

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

Jordi Brandts

Institut d'Anàlisi Econòmica (CSIC), Bellaterra 08193, Spain; and Barcelona GSE, Bellaterra 08193, Spain,
jordi.brandts@iae.csic.es

Valeska Groenert

Markets, Organizations and Votes in Economics, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; and
Directorate General for Economic and Financial Affairs, European Commission, Brussels 1049, Belgium,
valeska.groenert@ec.europa.eu

Christina Rott

Department of Business, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; and
Department of Economics (AE1), School of Business and Economics, Maastricht University, 6200 MD Maastricht,
The Netherlands, christina0rott@gmail.com

We conduct a laboratory experiment to study how advice by a more experienced and better-informed person affects an individual's entry into a real-effort tournament and the gender gap. Our experiment is motivated by the concerns raised by approaching the gender gap through affirmative action policies. Overall, advice improves the entry decision of subjects, in that forgone earnings due to wrong entry decisions go significantly down. The improvements are mainly driven by increased entry of strong-performing women, who also become more confident, and reduced entry of weak-performing men. We find that the overall gender gap persists even though it disappears among low and strong performers. The persistence is due to an emerging gender gap among intermediate performers driven by women (men) following more the advice to stay out of (enter) the tournament in this performance group.

Keywords: experiments; advice; gender gap in competitiveness

History: Received December 13, 2012; accepted September 11, 2013, by Uri Gneezy, behavioral economics.

Published online in *Articles in Advance*.

1. Introduction

The underrepresentation of strong-performing women in top-level jobs continues to be a major concern of managers, 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). 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 noncompetitive 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. This gender gap is mainly driven by strong-performing women enter-

ing too little and weak-performing men entering too often.

Subsequently, a number of papers have explored mechanisms 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 et al. 2013, 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 strong-performing men. Another concern is that they actually harm women because women who obtain a job under an affirmative action scheme may be stigmatized of being selected only because of this action. Another set of papers shows that the provision of relative performance feedback leads to a significant improvement of the competition entry behavior and a reduction of the gender gap (Wozniak et al. 2014).

The idea behind our experiment is inspired by the literature on naive advice and the fact that insiders or seniors in firms and institutions possess information and experience about the qualifications necessary for a post. The advice literature consistently shows that advice improves decisions in games such as the ultimatum game or the battle of the sexes game (for a survey, see Schotter 2003). A key difference in our experimental design is that beliefs about own relative performance play an important role for the decision, which is, to our knowledge, not the case in previous studies on the effect of advice. From the results in previous studies, it is not clear whether advice can improve decisions when confidence about own relative performance matters. 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. Advice can occur naturally in everyday situations, or one might channel advice through institutions such as mentoring programs.¹

For purposes of maximal comparability, our experimental design closely follows the design in Niederle and Vesterlund (2007). 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 earlier in the session 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 usually given: in most situations, both previous experience with the same decision as well as information about their own generation's and the advisee's performance influence advisors' recommendation. Wozniak et al. (2014), Ertac and Szentes (2011), and Ewers (2012) show that the

effect of information feedback of different types on entry behavior is positive. Our focus is not on information feedback. Instead, we want to see how people react to advice from another (experienced) person who is somewhat better, but not perfectly informed. In particular, advisors do not know how the advisee performs compared to the advisee's group members, but only compared to a small sample of the advisor's own generation, meaning that the information content is lower than in the design in Wozniak et al. (2014). In situations in which an objective performance ranking is not available in the short run, an experienced and better-informed insider might be able to compare a potential applicant's performance to previous observations and give good advice. Examples for such situations are performance requirements for top-level jobs. In addition, our design can be seen as a first step toward studying advice more in depth, involving matching advisees and advisors of the same or the other gender, more interactive forms of advice such as free-form communication, and people's advice seeking and giving behavior.

Our results show that advice improves the selection into competition by various measures. Without advice, women and men 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. 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. The forgone earnings from the two types of mistakes are significantly lower in our treatment 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 40% to over 80%) and the weak-performing men who enter significantly less (a decrease from over 60% to about 25%). The gender gap persists with advice, because a gender gap emerges among intermediate performers—men (women) follow more the advice "tournament" ("piece rate")—whereas it disappears among weak and strong performers. We also find that advice changes confidence levels and that this is an important determinant of the entry decision.

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

¹ Two examples of mentoring programs are the one organized by the organization Women in Technology and the one by the Committee on the Status of Women in the Economics Profession (CSWEP; see Blau et al. 2010 for an analysis of the effects of the mentoring program run by CSWEP). Our focus is on the advice component rather than the training or role model component of mentoring programs. It is worth mentioning here that mentoring programs typically have costs associated with them, which would need to be considered in a more complete analysis.

2. Literature Review

We first discuss the relevant literature on gender differences in competitiveness and then review the relevant literature on advice and mentoring.

2.1. Competitiveness and Gender Differences

An important stream of the literature on gender differences in competitive behavior (for a survey, see Croson and Gneezy 2009) has focused on the tendency to select into competition. These studies started with the discussion of how incentive schemes affect women's and men's performance differently (Gneezy et al. 2003). The gender gap in entry into competition found by Niederle and Vesterlund (2007) seems to be quite robust, as the results in a number of papers with similar designs show (see, e.g., Balafoutas and Sutter 2012, Booth and Nolen 2012, Cason et al. 2010, Dargnies 2012, Niederle et al. 2013, Sutter and Glätzle-Rützler 2014, Wozniak et al. 2014). In addition, similar gender gaps were found under a variety of different designs (see, e.g., Gupta et al. 2013, Gneezy et al. 2009, Dohmen and Falk 2011) and in the field (Flory et al. 2010).

One important reason for the gender gap in entry into competition seems to be that men are relatively more overconfident (for a review, see Croson and Gneezy 2009), in particular in a tournament environment (Charness et al. 2011, Reuben et al. 2012). Beliefs about one's performance are important not only on the supply side of the labor market. Evidence shows that hiring decisions can be discriminatory against women due to biased beliefs about their abilities (Reuben et al. 2014). Niederle and Vesterlund (2007) underline that, in addition, preferences for competing play an important role. Some studies find that the gender gap remains after controlling for some relevant factors. In their study of young children, Sutter and Glätzle-Rützler (2014) find that the gender gap between boys and girls in entering a competition is robust to controlling for gender differences in risk attitudes, overconfidence, and past performance. In their study with students around 15 years old, Booth and Nolen (2012) also observe a gender gap after controlling for risk attitudes and past performance, 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 et al. 2012), or for confidence levels (Charness et al. 2011). All of these studies use math tasks. Grosse and Riener (2010) show that, after controlling for differences in performance, risk attitudes, and overconfidence, the gender gap persists only in a quantitative, but not in a verbal task, suggesting gender task stereotypes as another explanation for the gender gap.

A number of papers have looked at designs to mitigate or even overcome the gender gap in entry. Niederle

et al. (2013) 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 et al. (2014) show that providing feedback on relative performance can eliminate the gender gap. Shurchkov (2012) shows that reduced time pressure associated with the competitive setting increase female competitive choices. Finally, Dargnies (2012) 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 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 nonoverlapping "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.²

In a similar vein, several field experiments and theoretical papers study the effect of mentoring and coaching on performance, success, and behavior. In particular, Blau et al. (2010) suggest that mentoring programs for female assistant professors lead to a significant increase in performance in terms of publications and grants obtainment.³ Athey et al. (2000) study theoretically how mentoring can lead to different steady states, including a "glass ceiling." They assume that mentor-mentee matching takes place among employees of the same type (e.g., gender), whereby one type (e.g., men) represents the majority at the upper level. In our

² These conclusions are drawn from the results in Celen et al. (2010), Iyengar and Schotter (2008), Schotter and Sopher (2003, 2007), and Chaudhuri et al. (2009).

³ See also Bettinger and Baker (2011) and Rodriguez-Planas (2010) for some more loosely related field experiments.

experimental design, advice does not affect ability, and all participants receive advice independent of their gender.

