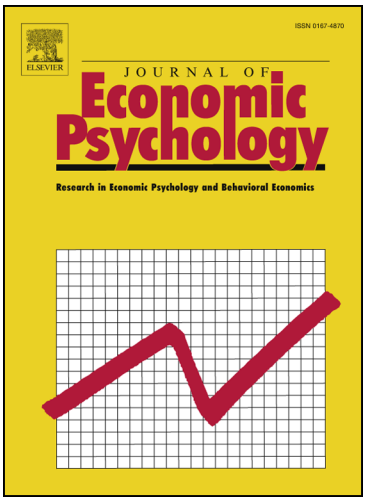
Journal Pre-proofs



Advice from Women and Men and Selection into Competition

Jordi Brandts, Christina Rott

PII: S0167-4870(20)30092-1
DOI: https://doi.org/10.1016/j.joep.2020.102333
Reference: JOEP 102333

To appear in: *Journal of Economic Psychology*

Received Date: 31 March 2020
Revised Date: 15 October 2020
Accepted Date: 20 October 2020

Please cite this article as: Brandts, J., Rott, C., Advice from Women and Men and Selection into Competition, *Journal of Economic Psychology* (2020), doi: https://doi.org/10.1016/j.joep.2020.102333

ADVICE FROM WOMEN AND MEN AND SELECTION INTO COMPETITION[◊]

October 2020

Jordi Brandts[*], Christina Rott[**]

ABSTRACT

Advice processes are omnipresent in our professional and private lives. We use a laboratory experiment to study how gender and gender matching affect advice giving and how gender matching affects advice following about entry into a real-effort tournament. For advice giving we find that women are less likely than men to recommend tournament entry to advisees than are intermediate performers. Furthermore, women maximize less often the expected earnings of advisees than intermediate performers. For advice following we find that men enter the tournament significantly more often than women in the intermediate-performance group do. Gender matching does not seem to affect advice giving or following. Overall, when it is less clear what the better advice or decision is, gender differences emerge. These results are consistent with findings in other areas that document that gender differences emerge in situations that are more ambiguous.

KEYWORDS: experiments, advice, gender gap in competitiveness

JEL CLASSIFICATION NUMBERS: C91, J08, J16

[◊] We thank Marco Castillo, Catherine Eckel, Ragan Petrie, Lise Vesterlund, Alistair Wilson, the seminar attendants at the University of Pittsburgh, Texas A&M University, KIT Karlsruhe, and RWTH Aachen University, the participants at the Barcelona GSE Summer Forum 2015, the ESA Conference 2015, the ESA North American ESA meeting 2016, and IMEBESS 2017 for very helpful comments on how to improve the paper. The authors gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (Grant: ECO2017-88130), the Severo Ochoa Program for Centers of Excellence in R&D (SEV2015-0563) the *Generalitat de Catalunya* (Grant: 2017 SGR 1136) and the Antoni Serra Ramoneda (UAB – Catalunya Caixa) Research Chair. The data are available at https://sites.google.com/site/christinaerott/.

[*] Instituto de Análisis Económico (CSIC) and Barcelona GSE. Address: Campus UAB, 08193 Bellaterra (Barcelona), Spain. Email: jordi.brandts@iae.csic.es (corresponding author).

[**] Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam, Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: c.e.rott@vu.nl.

1. INTRODUCTION

Increasing women's representation in top-level jobs is one of the main goals of gender equality policies in many countries. Identifying the causes of their current under-representation is a crucial input for the design of policies that can change this situation. There is now an established strand of experimental research (starting with Niederle and Vesterlund, 2007 and Croson and Gneezy, 2009) that studies this issue under controlled conditions where the decision to enter a real-effort tournament is used as a vehicle to study women's attitudes towards competing for high-ranking jobs. The main result that comes out of this line of research is that, compared to men, 'women shy away from competition' and that they underestimate themselves. Niederle (2016) reports that this result has been replicated many times, for a newer replication in China see Carlsson et al. (2020). Buser et al. (2014), Buser et al. (2017), and Reuben et al. (2017) have shown that this laboratory measure of competitiveness is significantly correlated with real world outcomes such as career choices and income in the labor market.[1]

