournament whereas men need notably more time (29.7s). The differences in decision times are all insignificant in generation 1sM. We excluded a male outlier from the analysis who needed 139 seconds to reach a decision.

The hesitation of women becomes clearer when we look at the time women and men need to choose reasons for the advice. While women in generation 1noM are on average 5.7 seconds faster than men if the advice is "piece rate" ($p = 0.088$, one-sided Mann-Whitney U test), men are even slightly faster than women when it comes to choosing reasons for the advice "tournament".[50]

When we add the advisor's own entry decision in task 3 to the logit regression in table 9 (model two), where the value 1 stands for "tournament" and 0 for "piece rate", and control for her own performance in task 2, then the estimates for the coefficient of the remaining explanatory variables do not change substantially compared to the results in model one. However, the coefficient estimate of the gender dummy becomes insignificant and the coefficient estimate of the dummy for the own entry decision is significant on the 10 percent level. This indicates that reasons related to an individual's own entry decision seem to be also important for the advice giving. Participants who entered themselves the tournament are more likely to recommend tournament entry to an advisee. In this vein, mentoring programs make sense in that individuals who decided for the competition are not only examples, but also seem to be more likely to advise facing the challenges of the competition.

Advisors can choose as many reasons as they wish for their advice (including none). For each recommendation, we provide three pre-formulated reasons advisors can choose from. If the advisor recommends the tournament, she can choose from: (1) Because it is fun to compete with others; (2) because you should be confident that you will do well; and (3) because in the competition you can gain much more. Similarly, if the advisor recommends the piece rate, she can choose from: (1') Because it is not fun to compete with others; (2') because you should not be confident that you will do well; and (3') because under the piece rate you gain something for sure. As is evident, reasons (1) and (1') emphasize the (dis)utility from competition, reasons (2) and (2') target the advisee's confidence in her performance, and reasons (3) and (3') use a risk argument. Overall, advisors who recommend to enter the tournament provide more reasons: The most commonly chosen reasons are (3) and (3'), using a risk argument (58 out of 60 if advice is 'tournament', 92 out of 108 if advice is 'piece rate'), followed by reasons (2) and (2'), alluding to the advisee's confidence (52 out of 60 if advice is 'tournament', 64 out of 108 if advice is 'piece rate'). Reasons related to preferences for competition are not chosen very often (25 out of 60 if advice is 'tournament', 7 out of 108 if ad-

[50]Independent of their advice, men need longer than women to choose reasons in generation 1sM.

vice is ‘piece rate’). There are no significant gender differences as to the selection of a particular reason ($p > 0.431$ for any reason separately for advice ‘tournament’ and ‘piece rate’, two-sided Fisher’s exact test) with one exception: If the advice is ‘piece rate’ significantly more men discourage the advisee to believe in her ability ($p = 0.031$, two-sided Fisher’s exact test).[51]

In summary, we have shown that the advice given is good, i.e. participants who perform better receive mainly the advice ‘tournament’ and participants with weaker performance receive mainly the advice ‘piece rate’. As to our fourth research question "Do men and women give advice in different ways?" we find that overall female advisors are more hesitant to recommend tournament entry if we control for important factors such as performance information of the advisee and the best performance in the advisor’s own group. The regression analysis suggests that there is no gender gap in advice giving if we compare women and men who entered the tournament themselves and women and men who did not enter the tournament.

5 Conclusion

We propose advice as a ‘soft intervention’ to improve the (self-)selection into competition and overcome the gender gap in entry into competition. While we have shown that advice indeed improves the efficiency of the selection process (high-ability individuals enter more, low-ability individuals enter less), the gender gap in entry is, at least on the surface, unchanged in our main advice treatment. A closer analysis has shown that the gap goes away among weak and strong performers, but a gap emerges under intermediate performers. This emerging gap is due to different reactions of men a women to the advice they receive.

Overall, our results suggest that advice increases significantly efficiency in the entry behavior, although it may not be suitable to overcome the gender gap in tournament entry entirely. Advice might therefore be a soft alternative to affirmative action, such as quotas or other forms of positive discrimination. Advice has the potential of improving efficiency, rather than only ‘not hurting’ it. The improvement

[51] For generation 1sM separately, there is no gender gap in discouraging the advisee ($p = 0.265$, two-sided Fisher’s exact test), but the direction is the same. When the advice is ‘tournament’ and the advisor encourages the advisee’s confidence in her performance, then male advisees evaluate their task 2 performance more positively than female advisees ($p = 0.000$). When the advice ‘piece rate’ is not supported by discouraging the advisee’s confidence, men are again significantly more positive about their performance than women ($p = 0.056$). However, when the advice is ‘piece rate’ and the advisor discourages the advisee’s confidence in her performance, men and women do not evaluate their task 2 performance significantly differently ($p = 0.531$).