Finally, advice has been extensively studied in the organizational psychology literature. Bonaccio and Dalal (2006) provide a review of that literature. In the typical setup, 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 setups there is a correct answer to the problem such as estimating the year of a specific event in U.S. 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 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 judgment 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, the quality of the 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 and some further information on the design.

3.1. The Basic Setup

For the purpose of maximal comparability, we keep the experimental design regarding the participation decision as close as possible to the one in Niederle and Vesterlund (2007), which involves three real-effort tasks, two entry decisions, and two self-evaluations. In the real-effort task, participants have five minutes to add up sets of five two-digit numbers without using a calculator. (See the screenshot provided with the instructions in the online appendix, available at <http://docs.google.com/file/d/0B7zDSay6dz3SWi0yWmg2OC1OYVE/edit?pli=1>.) Subjects first perform the task under the noncompetitive 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 noncompetitive payment scheme (task 3, choice). In addition, in task 4 subjects have to decide whether to apply the competitive or the noncompetitive payment scheme to their (past) task 1 performance. Finally, subjects have to rank their performance in tasks 1 and 2 relative to the group members' performances on a scale from 1 (best) to 4 (worst), respectively.

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, her task 3 performance is evaluated against the task 2 performance of her group members. Thus a subject "wins" the tournament in tasks 2 and 3 if she solves more problems correctly than each of her group members in task 2. Ties are broken randomly among the best performers. 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 subjects' entry decisions.

3.2. Advice

Upon arrival to the experiment, subjects are divided randomly and 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 group of advisors, who complete tasks 1–4 as in the original Niederle and Vesterlund (2007) 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.

The advice stage for advisors follows after they have completed tasks 1–4 and the self-evaluation task, whereas 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. Advisor and advisee do not learn the gender of each other. An advisor is paid 50% of her advisee's task 3 earnings. We reward advisors because the main objective of the advice incentive system is to make the advisors give "good" advice. On average, €4.02 was earned for the advice, with payoffs ranging from €0 up to €27. Thus incentives to give "good" advice were substantial. In the real world, the reward of advisors can be nonmonetary in form of building a reputation

Table 1 Timeline of Tasks and Compensation Schemes in the Experiment

Generation 1 (advisors)	Generation 2 (advisees)
Task 1: Five-minute addition task—Piece rate (€0.5)	
Task 2: Five-minute addition task—Tournament (€2, winner takes all)	
Task 3: Five-minute addition task—Selection of compensation scheme	
Task 4: Selection of compensation scheme task 1	Task 1: Five-minute addition task—Piece rate (€0.5)
Self-evaluation, tasks 1 and 2 (€1 per correct guess)	Task 2: Five-minute addition task—Tournament (€2, winner takes all)
Performance feedback, own group	
<i>One advisor randomly matched with one advisee (gender not revealed)</i>	
Receive 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: Five-minute addition task—Selection of compensation scheme
	Task 4: Selection of compensation scheme task 1
	Self-evaluation, tasks 1 and 2 (€1 per correct guess)
	Performance feedback, own group

or in form of a good feeling because of giving good advice to somebody.

The exact sequencing of the advice stage is as follows. The advisee sends information about his task 1 and 2 performances to his advisor.⁴ 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 preformulated reasons for each recommendation (“tournament” or “piece rate”) from which the advisor can select as many as she wishes to. We chose the two-way design of the advice stage to create a feeling of interaction between advisor and advisee.⁵ 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.

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 task 1 and task 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 setup. 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 his advisor has just completed all tasks and has information about the task 1 and task 2 performances of her own group. However, the advisee does not know if, or how, his advisor is compensated for giving advice. We wanted to eliminate the influence of social preferences on the advisee's entry decision and therefore do not provide the advisee with information about the monetary consequences for the advisor. In addition, advisees are put in a more natural advice environment if they do not know the advisor's incentives. Advisees do know that their advisor has some informational advantage, but they need to trust that the advisor will advise them correctly. In previous experimental studies on the effect of information on the relative performance on tournament entry, participants learn their relative performance with respect to the other group members (Wozniak et al. 2014, Ewers 2012) or the performance distribution in general (Ewers 2012) and know that this information is correct. Three key differences from these studies are that advisees in our experiment receive some (imperfect) information indirectly through another person who has made the same decision previously, the information is less accurate than in Wozniak et al. (2014), and advisees need to trust in the correctness of the advisor's recommendation.⁶

⁴ A few weak-performing (strong-performing) advisees sent the advisor higher (lower) performance information than the actual performance level. This did not reduce the quality of advice.

⁵ Ideally, we would have liked the interaction between advisor and advisee to be less structured, with them being able to communicate in a free-form chat with each other. 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.

⁶ We decided not to ask advisees for a tentative entry decision, as is common in the organizational psychology literature, because we were concerned of a possible confirmation bias. Without a tentative entry decision we cannot measure advice utilization in a direct, within-subject, fashion, so instead, we measure it in a between-subject comparison, contrasting entry decisions in our control group with entry decisions of those who receive advice.

3.3. Group Composition, Procedure, and Subject Pool

Each group of four participants who compete against each other is composed of two women and two men.⁷ We made sure, however, that participants were not aware that we controlled for the gender composition because there is evidence that making this information salient changes people's behavior (Iriberry and Rey-Biel 2013).⁸ 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.

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 advice stage at roughly the same time. During this 15 minutes waiting period, advisees are not yet informed about the content of the experiment because we wanted to ensure that the waiting period has no effect on the choice of the compensation scheme in task 3. 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 €18.28. The average duration of a session was 1 hour 30 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 January 2012 at the Universitat Autònoma de Barcelona (UAB). Subjects were recruited from a pool of subjects via the online recruitment system ORSEE (Greiner 2004) and were mainly undergraduate students from UAB. Students were invited to subscribe to ORSEE via flyers distributed and posted in different departments at UAB. We tried to have an equal distribution of subjects from different departments. UAB has a total of 40,000 students, and our subject

pool contains approximately 1,500 students.⁹ There is a very low likelihood that participants of the same group knew each other because all 1,500 students of the subject pool received the invitations for the sessions at the same time, and we assigned participants randomly to generation and group. A total of 224 subjects, 112 men and 112 women, participated in the experiment. Of the 224 subjects, 112 were assigned the role of an advisor (generation 1) and 112 the role of an advisee (generation 2). The experiment was programmed and conducted with the experimental software z-Tree (Fischbacher 2007).

3.4. Research Questions and Hypotheses

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

Does advice lead to efficiency gains and does advice reduce the gender gap in competition entry? This is the central question we would like to answer. A reduction in the gender gap is desirable only if the reduction is achieved in an efficiency-enhancing way. It would, for example, not be desirable if weak-performing women (strong-performing men) entered the competition more (less) often. We think of efficiency mainly in terms of a good self-selection process in the sense that strong performers enter frequently, whereas weak performers tend to stay out, thus maximizing expected earnings (efficiency gains in economic terms).

Advisors are expected to give “good” advice (maximizing the advisee's expected earnings), and advisees are expected to trust their advisor's entry recommendation and select better into competition. Consequently, we expect forgone earnings to be lower with advice. If advice leads to efficiency gains in economic terms, we should see a reduction of the gender gap in tournament entry with advice.