Gender differences in competitiveness may have a basis in evolution as proposed, for example, in Bateman (1948) and Trivers (1972).[2] However, this possibility does not mean that, in humans, competitiveness is unalterable. In particular, information and communication can be powerful tools to change attitudes towards competitiveness. In a previous study (Brandts et al., 2015), we reported on the results of a laboratory experiment in which we analyzed how advice affects men's and women's entry into tournaments. In that study we analyzed women's and men's reaction to advice in a setting in which participants did not know the gender of their advisor or advisee. Overall, we found that advice improved the entry decision of subjects, in that forgone earnings due to wrong entry decisions went significantly down, but that, at the same time, an overall competition gender gap persisted.

In this paper we address two main research questions: first, we analyze how women and men give advice. To our knowledge, this is the first study on the effect of gender and gender matching on advice *giving*. Second, we study the other dimension of the advice process when advisors and advisees know each other's gender; that is, advice following focusing on the effect of gender matching. We compare all four cases of gender matchings, in which advisors and advisees of the same and opposite gender can be matched (i.e., female advisors with female

[1] Apart from competitiveness, risk-taking is the other aspect of behavior for which important gender differences have been documented. For a recent study see Friedl et al. (2020).

[2] A different issue is whether aversion to competition can be related to models of social preferences. On this issue see the recent study by Dasgupta et al.(2019).

advisees and male advisors with male advisees as well as female advisors with male advisees and male advisors with female advisees).

One of the reasons why women are challenged by competitive environments is their lower self-evaluation compared to men.[3] This is a particular type of gender-specific perception bias that needs to be addressed. Actively promoting that people, in particular women, who face a competitive situation receive advice from others who have experience with the situation is a "soft and low-cost" policy intervention. It relies more on raising the awareness of women's internal obstacles to correct self-assessment through a process in which more and more people become aware of the bias and contribute to correcting it than on the change of explicit institutional rules, like the introduction of affirmative action measures. Advice is one of the social mechanisms through which women's low self-evaluation bias can be corrected.

In our set-up advice consists in recommending choosing either a competitive (tournament) or a non-competitive payment scheme (piece rate). The advisor can furthermore add one or more pre-formulated reasons for the recommended payment scheme relating to preferences for competition, self-confidence, and earnings risk. All communication takes place through the computer. The gender of the matched person is revealed explicitly at the beginning of the advice stage. This information is displayed on the computer screen, without any information about the identity of the partner. Given this very stylized design it is clear that our focus can only be on the 'pure effect' of knowing the other person's gender without allowing for a more intense interaction involving, for example, free-form and face-to-face communication. We see this as a necessary step in understanding how social interaction can lead to more informed decisions vis-à-vis entry into competition.

The first dimension we focus on is on how advice *giving* interacts with *gender and gender matching*. In particular, we study whether women and men differ in the advice they give, whether one gender's advice is better and whether the advisee's gender affects advice giving. We see two reasons for being interested in gender effects in how advice is given. First, it might have direct practical implications for advisees as to whom to ask for advice. Second, it will yield insights into another dimension of gender differences in a competitive environment. Focusing on the *advisor*'s choices is a novel and important contribution to the extensive literature on advice and mentoring processes in competitive settings, see for instance Schotter (2003), Blau et al. (2010), Bettinger and Baker (2014), and Eskreis-Winkler et al. (2019).

[3] Jung and Vranceau (2020) analyze gender effects in the relation between competition and well-being in France and Korea.