is possible because, as we showed, advice given tends to be good and therefore helps individuals adjust their perception of their relative performance, in addition to helping them think about the decision more carefully. Comparing more closely the changes in entry rates in our experiment compared to others, we find that among strong performing women, the increase in entry through advice (more than 100%) is roughly the same as with the affirmative action "quota" (Balafoutas and Sutter, 2012), and more than through provision of relative performance feedback (Wozniak, Harbaugh, and Mayr, 2010). With advice, the entry rate of strong performing men increases if anything, while high ability men even reduce their entry rates slightly with both, the quota and relative performance feedback. Comparing the impact on the entry decisions of weak performing men, we see a decrease of entry through advice (about 50%), which is similar to the decrease through the provision of relative performance feedback, but much better than the effect of the policy intervention "quota", which leads to a slight increase in entry.

One might argue that advice can be profitable not only for the advisee but also for the advisor (as it is by design in our experimental). In a firm, the senior can profit from a high ability employee in the future through building up a good relation with a potential star. Academic advisors benefit from advising high ability students to enter academia in form of good coauthors in future research projects. In the personal environment, advice of more experienced relatives and friends might benefit them because they care about the well-being of younger family members and friends.

For the moment, we can only speculate whether it matters that the advice in our experiment is *personalized*. Having somebody (exclusively) by the side could trigger an improvement in confidence, and the motivational part of advice that goes beyond the informational content could be important. This remains to be shown. The results from the organizational psychology literature lead us to conjecture that a more personalized form of communication, for example a free chat, would have an even higher potential for efficiency gains. Other naturally arising questions in this context are how the results would be if only (high-ability) women received advice and whether individuals solicit advice and, if so, from whom they ask advice if advice is an option (maybe even making it costly to obtain it). Additionally, analyzing the roles of intensity of the interaction between advisor and advisee and repetition of advice and the willingness of advisor to provide advice would be interesting questions. Finally, our design can easily be extended to allow for free-form communication as for example in Brandts and Cooper (2007) and Brandts, Charness and Ellman (2012).

From a policy perspective, our paper provides some justification for consulting centers which give advice to individuals who are about to make career decisions. Especially educational institutions such as schools and universities can help students to improve their career decisions and choose alternatives that are adequate for their ability. Students should be encouraged to not shy away from asking a senior's advice. Firms can improve the quality of applications, in particular get more applications from highly qualified women and less applications from poorly performing men if they actively approach individuals or students and offer support in their decisions. Particularly valuable might be the advice of recently hired employees who went through the same decision process.

References

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

Balafoutas, L., Kerschbamer, R. and M. Sutter (2011), Distributional preferences and competitive behavior, Journal of Economic Behavior and Organization, in press.

Bettinger, E. and R. Baker (2011), The effects of student coaching in college: An evaluation of a randomized experiment in student mentoring, NBER working paper no. 16881.

Blau, F.D., Currie, J.M., Croson, R.T.A. and D.K. Ginther (2010), Can mentoring help female assistant professors? Interim results from a randomized trial, NBER working paper no. 15707, available at http://www.nber.org/papers/w15707.

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

Booth, A. and P. Nolen (2012), Choosing to compete: How different are girls and boys?, Journal of Economic Behaviour and Organization 81, 542-555.

Brandts, J. and D. Cooper (2007), It's What You Say, Not What You Pay: An Experimental Study of Manager-Employee Relationships in Overcoming Coordination Failure, Journal of the European Economic Association 5, 1223-1268.

Brandts, J., Charness, G. and M. Ellman (2012), Let's Talk: How Communication Affects Contract Design, working paper.

Cason, T.N., Masters, W.A., and R.M. Sheremeta (2010), Entry into winner-take-all and proportional-prize contests: an experimental study, Journal of Public Economics 94, 604-611.

Celen, B. Kariv, S. and A. Schotter (2010), An Experimental Test of Advice and Social Learning, Management Science, Articles in Advance, 1-15.

Charness, G., Rustichini, A. and J. van de Ven (2011), Self-confidence and strategic deterrence, working paper.

Chaudhuri, A., Schotter, A. and B. Sopher (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 (January), 91–122.

Croson, R. and U. Gneezy (2009), Gender Differences in Preferences, Journal of Economic Literature, 47(2), 1–27.

Dargnies, M.P. (2011), Men too sometimes shy away from competition: The case of team competition, working paper, availabel at SSRN: http://ssrn.com/ab stract=1814 989 or http://dx.doi.org/10.2139/ssrn.1814989.