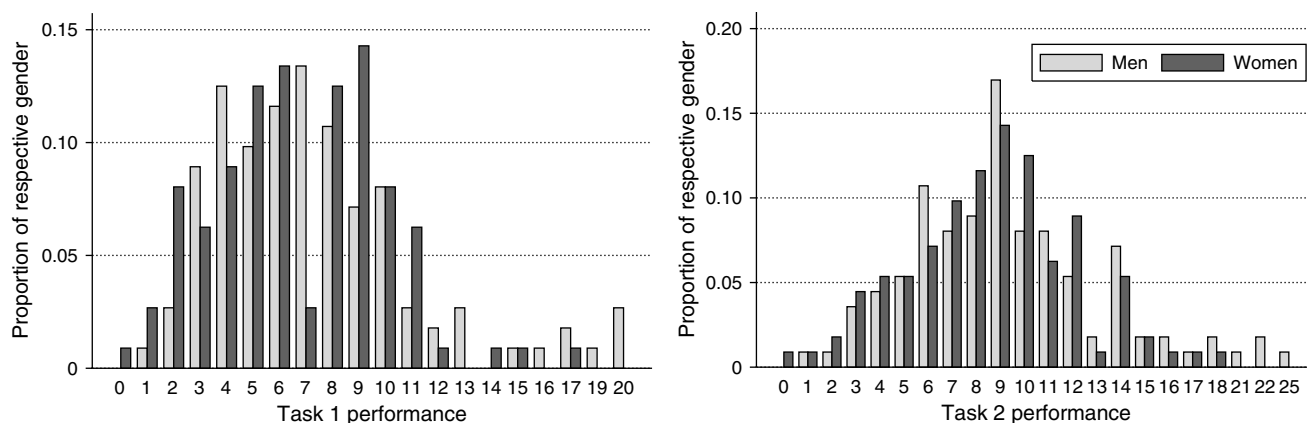
Does advice affect the entry decision of high-performing women and low-performing men in a correcting way? We will analyze the effect of advice in more detail. We will examine whether the advisees' entry decisions differ systematically from the advisors' entry decisions, and have a closer look at whose entry decisions are affected (men's, women's, those of the strong performers, those of the weak performers, etc.). We would also like to know how men and women react to the specific advice they receive (“piece rate” or “tournament”). Splitting participants into three performance groups (weak, intermediate, and strong), we examine whether and how the gender gap is affected in each of the three subgroups.

⁷ It has been shown that the gender composition of (potential) competitors can affect the willingness to participate in a competition (see, inter alia, Gneezy et al. 2003, Booth and Nolen 2012, Gupta et al. 2013). We use an equal gender composition because it is the most “neutral” setup and it is the one used in Niederle and Vesterlund (2007).

⁸ In the psychology literature, this effect was coined “stereotype threat” (Steele 1997). The idea is that if a task is stereotypically being 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.

⁹ One might be concerned about the representativeness of our subject pool (undergraduate university students) with respect to the general population and thus about the external validity of our results. We would like to emphasize that our results and conclusions hold for our subject pool, but we do not claim that the same is true for a different population.

Figure 1 Probability Density Distribution of Number of Correctly Solved Problems by Gender



Since we expect that advisees improve their entry decision compared to the advisors, the two main errors in tournament entry (strong-performing women entering too little and weak-performing men entering too often) are supposed to become less predominant. Consequently, the gender gap in the weak and in the strong performance groups is expected to shrink substantially or disappear with advice. In the intermediate performance group with advice, we expect no differences in the entry decision between women and men.

How does advice affect confidence levels? All aspects that are considered in an individual's entry decision (confidence in relative performance, preferences for competition, or risk attitudes) are natural candidates for the channels of advice. We have 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). Nonetheless, we are able to look at (changes in) confidence levels and their consistency with the entry decisions.

4. Experimental Results

This section presents the results of our experimental study. In §4.1, we report a few preliminary results, comparing performances across gender and generations, stating the entry rates, and how advice is given. Section 4.2 presents our main results, demonstrating how advice leads to efficiency gains in the entry decisions and that, nonetheless, the overall gender gap persists. In §4.3, we examine how advice affects women's and men's entry decisions in three performance groups (low, intermediate, and high) and what happens to the gender gap in entry in each of the three performance groups. We also show that the advice "tournament" increases the confidence level and entry rates significantly, which is particularly important for high-performing women.

Throughout this 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 noted otherwise, we use all 112 observations for generations 1 and 2, respectively.

To test for differences in the performance between (within) subjects, we use Mann–Whitney U (Wilcoxon rank sum) tests.¹⁰ Tests show that the performance variables are overall not normally distributed.

4.1. Preliminary Results and Advice Giving

4.1.1. Performance Distributions and Entry Rates.

To get a sense for the performance in the addition task, Figure 1 shows the probability density distributions of men's and women's performance in task 1 and task 2, respectively, using all 224 observations. On average, in task 1, women solve 6.64 problems and men solve 7.40, and in task 2, women solve 8.59 problems and men solve 9.42. These performance differences between women and men are not statistically significant. 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 online appendix. We also compare performances in tasks 1 and 2 in various other ways: across generations, between women and men for each generation separately, and across generations for each gender separately. We find statistical differences in only one instance: Participants in generation 1 perform significantly better than participants in generation 2 in task 1;¹¹ the difference is not significant for women and men separately. However, differences in task 1 performance do not affect our results, because the more

¹⁰ For an easier reading, we will not explicitly describe the Mann–Whitney U test results with "differences in the distribution of the tested variable."

¹¹ The differences constitute less than one-third of the standard deviations, an indication that this difference is a random event.

Table 2 Probability of Winning Given a Certain Performance Level in Task 2

Task 2 performance	<5	5	6	7	8	9	10	11	12	13	>13
Men (in %)	<0.1	0.3	1.0	2.8	6.6	15.4	29.5	43.0	56.7	66.3	>74.9
Women (in %)	<0.1	0.2	0.9	2.7	6.1	14.6	27.7	40.0	52.5	60.7	>69.9
Performance group (% of participants)	Weak (26%)			Intermediate (52%)				Strong (22%)			
Most likely (= optimal guessed) rank	4 (worst)			3 or 2 ^a				1 (best)			

Note. With 10 correct answers (in bold), expected earnings for the tournament are similar to those for the piece rate.

^aIn the intermediate performance group, men with 11 correct answers are ranked second with a probability of 42.0%, compared to first with 43.0%. For all other men and women in the intermediate group, the probability of being ranked second or third is highest.

relevant performance level for the tournament entry is task 2 performance, and because we will condition our further analysis on task 2 performance anyway.

Performances in tasks 1 and 2 are highly correlated, with subjects performing significantly better in task 2 than in task 1. This is also the case if we test separately for men and women. It is not clear to what extent this improvement can be attributed to a learning effect or to the change in incentives (moving from the piece rate to a tournament scheme), and we will not attempt to separate the two potential causes.

In generation 1, 58.9% of men (33 of 56) enter the tournament in task 3, whereas only 30.4% of women (17 of 56) do so. In generation 2, 58.9% of men (33 of 56) and 37.5% of women (21 of 56) choose the tournament. We will get back to the comparison of men's and women's entry decisions in §4.2.2.

4.1.2. Advice Giving. The advice given is, overall, in line with the performance of the advisees.¹² Remember that we incentivized advisors to give "good" advice without telling the advisees about the monetary incentives for the advisors (see experimental design).

We need two elements to be able to classify whether an advice given is "good" in economic terms: 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. Table 2 summarizes the results of the corresponding probability calculation. Since men and women perform quite similarly, these differences are small. Since the tournament rate (€2) is four times the piece rate (€0.5), expected earnings are the same in the tournament and under the piece rate if the probability of winning is 25%. This is more or less the case if a participant solves 10 problems in task 2; see Table 2. We will come back to Table 2 later

in the results section. The interested reader can find a detailed description of the calculation of the probability of winning in the online appendix.

From the advisors in generation 1, 88.4% (61 of 69) of those whose advisee's number of correct answers is at most nine in task 2 recommend correctly to choose the piece rate in task 3, whereas 67.6% (23 of 34) of those whose advisee gave at least 11 correct answers correctly recommend tournament entry. The task 2 performance is significantly larger if advisors recommend entering the tournament ($p = 0.000$, two-sided Mann–Whitney U test).¹³

As expected, the information about the performance in the advisor's 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, only 19.5% of the advisors (17 of 87) suggest entering the competition. If the advisee's reported task 2 performance is better than the highest performance level in the advisor's own group, 86.7% of the advisors (13 of 15) recommend choosing the tournament.

4.2. The Effect of Advice on Selection into Competition

4.2.1. Efficiency Gains in the Entry Decision. Our general focus is on whether advice improves the efficiency of people's assignment to jobs. In measuring efficiency gains, one can focus only on what happens for high-ranking jobs, or one can take a more global view and consider changes in both high- and low-ranking jobs. We decided in favor of the more global view and measure efficiency gains in terms of gains in expected earnings due to the changes in choosing the payment scheme.