The second dimension we study is advice *following* in the sense of how *gender matching* affects the reaction to the particular piece of advice received. The effect of gender on advice following has been analyzed in a gender-blind setup in Brandts et al. (2015). Here we focus on the effect of gender matching on advice following. We first look at the overall impact of advice with gender matching and compare it to the gender-blind setup in Brandts et al. (2015). We then focus on analyzing whether, to encourage a potentially high-performing woman to enter a competition, the advice from another woman is more effective than that from a man. Similarly, in convincing a potentially low-performing man not to make the socially wasteful decision of entering a competition it is possible that advice from a man has a bigger impact. Note that studying advice following consists here in going into some of the details of the advice process, beyond the overall impact of advice.

With respect to advice giving, we have two results. First, women are less likely than men to recommend tournament entry to advisees that are intermediate performers, regardless of the advisee's gender. Second, women maximize less often the advisee's expected earnings of advisees that are intermediate performers, again, regardless of the advisee's gender. In summary, for advice giving we find gender effects, but no gender-matching effects and, interestingly, they pertain to advisees that are intermediate performers. Women are more cautious in advice giving than men but only when advising intermediate performers. When it is less clear whether an advisee should be encouraged to enter the tournament or not, women recommend tournament less frequently and potentially hurt advisees economically by doing so.

With respect to the impact of gender matching on advice following, we find that advice improves entry decisions, both with male and female advisors, but only significantly for the latter case. However, the overall gender gap in tournament entry persists, both with male and female advisors because, in the intermediate-performance group with advice, men enter the tournament significantly more often than women do. The overall effect of advice is robust to knowing the gender of the counterpart and the gender differences in advice following are even more pronounced when the gender of the advisor is revealed. Men tend to refuse to follow the received advice, in particular the recommendation *not* to enter the tournament, when they know the advisor's gender. By contrast, women follow the received advice. Overall, the paired advisor's gender does not seem to play a role. Note that gender differences in tournament entry and advice following pertain, like the results for advice giving, to the intermediate-performance group. When it is less clear what the better decision is, then gender differences emerge.

2. BACKGROUND ON THE EFFECTS OF ADVICE AND GENDER MATCHING

The literature on advice and mentoring programs is broad and has focused mainly on the *impact*, i.e., on the effect of advice, mentoring, and education on advisees' behavior. Schotter (2003) provides a summary of the experimental literature on naïve advice showing consistent support for improved decisions with advice. Individuals give naïve advice to a next (non-overlapping) "generation" of individuals if they have simply made the same decision themselves prior to giving advice (as opposed to informed expert advice). In their field experiment, Blau et al. (2010) find that mentoring programs improve the performance of female assistant professors. In contrast to the vast majority of the literature on advice, in this paper we study the advice process and in particular the advisor's decision. To our knowledge, we are one of the first to analyze in detail the advice *giving* behavior. Closely related is the field experiment by Tungodden (2019) on parents' advice to their children in Norwegian high schools. The results show that parents make more competitive choices for boys than for girls, that parents are more responsive to boys' than to girls' ability and that fathers are more likely than mothers to enter their children into competition.

Similarly to overall gender differences in competitiveness, the possible effect of gender matching on behavior can be related to findings in evolutionary psychology. In their experimental study on gender matching and bargaining, Sutter et al. (2009) write that: "Trivers' (1972) theory of parental investment and sexual selection predicts that, as a consequence of the competition for a mate, rivalry and aggression in behavior should be more intense within the same sex (intra-sexual competition) than against the opposite sex (intersexual competition)." It is not obvious that this statement applies directly to all aspects of gender matching in humans, but it does provide a solid motivation for investigating such effects. In addition, we think that paying attention to the effects of gender matching is rather natural in our context. If women and men react differently to the choice between competition and the absence of it as well as to advice about this choice, then it seems plausible that being advised by women or men may also have an impact.