Dohmen, T. and A. Falk (2011), Performance pay and multi-dimensional sorting: productivity, preferences, and gender, American Economic Review 101, 556-590.

Ewers, M. (2012), Information and competition entry, IZA Discussion Paper No. 6411. Available at SSRN: http://ssrn.com/abstract=2028211.

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics 10(2), 171-178.

Flory, J.A., Leibbrandt, A. and J.A. List (2010), Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions, NBER working paper no. 16546.

Gino, F. (2008), Do we listen to advice just because we paid for it? The impact of advice cost on its use, Organizational Behavior and Human Decision 107, 234-245.

Gneezy, U., M. Niederle, and A. Rustichini (2003), Performance in competitive environments: Gender differences, Quarterly Journal of Economics, August, 1049-1074.

Gneezy, U., Leonard, K.L. and J.A. List (2009), Gender differences in competition: evidence from a matrilineal and a patriarchal society, Econometrica 77, 1637-1664.

Greiner, B. (2004), The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics, University of Cologne, Working Paper Series in Economics 10.

Gupta, N.D., Poulsen, A. and M.-C. Villeval (2011), Gender matching and competitiveness: experimental evidence, Economic Inquiry, forthcoming, http:// dx.doi.org/ 10.1111/j.1465-7295.2011.00378.x

Irriberri, N. and P. Rey-Biel (2011), Let's (Not) Talk about Sex: The Effect of Information Provision on Gender Differences in Performance under Competition, working paper.

Iyengar, R. and A. Schotter (2008), Learning under supervision: an experimental study, Experimental Economics, 11, 154–173.

Möbius, M., Niederle, M., Niehaus, P. and T. Rosenblatt (2011), Managing Self-Confidence: Theory and Experimental Evidence, working paper.

Niederle, M. and L. Vesterlund (2007), Do women shy away from competition? Do men compete too much?, Quarterly Journal of Economics, 122, 1067-1101.

Niederle, M. and L. Vesterlund (2011), Gender and competition, Annual Review of Economics, 3, 601-630.

Niederle, M., Segal, C. and L. Vesterlund (2010), How costly is diversity? Affirmative action in light of gender differences in competitiveness, working paper, Stanford University.

Reuben, E., P. Sapienza, and L. Zingales (2010), The glass ceiling in experimental markets, working paper.

Reuben, E., P. Rey-Biel, P. Sapienza, and L. Zingales (2012), The Emergence of Male Leadership in Competitive Environments, Journal of Economic Behavior and Organization, forthcoming.

Rodríguez-Planas, N. (2010), Mentoring, Educational Services, and Economic Incentives: Longer-term Evidence on Risky Behaviors from a Randomized Trial, working paper, Universitat Autònoma de Barcelona.

Schotter, A. (2003), Decision making with naive advice, American Economic Review, 93(2), 196-201.

Schotter, A. and B. Sopher (2003), Social Learning and Coordination Conventions in Inter-generational Games: An Experimental Study, Journal of Political Economy, 111(3), 498-529.

Schotter, A. and B. Sopher (2007), Advice and behavior in intergenerational ultimatum games: An experimental approach, Games and Economic Behavior, 58, 365-393.

Steele, C.M. (1997), A threat in the air: How stereotypes shape intellectual identity and performance, American Psychologist 52(6), 613-629.

Sutter, M. and D. Rützler (2010), Gender differences in competition emerge early in life, IZA discussion paper no. 5015.

Wozniak, D., Harbaugh, W.T. and U. Mayr (2010), Choices about competition: Differences by gender and hormonal fluctuations, and the role of relative performance feedback, working paper, University of Oregon.

For online publication:

Appendix A: Experimental Design

We used the following procedure to control for the gender composition of each group while minimizing the possibility that participants would take note. In our online recruitment system ORSEE (Greiner, 2004), we created two seperate experiments, one for women and one for men, which for them looked like exactly the same experiment. By doing so we ensured that roughly the same number of women and men showed up to each session. Upon arrival to the session, each male participant was given an odd number and each female participant was given an even number (in a way that looked random to them). We told them we used the numbers mainly to do a lottery, determining who could participate and who could not since we needed multiples of eight to participate. We numbered the computer terminals with 1-4 for the first group in the first row, 5-8 in the second row, etc, and asked participants to sit where they saw their number.