We could measure gains in earnings only among those participants who win the tournament. In broader terms, this would correspond to looking at earnings of successful applicants for high-ranking jobs who are effectively hired. A second possibility would be to

¹² A few weak-performing (strong-performing) advisees informed the advisor about having performed better (worse) than they actually did. From the nine participants in generation 2 who did not send their actual task 2 performance levels, only one participant with 11 correct answers received advice that did not fit his actual task 2 performance. The participant did, however, enter the competition, which is why we believe that the incorrect information provision by some advisees did not affect the quality of advice.

¹³ The improvement in performance from task 1 to task 2 is perceived positively: Advisors who recommend to enter the tournament observe a significantly larger change in advisee performance from task 1 to task 2 than advisors who suggest to select the piece rate ($p = 0.030$, two-sided Mann–Whitney U test).

Table 3 Performance by Choice of Compensation Scheme in Task 3

Choice of compensation	Generation 1 (without advice)		Generation 2 (with advice)	
	Piece rate	Tournament	Piece rate	Tournament
Men				
Task 1 (piece rate)	7.61 (3.4)	8.09 (4.5)	5.09 (2.2)	8.18 (4.3)***
Task 2 (tournament)	8.57 (4.0)	10.15 (4.4)	6.57 (2.6)	11.27 (4.4)***
Task 3 (choice)	8.78 (3.8)	10.73 (5.9)	7.00 (2.4)	11.73 (5.2)***
Women				
Task 1 (piece rate)	6.69 (3.2)	7.71 (2.9)	5.83 (3.1)	7.05 (3.7)
Task 2 (tournament)	8.31 (3.1)	9.41 (3.3)	7.74 (3.1)	9.86 (4.4)*
Task 3 (choice)	8.97 (2.8)	10.12 (3.4)	7.94 (2.8)	10.14 (4.2)**

Notes. The average number of solved problems is shown for each subgroup (standard deviation in parenthesis). Bold piece rate–tournament value pairs are statistically significantly different.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ (two-sided Mann–Whitney U test).

consider gains in expected earnings for all those who enter the tournament, corresponding to all applicants—successful or not—to high-ranking jobs. A third possibility is to measure the gains of expected earnings for all experimental participants, i.e., for the entire labor force, regardless of whether or not they apply for high-ranking jobs. If one is not only concerned about the quality of those who obtain a high-ranking job or who apply for one (i.e., who are available for a high-ranking job), but also wants those who have little or no chance to obtain a high-ranking job to assess their chances correctly, not waste resources on an application for the job, and accept a low-ranking job instead, the entire labor force is the right reference group to look at. Efficient decisions in this group are reflected by “correct” self-selection of all participants in our experiment: weak performers refrain from entering the competition, whereas strong performers do enter. We refer the interested reader to the online appendix for an extended discussion of the efficiency gain analysis.

To assess the efficiency of the self-selection process among the entire potential labor force, we consider two measures. First, we show that, with advice, the spread between the performance of those who enter the competition and those who stay away widens, indicating 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.¹⁴

¹⁴ Ideally, we would like to assess efficiency also in terms of actual task 3 earnings. However, actual earnings do not necessarily reflect the “correctness” of the entry decisions. Since we have only few observations for most performance levels and the spread of payoffs is particularly large among those who (correctly) enter the tournament (€2 in case of winning versus zero payoffs in case of at least one better performer in the same group), the random group composition element is not averaged out.

To show that the spread in performance widens between those who enter the competition and those who do not, we use three different measures of performance: performance in tasks 1, 2, and 3. The corresponding average performance values are presented in Table 3 for women and men in generations 1 and 2. For each of the three subgroups, the average number of solved problems is calculated separately for those who choose the piece rate and those who choose the tournament in task 3. As expected, the average performance of those who enter the tournament is better in all subgroups for each of the three performance measure. 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 ($p > 0.219$ for each of the three performance measures, two-sided Mann–Whitney U test). That means that the strong- and weak-performing women do not separate well into those who enter the competition and those who stay away. For men in generation 1, none of the three performance measures of those who enter the tournament is significantly better than the performance measures of those who choose the piece rate even though the gap in performance is larger than among women in generation 1 ($p > 0.152$ for each of the three performance measures). In generation 2 (with advice), women and men make their decisions more in line with their performance: In tasks 2 and 3, women who enter the tournament now perform significantly better than women who choose the piece rate ($p < 0.051$ for both measures). For task 1 performance, the gap is not statistically significantly different ($p = 0.198$). For men, the gaps clearly widen and become statistically significant, no matter which performance measure we consider ($p < 0.014$ for each of the three performance measures).

Table 4 presents the results of logit regressions of the tournament entry decision in task 3—models (1) through (4)—and of an ordered logit regression of the estimated rank in task 2—model (5). In regression models (1) and (2), separate regressions are presented for generations 1 and 2, respectively. The tournament entry decision in task 3 is regressed on a gender dummy (which takes the value 1 for a female participant and 0 for a male participant), task 2 performance, and the performance change from task 1 to task 2. Remember that the number of observations is the same in models (1) and (2). In model (1) without advice, the coefficient estimate of the performance in task 2 is not significant. In model (2) with advice, the performance in task 2 becomes a highly significant predictor for entry into competition (larger coefficient estimate and smaller p -value). This indicates that advice improves selection into competition in that better-performing participants are more likely to enter the tournament than worse-performing participants.

Table 4 Logit of Tournament Entry Decision in Task 3 (Models (1)–(4)) and Ordered Logit of Guessed Task 2 Rank (Model (5)) for Generations 1 and 2

Variables	Tournament entry (task 3)				Guessed rank (task 2)
	(1)	(2)	(3)	(4)	(5)
<i>Female</i>	−1.153*** (0.405)	−0.906** (0.433)	−1.042*** (0.296)	−1.091*** (0.303)	1.234*** (0.289)
<i>Task 2 performance</i>	0.092 (0.063)	0.252*** (0.0765)	0.0855 (0.0601)	0.112** (0.0509)	−0.353*** (0.0541)
<i>Task 2 – Task 1 performance</i>	0.0268 (0.0788)	0.092 (0.1116)	0.0473 (0.0641)	0.0628 (0.0655)	−0.280*** (0.0673)
<i>Generation 2</i>			−1.476* (0.853)		
<i>(Generation 2) * (Task 2 performance)</i>			0.183** (0.0898)		
<i>Advice “tournament”</i>				1.367*** (0.494)	−1.120** (0.465)
<i>Advice “piece rate”</i>				−0.383 (0.340)	0.332 (0.310)
Observations	112	112	224	224	224
Log-likelihood	−70.5	−63.5	−134.2	−130.4	−201.0

Notes. Coefficient estimates are shown (standard errors in parentheses). The dependent variable *Tournament entry (task 3)* takes the value 1 for tournament and 0 for piece rate. The dependent variable *Guessed task 2 rank* takes values between 1 (best) and 4 (worst). The variable *Advice “tournament”* (*Advice “piece rate”*) takes the value 1 if the advice received is “tournament” (“piece rate”) and 0 otherwise. The constants are not reported. The sample is generation 1 for regression (1), generation 2 for (2), and generations 1 and 2 for regressions (3)–(5).

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The sample for regression models (3)–(5) is generation 1 and generation 2. In model (3), the entry decision is regressed on a dummy variable for gender, 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 2 and 0 for generation 1), and an interaction term between generation and task 2 performance. We will discuss regression models (4) and (5) later. At this point note only that in model (3), the positive coefficient estimate for the interaction term of generation and task 2 performance is significant, confirming that participants in generation 2 consider more their own performance than participants in generation 1 when making their entry decision.¹⁵ 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. We define the forgone earnings as the difference between expected earnings under the payment scheme the participant did *not* choose and those under the one she chose if a participant chose the (for her) inferior payment scheme. Forgone earnings are comparable to the difference between the opportunity cost and the actual earnings of a decision and are therefore a measure of efficiency gains in economic terms. We expect that these forgone earnings will be significantly reduced if participants receive advice and will show that this is indeed the case by comparing forgone earnings in generations 1 and 2.

To calculate forgone earnings under each payment scheme, we need the same two ingredients as in the analysis of the correctness of advice: 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. Table 2 summarizes

¹⁵ We include the interaction term *(Generation 2) · (Task 2 performance)*, because task 2 performance seems to be more important for the entry decision in generation 2 than the performance change; see model (2) of Table 4. We ran a logit regression including the interaction term of generation and performance change in addition to the variables in model (3) of Table 4. The coefficient estimate of the interaction term *(Generation 2) · (Task 2 – Task 1 performance)* was insignificant ($p = 0.622$), and the coefficient estimate of the interaction term *(Generation 2) · (Task 2 performance)* was slightly insignificant ($p = 0.103$). Spearman's rank correlation tests showed that task 2 performance and the performance change are highly correlated, and we think that this is the reason for the slightly insignificant interaction term between generation 2 and task 2 performance in the regression with both interaction terms. We also ran the logit regression (3) of Table 4 with the interaction term *(Generation 2) · (Task 2 – Task 1 performance)* instead of the interaction term *(Generation 2) · (Task 2 performance)* and got an insignificant coefficient estimate for the

interaction term ($p = 0.183$), underlining that task 2 performance is somewhat more important for the entry decision in generation 2 than the performance improvement.

the results of the probability calculation. Remember that expected earnings are the same in the tournament and with the piece rate if the probability of winning is 25%. If a subject solves 11 or more problems in task 2 and does not enter the tournament, we count this as underentry. The forgone earnings from underentry are the difference between expected tournament earnings and the expected piece rate earnings. An example for underentry on the job market would be a highly qualified person who does not apply for a high-ranking job and thereby loses potential high earnings. If a participant with nine or fewer correct answers in task 2 enters the tournament, we count this as overentry. The forgone earnings from overentry are the difference between the expected piece rate earnings and the expected tournament earnings. On the job market, a weakly qualified person who has low probabilities of getting a high-ranking job but nevertheless applies, rejecting sure earnings from a low-ranking job, would be an example for overentry.

Under- and overentry rates become, as expected, smaller with advice. In generation 1, 23.5% (4 of 17) of men and 57.1% (8 of 14) of women “underenter,” i.e., do not enter though they should, whereas the rates are only 10.5% (2 of 19) of men and 33.3% (5 of 15) of women in generation 2. The reduction in the underentry rate with advice is significant for men and women together ($p = 0.091$, one-sided Fisher’s exact test). Among those who do enter the tournament though they should not, there is only a reduction among men: 50.0% (17 of 34) of men and 24.2% (8 of 33) of women in generation 1 “overenter,” whereas 36.4% (12 of 33) of men and 27.8% (10 of 36) of women in generation 2 do so. The reduction of the overentry rate in generation 2 is not significant ($p = 0.314$, one-sided Fisher’s exact test).

The forgone earnings due to under- and overentry are summarized in Figure 2. The bars represent the sum of forgone earnings for generation 1 (175.2) and generation 2 (102.7) and show the forgone earnings

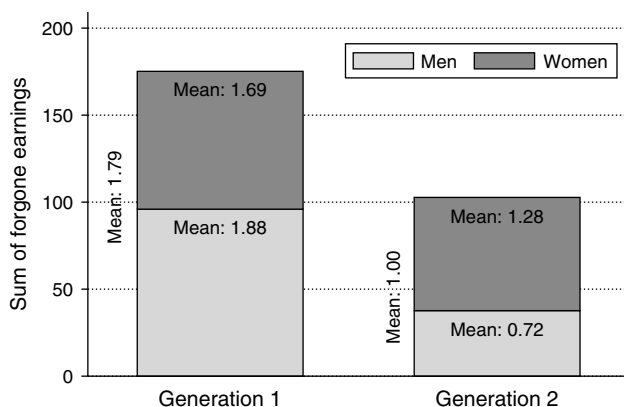
of men (96.0) and women (79.2) in generation 1 as well as of men (37.7) and women (65.1) in generation 2. The average forgone earnings are indicated for each subgroup and are calculated over all those who potentially could make a “wrong decision” in each subgroup, respectively. Note that the level of forgone earnings is not interesting by itself since it depends on how much is paid for the experiment. We are only interested in the relative comparison across the generations. Average forgone earnings are 1.79 (1.88 for men, 1.69 for women) in generation 1, and 1.00 (0.72 for men, 1.28 for women) in generation 2. Pooling men and women, advice reduces forgone earnings significantly ($p = 0.055$, 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 men. For them, forgone earnings are significantly lower if they receive advice ($p = 0.087$, one-sided Mann–Whitney U test). For women, forgone earnings become also smaller if they receive advice, but the reduction is not significant ($p = 0.220$, one-sided Mann–Whitney U test).

In summary, we can provide a positive answer to our expectation that advice leads to efficiency gains in economic terms. Advice improves the self-selection process as evidenced through a widened performance spread between those who enter and those who do not, and lower forgone earnings due to wrong entry decisions. 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 the answers 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 answers to the question are 4.5 and 5.2 for those who choose, respectively, the piece rate and the tournament. In generation 2, the average answers are 4.2 and 5.5; that is, the gap widens. Moreover, the general insights from the advice literature we discuss at the end of the literature section lead us to conjecture that in practice efficiency gains could be increased by the right design choice.

4.2.2. The Gender Gap With and Without Advice.

After having seen that advice improves the efficiency in tournament selection, we would expect that the commonly found gender gap in tournament entry is reduced with advice. First we have a look at the gender gap in generation 1 (without advice). In generation 1, 58.9% of men (33 of 56) enter the competition in task 3, whereas only 30.4% of women (17 of 56) do so. This difference is statistically significant ($p = 0.004$, two-sided Fisher’s exact test). In generation 2 (with

Figure 2 Forgone Earnings of Tournament Over- and Underentry in Task 3 for Generations 1 and 2 (Based on Task 2 Performance)



advice), 58.9% of men (33 of 56) and 37.5% of women (21 of 56) choose the tournament, still a significant difference ($p = 0.037$, two-sided Fisher's exact test). The gender gap, defined as the percentage point gender difference in entry rates, is 28.5 percentage points in generation 1, whereas it is still 21.4 percentage points in generation 2. The gender gap is also confirmed in the regression models (1) and (2) in Table 4, where the coefficient estimates for the female dummy are negative and highly significant. The overall gender gap is not changed considerably through advice, a surprising result after having confirmed efficiency gains in tournament entry through advice. We will first look at how the advice received changes participants' entry behavior and then analyze in more detail whose decision is changed with advice depending on the gender and the performance level.

4.2.3. Effects of the Two Types of Advice. First of all, let us note that the advice that men and women receive does not differ ($p = 0.840$, two-sided Fisher's exact test): 37 men and 39 women (of 56, respectively) receive the advice to choose the piece rate; the remaining 19 men and 17 women are advised to enter the competition. For our analysis of how participants react to the advice, we compare entry rates of those who receive a certain piece of advice, for instance "piece rate," with entry rates of those who do not receive any advice (i.e., generation 1). A difficulty with this comparison is that the performance of those who receive the advice "piece rate" ("tournament") is, on average, worse (better) than the performance in generation 1. To account for these performance differences, we compare entry rates of those in generation 2 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 distribution is the same as in the subgroup of generation 2 who received a certain piece of advice (e.g., "piece rate").¹⁶

Looking at the data in Table 5, we see that participants react in the expected way to the advice they receive, i.e., they enter, on average, less if they receive the advice "piece rate" (a reduction from 38.5% to 33.3%), and they enter, on average, more if they receive the advice "tournament" (an increase from 49.4% to 79.4%). The reaction to the advice "tournament" is considerably stronger than the reaction to the advice "piece rate" (an increase of 59.3% versus a decrease

Table 5 (Expected) Entry Rates for Generations 1 and 2 (in %)

	Generation 1 ^a	Generation 2	% change wrt max ^b
Pooled (men and women)			
Advice "piece rate"	38.5	33.3	−13.4
Advice "tournament"	49.4	79.4	59.3
Men			
Advice "piece rate"	44.9	41.7	−7.3
Advice "tournament"	53.3	93.3	85.7
Women			
Advice "piece rate"	25.8	25.0	−3.1
Advice "tournament"	45.0	62.5	31.8

^aExpected entry rate in a subgroup of generation 1 with the same task 2 performance distribution as in the corresponding group of generation 2 that received the advice "piece rate" ("tournament").