Appendix B: Performance in task 1 (piece rate) and task 2 (tournament)

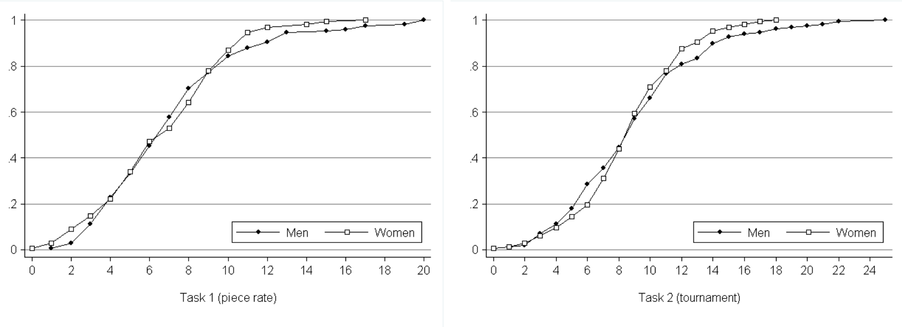
The performances of women and men do not differ, neither under the piece rate nor under the tournament payment scheme. Pooling the data for generations 1, 2noM and 2sM, women solve on average 7.02 and 8.95 problems in tasks 1 and 2, respectively.[52] The average number of problems solved by men is 7.42 in task 1 and 9.29 in task 2.[53] The distribution of performance is not significantly different between women and men ($p > 0.873$ for either task, two-sided Mann-Whitney U test).[54]

[52]The corresponding standard deviations of performances in task 1 and 2 are 3.18 and 3.31. We can pool the data across generations for the analysis of task 1 and 2 performances because there are no treatment differences for these two tasks.

[53]The corresponding standard deviations of performance in tasks 1 and 2 are 3.72 and 4.31.

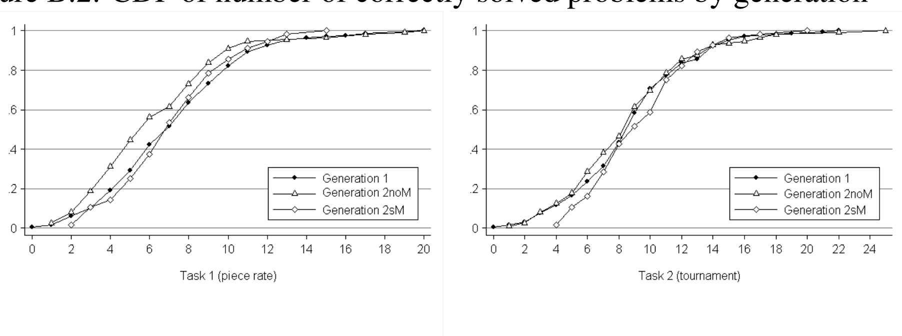
[54]The distribution of the change in performance between task 1 and 2 and the distribution of performance in task 3 are not significantly different between women and men either ($p > 0.822$ for either of the two performance measures, two-sided Mann-Whitney U test).

Figure B.1: CDF of number of correctly solved problems by gender



Comparing performance in generations 1, 2noM and 2sM, we find that, overall, the performance is not very different. In generation 1, participants solve on average 7.55 and 9.07 problems in tasks 1 and 2.[55] The average number of problems solved in generation 2noM in tasks 1 and 2 is 6.60 and 8.94.[56] In generation 2sM, participants solve on average 7.45 problems in task 1 and 9.63 problems in task 2.[57] Participants in generation 1 do not perform differently from participants in generation 2sM in task 1 (p = 0.970, two-sided Mann-Whitney U test). However, participants in generations 1 and 2sM outperform participants in generation 2noM significantly in task 1 ($p < 0.035$ for either of the two generations, two-sided Mann-Whitney U test), even when using the Bonferroni correction. An indication that this performance difference is a random (though somewhat unlikely) event is that it constitutes less than one third of the standard deviation. The distributions of task 2 performance do not differ significantly across generations 1, 2noM and 2sM ($p > 0.167$ for each of the three tests, two-sided Mann-Whitney U test).

Figure B.2: CDF of number of correctly solved problems by generation



[55]The corresponding standard deviations are 3.52 and 3.84 in tasks 1 and 2.
[56]The corresponding standard deviations are 3.60 and 4.10 in tasks 1 and 2.
[57]The corresponding standard deviations are 2.85 and 3.28 in tasks 1 and 2.

The cumulative distribution functions (cdf) for the number of problems solved in task 1 and 2 are presented in figures 1 (Appendix) and 2 (Appendix). The graphs give the probability that women, men, participants in generation 1, 2noM or 2sM solve that many or less problems. Figure 1 (Appendix) displays the cdfs for women and men; Figure 2 (Appendix) displays the cdfs for generations 1, 2noM and 2sM. As is evident, task 1 and 2 performances are very similar across all subgroups.

Looking at each generation separately, there are no significant gender differences in task 1 and 2 performances ($p > 0.425$ for all six tests). If we look at each gender separately and examine differences across generations, we find that most differences are small and insignificant ($p > 0.135$ for ten of the twelve tests. However, women in generations 1 and 2sM perform significantly better than women in generation 2noM in task 1, $p = 0.019$ for generation 1 and $p = 0.081$ for generation 2sM).[58] All together, we conclude that there are no relevant performance differences between women and men so that, all else equal, one would expect them to enter the tournament at similar rates. The small differences that we find across generations are not a problem for us as we use task 2 performance mainly (where we find no generation differences in performance) and will condition our analysis on performance anyway.

As in NV07, performance in task 1 and 2 is highly correlated (Spearman rank correlations of 0.755 for women and 0.675 for men) and subjects perform significantly better in task 2 than in task 1 ($p = 0.000$ for women and men separately, Wilcoxon signed-rank test).[59] This improvement can be due to a learning effect but also be cause by the change in incentives when moving from the piece rate to a tournament scheme, see also NV07.

Appendix C: Efficiency discussion

We suggest three different reference groups to evaluate efficiency changes: 1) The "winner pool" consists of all participants who enter the tournament and win it. On the job market, these are the applicants who successfully apply for a job. If one cares only about the quality of successful candidates, as maybe employers

[58]Comparing task 1 performance of women in generation 2noM and women in generation 1 who give advice to participants in generation 2noM only, there are no significant differences ($p = 0.117$).

[59]The same is true if we test for correlation and performance differences in generation 1, 2noM and 2sM and for women and men in generation 1, 2noM and 2sM separately (Spearman rank correlations between 0.603 and 0.800 for all nine tests; $p < 0.003$ for all nine tests, Wilcoxon signed-rank test).

would do, a natural measure of the efficiency of the entry decisions taken by our participants would be the performance distribution (as a measure of ability) within this winner pool. 2) The "applicant pool" consists of all participants who enter the tournament (who "apply for the job"). If one is concerned about the quality of the applicants (for example because it is difficult to discern the good from the bad ones, or because one wants to minimize the chance of a complete mismatch), a good measure of efficiency would be the performance distribution within this applicant pool). 3) The "labor pool" consists of all our participants - in the job market analogy, we think of them as the labor force. If one is not only concerned about the quality of those who obtain a job or who apply for one, but also wants those who have little or no chance to obtain the job to assess their chances correctly and not waste resources on an application, this is the right pool to look at. Efficient decisions in this group are reflected by "correct" self-selection: weak performers refrain from entering the competition, while strong performers do enter.

While the under-entry of strong performing women is a major concern, we would like to draw attention to the inefficient decisions of those who have little chance of winning, but still enter the tournament. In reality, this is not only a waste of resources for the applicant, but may also harm the potential employer who in turn has to provide more resources to select the best candidate among the applicants. Therefore, we will be mainly concerned with the efficiency of the entire labor pool. An additional reason why, in contrast to Balafoutas and Sutter (2012) and Niederle et al (2010), we do not examine efficiency in the winner pool is owed to our experimental design. Recall that we have adopted from NV07 the feature that a participant who enters the tournament in task 3 competes against the task 2 performance of her group. As a consequence, those who enter the tournament do not compete for a fixed number of "openings." To see this, note that it is possible that everybody who enters the tournament wins it (or that everybody loses it). Comparing the quality of winners can therefore be misleading. For instance, suppose that without advice only the top performer among *all* our participants enters (and wins) the tournament. If with advice many other strong performers enter the tournament (as should be desirable), the quality of the winners must decrease. Therefore we think that this measure does not reflect well efficiency in our context. Furthermore, the conditions, under which winners are determined, change with the introduction of affirmative actions. With affirmative action, the performance of winners can be substantially lower compared to no intervention, in particular, if women perform worse than men. Comparing the quality of winners is therefore a reasonable efficiency analysis. With advice, the conditions, under which winners are determined, do however not change.