^b"% Change wrt max" refers to the change in the (expected) entry rate between generations 1 and 2 with respect to the greatest possible correcting change.

of 13.4% with respect to the greatest possible change, respectively). This is consistent with findings in the literature that show that individuals react more to positive feedback than negative feedback (see, e.g., Möbius et al. 2013).

The results from the logit regression in model (4) of Table 4 confirm this finding as well. There, we replace the generation 2 dummy and the interaction term 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 rate" and 0 otherwise. All 224 observations from generations 1 and 2 are included in the regression. The positive and significant coefficient estimate 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 "piece rate"* than for *Advice "tournament,"* which means that participants react more to the advice "tournament" than to the advice "piece rate."

The ordered logit regression model (5) in Table 4 provides evidence that advice affects confidence levels and that this is one reason why advice—in particular the advice "tournament"—has an impact on entry decisions.¹⁷ As in Niederle and Vesterlund (2007) and

¹⁶ 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 2, the corresponding entry rate in generation 1. If there is no observation in generation 1, we eliminate the corresponding observation in generation 2 as well. By doing so, we lose 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 2.

¹⁷ The general features of confidence levels we find are consistent with findings in the previous literature. In particular, in both generations, participants overestimate their rank in task 2 ($p < 0.001$

Niederle and Vesterlund (2011), preference for competition, risk attitudes, and self-confidence are three important factors for the entry decision among our participants.¹⁸ In model (5) in Table 4, the task 2 guessed rank is regressed on a dummy for gender, task 2 performance, the performance change from task 1 to task 2, and the dummy variables *Advice "tournament"* and *Advice "piece rate."* All 224 observations from generations 1 and 2 are included in the regression. Participants become significantly more confident if they get the advice "tournament" compared to receiving no advice (and controlling for performance), and become (insignificantly) 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." This parallels our result of the larger change in entry rates through the advice "tournament."

Breaking down the analysis of the reaction to advice by gender, we see that men react more strongly to the advice they receive, taking into account the ceiling problem that the minimum and the maximum entry rates are 0% and 100%, respectively. The entry rate of men (women) who receive the advice "piece rate" decreases from the expected rate of 44.9% (25.8%) in generation 1 to 41.7% (25.0%) in generation 2, a decrease with respect to the greatest possible decrease of 7.3% (3.1%). If the advice received is "tournament," the (expected) entry rates are 53.3% (45.0%) for men (women) in generation 1 and 93.3% (62.5%) in generation 2. Here, the increases in the entry rate with respect to the greatest possible increase are 85.7% for men and 31.8% for women.

for generations 1 and 2 separately, two-sided Wilcoxon signed-rank test) and also when testing separately for men and women ($p < 0.029$ for both genders and generations separately, two-sided Wilcoxon signed-rank test). Men are more overconfident about their task 2 performance than women in generations 1 and 2 ($p < 0.012$, two-sided Mann-Whitney *U* test).

¹⁸ We add three possible explanations for the tournament entry decision to the logit regression model (3) of Table 4: (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. The differential impact of task 4 entry as an explanatory variable should be (mainly) driven by risk attitudes because we control for all other known variables that could potentially affect task 4 entry. All three added variables have significant coefficient estimates (at the 5% and 1% levels) and are thus important determinants of the entry decision: the likelihood of entering the tournament goes up with smaller self-assessed rank, lower degrees of risk-aversion, and increased preferences for competition.

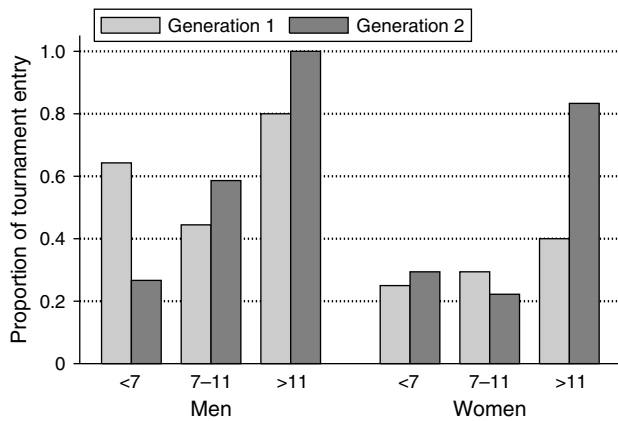
4.3. The Effect of Advice Depending on Performance

4.3.1. Whose Entry Decisions Are Affected? To examine more closely how advice improves the entry decision and why the gender gap in tournament entry persists with advice, we look at the entry decisions of women and men conditional on their performance levels. We use again the 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. We split participants into three groups according to the performance quartiles: (1) "weak" performers solve 6 or less problems (26% of all 224 participants), (2) "intermediate" performers give between 7 and 11 correct answers (52%), and (3) "strong" performers solve 12 or more problems (22%). For an overview, see also Table 2.

Another rationale for this split is the most likely rank in the competition given a certain task 2 performance level.¹⁹ Weak (strong) performers are most likely ranked fourth (first) without exception. For weak (strong) male and female performers, the probability of rank 4 (1) is at least 48.1% (52.5%). Remember that the fourth (first) rank stands for worst (best) performer in task 2. In the intermediate performance group, men with 11 correct answers are ranked second with a probability of 42.0%, compared to first with 43.0%. For all other men and women in the intermediate group, the probability of being ranked second or third is highest. The rank a participant is most likely to obtain given her performance is at the same time the optimal guessed rank in the self-evaluation of task 2 performance.

Ideally, one would want advice to increase entry rates among strong performers and decrease them among weak performers (whereas for the intermediate performers, it is not as clear whether one decision is superior to the other). Figure 3 shows the proportions of men and women who choose the tournament for each of the three performance groups and for generations 1 and 2. Clearly, advice reduces the two common errors in tournament entry: The effect of advice is particularly strong for the group of high-ability (i.e., strong-performing) women. In this group, only 40% of women who do not receive advice enter the tournament (4 of 10), as opposed to 83% of women who receive advice (10 of 12); the increase in the entry rate

¹⁹ Using the actual distribution of task 2 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 12 correct answers in task 2 obtaining ranks 1, 2, 3, and 4 are 52.5%, 37.9%, 8.9%, and 0.7%, respectively. Her most likely rank is thus rank 1.

Figure 3 Proportions of Men and Women Who Enter the Competition for a Given Range of Correct Answers in Task 2 by Generation

is statistically significant ($p = 0.048$, one-sided Fisher's exact test). This is particularly important because the group of strong-performing participants is the one that should ideally enter the tournament and "be available for high-ranking jobs." Also the entry rate of strong-performing men if anything increases (from 80% to 100%; $p = 0.156$, one-sided Fisher's exact test). The other common error in tournament entry (weak-performing men entering the tournament too often) is also reduced significantly as expected: entry by weak-performing men is reduced from 64% (9 of 14) to 27% (4 of 15; $p = 0.048$, one-sided Fisher's exact test). There are no significant effects in the other subgroups. Note that there are different effects for men and for women in the intermediate group: men enter *more* and women enter *less* when they receive advice. We will discuss this in the next subsection.

We find some indication that advice changes confidence levels particularly among strong performers. Recall that in the self-evaluation after tasks 1–4, we elicit beliefs about one's own rank in tasks 1 and 2, ranging from 1 (best) to 4 (worst). Note that, by design, rank 4 (weak-performing) and rank 1 (strong-performing) candidates can only over- and underestimate their performance, respectively. In Table 6, the self-assessment for task 2 is presented separately for men and women in generations 1 and 2 for each performance group. Remember that the optimal guessed

ranks for weak, intermediate, and strong performers are 4, 3 or 2, and 1, respectively. We expect that strong-performing women become more confident with advice (the guessed rank decreases), and that weak-performing men become less confident with advice (the guessed rank increases). As expected, strong-performing women and also strong-performing men become significantly more confident ($p = 0.053$ for men and $p = 0.020$ for women, one-sided Mann–Whitney U test). Surprisingly, the confidence level of weak-performing men does not decrease ($p = 0.455$, one-sided Mann–Whitney U test); weak-performing women adjust their confidence level slightly downward ($p = 0.092$, one-sided Mann–Whitney U test). Among intermediate performers, the self-assessment of women does not change with advice ($p = 0.4892$, two-sided Mann–Whitney U test), whereas men become more confident with advice ($p = 0.098$, two-sided Mann–Whitney U test).