Appendix D: Calculation of forgone earnings

To calculate forgone expected earnings under each payment scheme, we need two ingredients: the number of problems a participant is expected to solve in task 3 and the corresponding probability of winning the tournament in task 3. A participant's expected tournament earnings are equal to the expected number of solved problems times the expected probability of winning the tournament times €2. A particpant's expected piece rate earnings are equal to the expected number of problems solved times €0.5. For the number of problems a participant is expected to solve in task 3, we use the participant's task 2 performance.[60] Since one cannot know the ex-post performance in task 3 before deciding on the tournament entry (using the ex-post performance is a somewhat theoretical analysis), the ex-ante performance in task 2 should be the main indicator for the entry decision. We can interpret ex-ante performance as experienced performance under similar circumstances and ex-post performance as the actual performance when entering a competition. We determine the second ingredient, the probability of winning the tournament, as follows. Using the sample of all 336 participants, the probability calculation is done asuming that a participant with a given task 2 performance is randomly grouped with one participant of the same gender and two participants of the other gender. Thus the composition of each possible group is the same as in the experiment: two women and two men.[61] We can then calculate the probability that this performance level is higher than the task 2 performances of three other randomly drawn participants.[62]

[60]We could also use task 1, but evidently task 2 performance is a better predictor for task 3 performance in case of entry into competition. In addition, there is not significant change in performance from task 2 to task 3, so indeed, task 2 seems to be a good predictor for task 3 performance.

[61]Following NV07, we calculated expected earnings in two ways. The 'ex-ante' calculation is the one we just described and report here, but we also did an 'ex-post' calculation where we use a participant's actual task 3 performance for the calculation. The results are virtually the same as for the ex-ante calculation. This is because the subjects who enter the tournament compete against the task 2 performance of their group members. That means, that we compare a participant's task 3 performance with the task 2 performance of all the other participants when we calculate her probability of winning based on her task 3 performance.

[62]We use this approach not only for calculating the probability of winning, but also for the probability of obtaining each of the four possible ranks given a certain performance. NV07 use a bootstrap method to calculate the probabilities of winning. We do not expect the two approaches to lead to significant differences.

Because differences in performance between women and men are small and insignificant, the probabilities of winning the tournament conditional on a certain performance level are similar for women and men.

Because the tournament-rate (€2) is four times the piece rate (€0.5), expected earnings are higher in the tournament if the probability of winning given a specific performance level (i.e. a specific number of problems solved) is larger than 25%. A female participant who solves ten problems in task 2 wins the tournament with a probability of 24.8% and a male participant with ten correct answers wins it with 26.3%. For nine solved problems, the probabilities of winning decrease to 13.6% for a woman and 13.8% for a man, whereas they increase to 38.1% for a woman and 39.7% for a man if the participant gives eleven correct answers in task 2.[63] For lower and higher performance levels in task 2, the probabilities decrease and increase, respectively: With eight or less correct answers, the probabilities of winning are 6.1% or less for both women and men; with twelve or more correct problems, the probabilities of winning are higher than 51.6% for women and men.

A participant should enter the tournament if she solves at least eleven problems in task 2. If a subject solves eleven or more problems in task 2 and does not enter the tournament, we count this as under-entry. The corresponding forgone expected earnings from under-entry are the difference between expected tournament earnings and the expected piece rate earnings. The probability of winning the competition is lower than 25% if a participant gives nine or less correct answers in task 2. If such a participant nonetheless enters the tournament, then we count this as over-entry. The forgone expected earnings from over-entry are the difference between the expected piece rate earnings and the expected tournament earnings in the tournament.

Appendix E: Instructions

General Instructions

Only Generation 1:

In the experiment today you will be asked to complete six different tasks. The

[63] In task 1, women win the tournament with a probability of 24.2% if they solved eight problems. For men, this probability is 22.1%. With seven solved problems, the probabilities are 13.3% for women and 12.9% for men. The probability of winning with nine correct answers is 38.9% and 37.6% for women and men.

method we use to determine your earnings varies across tasks. Before each task, we will describe in detail how your payment of that task is determined. Your total earnings at the end of the experiment is the sum of the following components: (1) A €5 show up fee; (2) €4 for completing Tasks 1-4; (3) In addition, for Tasks 1-4, we will randomly select one of the four tasks and pay you based on your performance in that task; (4) You will be paid for Tasks 5 and 6. Once you have completed all tasks we determine which of the first four tasks counts for payment by drawing a number between 1 and 4. At the end of the experiment, we ask you to stay seated. We will come to you and pay you in private. During the duration of the experiment the use of cell phones is prohibited.

Only Generation 2:

The experiment today will begin with a waiting period of approximately 15 minutes. After these 15 minutes we will instruct you about the next steps. We are asking you to spend the waiting period silently at your assigned seats, without talking to each other or on the phone. You may read or engage in any other quiet activity as you wish. At the end of the experiment you will be paid €2 for having waited quietly.