4.3.2. Changes in the Composition of the Gender Gap. Across performance groups (see Figure 3), there are interesting differences regarding the gender gap and how advice affects it. Remember that the strong and the weak performers represent 22% and 26% of all observations, respectively, and the intermediate performers account for 52%. We first note that, in generation 1, men enter more often than women in all three groups, and the effect is significant among the strong and the weak performers ($p = 0.087$ and $p = 0.062$, two-sided Fisher's exact test), but not among the intermediate performers ($p = 0.287$, two-sided Fisher's exact test). The typical gender entry errors become very clear when participants do not receive advice, i.e., strong-performing women enter too seldom and weak-performing men enter too often. The overall gender gap in generation 1 is thus mostly driven by the differences in entry behavior among strong and weak performers. In contrast, the gender gap in generation 2 is now driven by the differences in entry behavior among intermediate performers, where 59% (17 of 29) men and 22% (6 of 27) women enter ($p = 0.007$, two-sided Fisher's exact test). The gender gap 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). Among weak-performing participants, women

Table 6 Average Guessed Task 2 Ranks of Men and Women in Generations 1 and 2

Performance group	Weak		Intermediate		Strong	
	Gen 1	Gen 2	Gen 1	Gen 2	Gen 1	Gen 2
Men	2.64 (1.2)	2.60 (1.0)	1.89 (0.9)	1.52 (0.7)*	1.20 (0.4)	1.00 (0.0)*
Women	2.58 (0.9)	3.00 (0.7)*	2.15 (0.6)	2.22 (0.6)	1.70 (0.5)	1.25 (0.5)**
Optimal guessed rank	4		3 or 2		1	

Notes. Standard deviations are in parentheses. Bold generation 1–generation 2 value pairs are statistically significantly different. Except for men with 11 correct answers, participants' optimal guessed task 2 rank in the intermediate group is second or third.

* $p < 0.1$; ** $p < 0.05$ (one-sided (two-sided) Mann–Whitney U test in the weak and strong (intermediate) performance groups).

Table 7 Entry Rates and Average Guessed Task 2 Ranks of Men and Women in Generation 2 Depending on the Advice Received

Performance group	Weak		Intermediate		Strong	
	Men	Women	Men	Women	Men	Women
Entry rates						
Advice “piece rate” (%)	29 (4 of 14)	31 (5 of 16)	43 (9 of 21)	15* (3 of 20)	100 (2 of 2)	67 (2 of 3)
Advice “tournament” (%)	0 (0 of 1)	0 (0 of 1)	100 (8 of 8)	43** (3 of 7)	100 (10 of 10)	89 (8 of 9)
Guessed task 2 ranks						
Advice “piece rate”	2.64 (1.0)	3.06 (0.7)	1.71 (0.8)	2.25 (0.7)**	1.00 (0.0)	1.33 (0.6)
Advice “tournament”	2.00 (0.0)	2.00 (0.0)	1.00 (0.0)	2.14 (0.4)***	1.00 (0.0)	1.22 (0.4)

Note. Bold men–women value pairs are statistically significantly different.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ (two-sided Fisher's exact test for entry rates, two-sided Mann–Whitney U test for guessed ranks).

enter even slightly more than men if they receive advice ($p = 1.000$, two-sided Fisher's exact test). The effects of advice in the strong- and weak-performing groups are as expected; however, the emergence of a gender gap in the intermediate performance group is a surprise. To find an explanation for the gender differences in tournament entry in generation 2, we look at men's and women's reactions to the advice they receive.

4.3.3. Reactions to the Advice Received. First of all, we confirm that the two types of advice are equally spread between men and women in all three performance groups: Among weak performers in generation 2, 6.7% (1 of 15) of men and 5.9% (1 of 17) of women receive the recommendation to enter the competition. For intermediate performers, the corresponding rates are 27.6% (8 of 29 men) and 25.9% (7 of 27 women), and for strong performers they are 83.3% (10 of 12 men) and 75% (9 of 12 women). None of these differences are statistically significant.

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. We find no gender difference in reaction among the weak or strong performers. Table 7 summarizes the entry rates and average guessed ranks of men and women separately for each performance group and depending on the advice received. Overall, the entry rates for men are larger than for women for both pieces of advice: With the advice “piece rate,” 41% (15 of 37) of men and 26% (10 of 39) of women enter the tournament; with the advice “tournament,” 95% (18 of 19) of men and 65% (11 of 17) of women do so. The difference is significant only for the advice “tournament” ($p = 0.037$, two-sided Fisher's exact test). Testing for gender differences in the reaction to advice within performance groups reveals that the gender gap in the reaction to advice depends on the performance level: There is no gender gap in the weak-performing group with the advice “piece rate” or the strong-performing group with the advice “tournament” ($p > 0.474$ for the two

groups separately, two-sided Fisher's exact test), but there is one in the intermediate group ($p = 0.026$ for the advice “tournament” and $p = 0.085$ for the advice “piece rate,” two-sided Fisher's exact test).²⁰

Interestingly, the finding that, among intermediate performers, women are more reluctant in following advice to enter the competition, whereas men are more reluctant in following 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 2 need, on average, 20.9 (21.0) seconds to reach a decision. Men who receive the advice “piece rate” need, on average, longer for their decision (23.4 s) than men who receive the advice “tournament” (16.1 s), and also longer than women who receive the advice “piece rate” (19.9 s). For women, we find the reverse: Upon receiving the advice “tournament,” they need longer to make up their mind (23.6 s) than if they receive the advice “piece rate” (19.9 s), and also longer than men who receive the advice “tournament” (16.1 s). The difference in response times between men who receive the advice “piece rate” and men who receive the advice “tournament” is significant ($p = 0.043$, two-sided Mann–Whitney U test). The other time differences are not significant.²¹ The time differences become larger for women if we restrict the analysis on the intermediate performers.

²⁰ We also find gender differences in the reaction to the reasons provided to support the advice given. If the advice “tournament” is supported by emphasizing potentially higher earnings/encouraging to trust in one's own ability, men enter the competition more often than women ($p < 0.041$, two-sided Fisher's exact test), particularly in the intermediate group ($p < 0.055$, two-sided Fisher's exact test).

²¹ Variances of response times are generally quite high.

Comparing the self-assessment of men and women in the three performance groups for each type of advice with the corresponding entry decisions indicates that self-confidence is an important reason for the entry decision. Table 7 summarizes the average guessed rank for task 2 performance of men and women depending on the advice they receive. Similar to the tournament entry decision where men show higher entry rates, men are more confident than women in generation 2 independent of the advice they receive, i.e., their guessed rank is lower. With the advice “piece rate,” men (women) guess a rank of 2.03 (2.51); receiving the advice “tournament,” the guessed rank of men (women) is 1.05 (1.65) on average ($p < 0.023$ for both types of advice separately, two-sided Mann–Whitney U test). Analyzing the self-assessment by performance group, we find that men are significantly more confident than women in the intermediate group independent of the advice they receive ($p < 0.030$ for each type of advice, two-sided Mann–Whitney U test). Although there is a slight difference among strong-performing participants who are advised to enter the competition ($p = 0.125$, two-sided Mann–Whitney U test), men and women in the weak-performing group do not differ in their relative self-assessment ($p = 0.239$, two-sided Mann–Whitney U test), which mirrors the results of the effect of advice on entry decisions.