In the experiment today you will be asked to complete five different tasks. The method we use to determine your earnings varies across tasks. Before each task we will describe in detail how your payment of that task is determined. Your total earnings at the end of the experiment is the sum of the following components: (1) A €5 show up fee; (2) €2 for the waiting period; (3) €4 for completing Tasks 1-4; (4) In addition, for Tasks 1-4, we will randomly select one of the four tasks and pay you based on your performance in that task; (4) you will be paid for Task 5. Once you have completed all tasks we determine which of the first four tasks counts for payment by drawing a number between 1 and 4. At the end of the experiment, we ask you to stay seated. We will come to you and pay you in private. During the duration of the experiment the use of cell phones is prohibited.

Both Generations:

It is important that you do not talk with one another for the duration of the experiment. We also ask you that you do not look at the screens of the other participants. You can ask us at any point in time. If you have a question, please raise your hand and one of the experimenters will come to you.

Task 1 – Piece rate

In Task 1 you have to calculate a series of sums of five two-digit numbers (see "Screenshot Task 1). You will be given 5 minutes to calculate the correct sum of a series of these problems. You cannot use a calculator to determine this sum. However, you are welcome to write the numbers down and make use of the provided scratch paper. You submit an answer by clicking the button "Next" with your mouse. When you submit an answer, the computer will immediately tell you whether the answer is correct or not and a new problem is generated. Your answers to the problems are anonymous.

If Task 1 is the one randomly selected for payment, then you earn 50 cents per problem you solve correctly in the 5 minutes. Your payment does not decrease if you provide an incorrect answer to a problem. We will refer to this payment scheme as the piece rate payment.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 1 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task. After clicking the "Continue" button, the task will begin immediately.

Suppose you have solved 2 problems correctly and 3 problems incorrectly, what is your payment for Task 1 if it is chosen for payment? Answer: €1.60.

Pantalla de la Tarea 1

Task 2 – Tournament

As in Task 1 you will be given 5 minutes to calculate the correct sum of a series of five 2-digit numbers. However, for this task your payment depends on your performance relative to that of a group of other participants. Each group consists of four people; the three other members of your group are located in the same row as you. If Task 2 is the one randomly selected for payment, then your earnings depend on the number of problems you solve compared to the three other people in your group. The individual who correctly solves the largest number of problems will receive €2.00 per correct problem, while the other participants receive no payment. If there are ties the winner will be randomly determined. We refer to this payment scheme as the tournament payment. You will not be informed of how you did in the tournament until you have completed all five tasks.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 2 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task. After clicking the "Continue" button, the task will begin immediately.

Suppose you have solved 2 problems correctly and 3 problems incorrectly, and that everybody else in your group solved 1 problem correctly. What is your payment for Task 2 if it is chosen for payment?

Suppose you have solved 2 problems correctly and 3 problems incorrectly, and that one person in your group solved 3 problems correctly. What is your payment for Task 2 if it is chosen for payment?

Task 3 – Choice

As in the previous two tasks you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. However, now you have to choose which of the two payment schemes, piece rate or tournament, you prefer to apply to your performance in the third task.

If Task 3 is the one randomly selected for payment, then your earnings for this task are determined as follows. If you choose the piece rate you receive 50 Cents per problem you solve correctly. If you choose the tournament your performance will be evaluated relative to the performance of the other three participants of your group in the Task 2-tournament. The Task 2-tournament is the one you just completed. If you correctly solve more problems than the other three members of your group in Task 2, then you receive €2.00 for each correct sum, which is four times the amount from the piece rate. You will receive no earnings for this task if you choose the tournament and do not solve more problems correctly now, than the other three members of your group in the Task 2-tournament. If there are ties the winner will be randomly determined. You will not be informed of how you did in the tournament until all five tasks have been completed.

Only Generation 1:

The computer screen following the control question will ask you to choose whether you want the piece rate or the tournament applied to your performance. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Only Generation 2:

(a) Advice

Before deciding on a payment scheme, you will receive some advice as to which one to choose. Your advisor is a person from the group next door who has already completed Tasks 1-4 and who knows how the members of his own group performed in Tasks 1 and 2. Each member of your group will be randomly assigned a different advisor. First, you are asked to send your advisor information on the number of problems you solved correctly in Tasks 1 and 2. Your advisor will then tell you whether he or she thinks you should enter the tournament and probably also give you a reason for his/her advice.

The next computer screen will ask you to enter the numbers of correct problems you solved in Tasks 1 and 2. You will then have to wait for a moment to receive a message from your advisor.

(b) Entry decision

The computer screen that informs you about the advice you received will ask you to choose whether you want the piece rate or the tournament applied to your performance. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Both Generations:

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 2 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task.

Suppose you have chosen the piece rate and that you solved 3 problems correctly and 1 problem incorrectly. What is your payment for Task 3 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that everybody else in your group solved 1 problem correctly in Task 2. What is your payment for Task 3 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that another person in your group solved 3 problems correctly in Task 2. What is your payment for Task 3 if it is chosen for payment?

Task 4 – Payment scheme for Task 1

You do not have to add any numbers for the fourth task of the experiment. Instead we will pay you again for the number of problems you solved in Task 1 – Piece Rate. However, you now have to choose which payment scheme you want applied to the number of problems you solved. You can either choose being paid according to the piece rate, or according to the tournament.

If the fourth task is the one selected for payment, then your earnings for this task are determined as follows. If you choose the piece rate you receive 50 Cents per problem you solved in Task 1.

If you choose the tournament your performance will be evaluated relative to the performance of the other three members of your group in the Task 1-piece rate. If you correctly solved more problems in Task 1 than the other three members of your group did then you receive four times the earnings of the piece rate, which is €2.00 per correct problem. You will receive no earnings for this task if you choose the tournament and did not solve more problems correctly in Task 1 than the other members of your group. If there are ties the winner is determined randomly.

The next computer screen will tell you how many problems you correctly solved in Task 1, and will ask you to choose whether you would like to apply the piece rate or the tournament rate to your performance.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 4 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose you have chosen the piece rate. Suppose further that you have solved 2 problems correctly and 3 problems incorrectly in Task 1. What is your payment for Task 4 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you have solved 2 problems correctly and 3 problems incorrectly in Task 1, and everybody else in your group solved 1 problem correctly in Task 1. What is your payment for Task 4 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that another person in your group solved 4 problems correctly in Task 2. What is your payment for Task 4 if it is chosen for payment?

Task 5 – Self-evaluation

In this task you are asked to guess your ranks of your performances in Tasks 1 and 2. Since there are four members in your group your rank may be between 1 and 4, with 1 being your rank if you (correctly) solved the largest number of problems in your group and 4 being your rank if you solved the lowest number.

For each correct guess you will receive €1. If your guess is not correct, you will receive no earnings for this guess. In case of ties in the actual ranks, we count every answer that could be correct as correct. For example, if the performance in the group was 5, 5, 4, 4, then an answer of "last position" and "third position" is correct for somebody who solved 4 problems correctly, and an answer of "first position" and "second position" is correct for somebody who solved 5 problems correctly. Note that the numbers used in this example are not indicative of actual performances in Tasks 1 and 2.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 5 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose that in Task 1 you solved 3 problems correctly and the other members of your group solved, respectively, 1, 2, and 3 problems. Suppose further that you estimated your rank to be "second position". What is your payment for this estimate?

Only Generation 1:

Task 6 – Advice

In the room next to us there are other groups who also complete Tasks 1-4 (the same ones you just completed). At this point they have completed Tasks 1 and 2, but have not yet started with Task 3, that is, their next task is to decide between the tournament and the piece-rate. You will be randomly matched to one of them, whom we will refer to as your "advisee", and your task is to advice your advisee in his or her choice between tournament and piece rate. Before you give your advice, your advisee will send you information on the number of problems he or she solved correctly in Tasks 1 and 2.

The first step is that you send your advisee a message telling him or her whether you recommend entering the tournament. In a second step you may give a reason for the advice you choose. For this purpose we provide you with a list of reasons. You may select as many reasons as you wish (including none, in case you don't wish to select any of the reasons provided).

As a payment for this task you will receive 50% of the Task 3 earnings of your advisee. This means that if your advisee chooses the piece rate you receive 25 Cents (50% of 50 Cents) per problem he/she solves correctly. If your advisee chooses the tournament and his/her performance is better than the Task 2 performance of his/her group members, you receive €1.00 (50% of €2.00) for each problem he/she solves correctly. Finally, if your advisee chooses the tournament and his/her performance is not better than the Task 2 performance of his/her group members, you will receive no earnings. Note that you will be paid even if your advisee does not receive a payment for Task 3 (because Task 3 was not the one randomly selected for payment).

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 5 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose your advisee has chosen the piece rate. Suppose further that your advisee solved 3 problems correctly and 3 problems incorrectly. What is your payment for Task 6?

Suppose your advisee has chosen the tournament. Suppose further that your advisee solved 2 problems correctly and 1 problem incorrectly, and everybody else in his/her group solved 1 problem correctly in Task 2. What is your payment for Task 6?

Suppose your advisee has chosen the tournament. Suppose further that your advisee solved 2 problems correctly and 3 problem incorrectly, and another person in his/her group solved 3 problems correctly in Task 2. What is your payment for Task 6?