Summarizing the last three subsections, we confirm significant improvements in self-selection; in particular, strong-performing women enter significantly more and weak-performing men enter significantly less often. We have seen that the persistence of the gender gap in tournament entry is due to a gender gap in following the advice received among intermediate performers: whereas women more often follow the advice “piece rate,” men more often follow the advice “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. Although we have shown that advice indeed improves the efficiency of the selection process (strong-performing individuals enter more, weak-performing individuals enter less), the gender gap in entry is, at least on the surface, unchanged. 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 and women to the advice they receive.

Overall, our results suggest that advice significantly increases efficiency in the entry behavior, although it may not be suitable to overcome the gender gap in tournament entry entirely. Advice (given equally to women and men or particularly to strong-performing women)

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 is possible because, as we showed, advice 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 to those found by others, we find that among strong-performing women, the increase in entry through advice (more than 100%) is roughly the same as through the affirmative action “quota” (Balafoutas and Sutter 2012), and more than through the provision of relative performance feedback (Wozniak et al. 2014). With advice, the entry rate of strong-performing men increases (insignificantly) from 80% to 100%, whereas with the quota and with relative performance feedback, high-ability men even reduce their entry rates slightly. Comparing the impact on the entry decisions of weak-performing men, we see a decrease in 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. We can conclude that advice—especially the advice “tournament” for strong-performing participants—improves the self-evaluation.

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 experiment). The profit can be monetary and nonmonetary: In a firm, the senior can profit from a high-ability employee in the future through building up a good relationship 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, more experienced relatives and friends might benefit from giving advice because they care about the well-being of younger family members and friends and feel good for giving appropriate advice. Building up a reputation can play a role in the professional and personal environment.

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-form chat, would have an even higher potential for efficiency gains. Our design can easily be extended to allow for free-form communication, as, for example, in Brandts and Cooper (2007) and Brandts et al. (2012).

Additionally, analyzing the roles of the intensity of the interaction between advisor and advisee, the repetition of advice, and the willingness of advisors to provide advice would be interesting questions. Other naturally arising questions in this context are how the results would be if only (high-ability) women received advice, whether individuals solicit advice, and, if so, from whom they ask advice if it is an option (maybe even making it costly to obtain it).

Acknowledgments

The authors acknowledge the financial support of the Antoni Serra Ramoneda Research Chair (Universitat Autònoma de Barcelona—Catalunya Caixa); the Spanish Ministerio de Economía y Competitividad [Grant ECO2011-29847-CO2-01]; and the Generalitat de Catalunya [Grant 2009 SGR 820]. Valeska Groenert gratefully acknowledges support from the Spanish Ministerio de Ciencia e Innovación [Grant “Consolidated Group-C” ECO2008-04756] and the European Regional Development Fund (FEDER).

Christina Rott will be at Maastricht University effective September 1, 2014.

References

- Athey S, Avery C, Zernsky P (2000) Mentoring and diversity. *Amer. Econom. Rev.* 90(4):765–786.
- Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* 335(6068):579–582.
- Balafoutas L, Kerschbamer R, Sutter M (2012) Distributional preferences and competitive behavior. *J. Econom. Behav. Organ.* 83(1):125–135.
- Bettinger E, Baker R (2011) The effects of student coaching in college: An evaluation of a randomized experiment in student mentoring. Working Paper 16881, National Bureau of Economic Research, Cambridge, MA.
- Blau FD, Currie JM, Croson RTA, Ginther DK (2010) Can mentoring help female assistant professors? Interim results from a randomized trial. *Amer. Econom. Assoc.* 100(2):348–352.
- Bonaccio S, Dalal RS (2006) Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organ. Behav. Human Decision Processes* 101(2):127–151.
- Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *J. Econom. Behav. Organ.* 81(2):542–555.
- Brandts J, Cooper D (2007) It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *J. Eur. Econom. Assoc.* 5(6):1223–1268.
- Brandts J, Charness G, Ellman M (2012) Let's talk: How communication affects contract design. CESifo Working Paper 3883, Institut d'Anàlisi Econòmica (CSIC), Bellaterra, Spain.
- Cason TN, Masters WA, Sheremeta RM (2010) Entry into winner-take-all and proportional-prize contests: An experimental study. *J. Public Econom.* 94(9–10):604–611.
- Celen B, Kariv S, Schotter A (2010) An experimental test of advice and social learning. *Management Sci.* 56(10):1687–1701.
- Charness G, Rustichini A, van de Ven J (2011) Self-confidence and strategic deterrence. Discussion Paper 11-151/1, Tinbergen Institute, Amsterdam.
- Chaudhuri A, Schotter A, Sopher B (2009) Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice. *Econom. J.* 119(534):91–122.
- Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Literature* 47(2):1–27.
- Dargnies MP (2012) Men too sometimes shy away from competition: The case of team competition. *Management Sci.* 58(11):1982–2000.
- Dohmen T, Falk A (2011) Performance pay and multi-dimensional sorting: Productivity, preferences, and gender. *Amer. Econom. Rev.* 101(2):556–590.
- Ertac S, Szentes B (2011) The effect of information on gender differences in competitiveness: Experimental evidence. Working Paper 1104, Koç University, Istanbul, Turkey.
- Ewers M (2012) Information and competition entry. Discussion Paper 6411, Institute for the Study of Labor (IZA), Bonn, Germany.
- Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.
- Flory JA, Leibbrandt A, List JA (2010) Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions. Working Paper 16546, National Bureau of Economic Research, Cambridge, MA.
- Gino F (2008) Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organ. Behav. Human Decision* 107(2):234–245.
- Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77(5):1637–1664.
- Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118(3):1049–1074.
- Greiner B (2004) The online recruitment system ORSEE 2.0—A guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne, Cologne, Germany.
- Grosse ND, Riener G (2010) Explaining gender differences in competitiveness: Gender-task stereotypes. Jena Economic Research Paper 2010-017, University of Jena, Jena, Germany.
- Gupta ND, Poulsen A, Villeval MC (2013) Gender matching and competitiveness: Experimental evidence. *Economic Inquiry* 51(1):816–835.
- Iriberrí N, Rey-Biel P (2013) On women's underperformance in competitive environments: When and why. Working paper, University of the Basque Country, Bilbao, Spain.
- Iyengar R, Schotter A (2008) Learning under supervision: An experimental study. *Experiment. Econom.* 11(2):154–173.
- Möbius MM, Niederle M, Niehaus P, Rosenblatt TS (2013) Managing self-confidence: Theory and experimental evidence. Working paper, Microsoft Research, Cambridge, MA.
- Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122(3):1067–1101.
- Niederle M, Vesterlund L (2011) Gender and competition. *Annual Rev. Econom.* 3(1):601–630.
- Niederle M, Segal C, Vesterlund L (2013) How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Sci.* 59(1):1–16.
- Reuben E, Sapienza P, Zingales L (2014) How stereotypes impair women's career in science. *Proc. Natl. Acad. Sci. USA* 111(12):4403–4408.
- Reuben E, Rey-Biel P, Sapienza P, Zingales L (2012) The emergence of male leadership in competitive environments. *J. Econom. Behav. Organ.* 83(1):111–117.

- Rodriguez-Planas N (2010) Mentoring, educational services, and economic incentives: Longer-term evidence on risky behaviors from a randomized trial. Discussion Paper 4968, Institute for the Study of Labor (IZA), Bonn, Germany.
- Schotter A (2003) Decision making with naive advice. *Amer. Econom. Rev.* 93(2):196–201.
- Schotter A, Sopher B (2003) Social learning and coordination conventions in inter-generational games: An experimental study. *J. Political Econom.* 111(3):498–529.
- Schotter A, Sopher B (2007) Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games Econom. Behav.* 58(2):365–393.
- Shurchkov O (2012) Under pressure: Gender differences in output quality and quantity under competition and time constraints. *J. Eur. Econom. Assoc.* 10(5):1189–1213.
- Steele CM (1997) A threat in the air: How stereotypes shape intellectual identity and performance. *Amer. Psychologist* 52(6):613–629.
- Sutter M, Glätzle-Rützler D (2014) Gender differences in the willingness to compete emerge early in life and persist. *Management Sci.* Forthcoming.
- Wozniak D, Harbaugh WT, Mayr U (2014) The menstrual cycle and performance feedback alter gender differences in competitive choices. *J. Labor Econom.* 32(1):161–198.