There are some previous experimental studies on how gender matching affects behavior in various environments. The results of Sutter et al. (2009) show that competition, retaliation, and low efficiency are more frequent when bargaining partners have the same gender than when they have the opposite. In an experimental study of competitiveness, Datta Gupta et al. (2013) report that the gender of the co-participant directly influences men's choices (men compete less against other men than against women), but only when the gender information is made

sufficiently salient. Brandts et al. (2020) study gender differences in preferences for performance ranking by another person. They find no gender differences when the ranker is a woman. By contrast, with a male ranker, men have a much stronger desire to be ranked than women do. Shurchkov and van Geen (2019) study whether men and women favor different incentive schemes for others and find that women are less likely to assign competitive incentives to others, particularly when their gender can be observed. Relatedly, Price (2012) reports that, when given information about worker ability, male managers choose the tournament significantly less often for a female worker, but that when no information about worker ability is given to the manager, there is no difference in compensation choice for the worker. Beugnot and Peterlé (2020) use laboratory data to study the effects of gender matching in job referrals. Their data suggest that women tend to favor women when choosing a job candidate, whereas men do not attach much importance to the gender of potential candidates.

There are also some studies about the effects of gender matching using observational data. Bettinger and Long (2005), Hoffmann and Oreopoulos (2009), and Carrell et al. (2010) provide evidence that the gender matchings of instructors and students affects students' outcomes, in particular those of female students. Gerdes and Gränsmark (2010) study behavior in chess tournaments and find that men choose more aggressive strategies against female opponents even though such strategies reduce their winning probabilities. De Sousa and Hollard (2016) also study chess tournaments and find that, when playing against men, women suffer a systematic handicap that prevents them from reaching top positions in the chess hierarchy. Card et al. (2020) compare the recommendations of male and female reviewers in assessing male and female-authored papers in four major economics journals. This paper studies an issue close to the one we study. It analyzes advice to editors and authors, controlling for gender matchings.

Finally, in education economics, the question of gender matching has been addressed extensively. Though evidence at the primary and secondary school level draws a mixed picture of the interaction of the student's and the teacher's gender,[4] there is some evidence for a positive effect of same-gender matching of student and teacher. In particular, female students seem to perform better when being taught by a female instructor at the post-secondary level (e.g., Bettinger and Long, 2005; Hoffmann and Oreopoulos, 2009; Carrell, Page, and West, 2010). In our setup, advisors are not role models as teachers and instructors (or mentors) in the before-mentioned studies can be. Whether the evidence on gender matching translates into the advice

[4] See for instance Dee (2004, 2005, 2007), Ehrenberg et al. (1995) Nixon and Robinson (1999), Holmlund and Sund (2008), Carrington et al. (2008), and Lahelma (2000).

process in a more controlled environment remains thus an open question and will be addressed in this study.[5]

3. EXPERIMENTAL DESIGN AND RESEARCH QUESTIONS

We first describe the basic experimental 'two-generation' design with regard to the choice of participating in a competition and then turn to the specifics of the advice part of the experiment and to some further information on the design. Finally, we address the specific research questions we ask.

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 Brandts et al. (2015) which extended the design of Niederle and Vesterlund (2007). The experiment took place at the UAB in Barcelona in two adjacent computer rooms, separated by a sound-proof glass wall. Upon arrival, subjects were divided randomly into two room groups of equal size. Participants in one room have the role of advisors, and those in the other room the role of advisees, but they do not learn about their roles until later in the experiment when the advice stage begins. The separation of the two rooms by the glass window makes it possible for participants to see that the participants in the other generation really exist.

Participants in both groups go through the same eight decisions in the same order: three real-effort tasks, two entry decisions, and two self-evaluations. There is a difference in the timing in which the groups go through their tasks. See Table 1 for a timeline of the tasks of the two generations. As illustrated in Table 1, advisors begin to make decisions 15 minutes earlier while the advisees wait. This waiting period was necessary to ensure that advisors and advisees reach the advice stage at roughly the same time. During this waiting period, advisees are not yet informed about the content of the experiment because we wanted to ensure that the waiting period had no effect on the choice of the compensation scheme in task 3.

Advisors	Advisees
Task 1: Five-minute addition task - Piece rate (€ 0.5)	

[5] We see advice as a particular form of communication. For a survey of laboratory studies on the effects of communication see Brandts et al. (2019).

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 Self-evaluation, task 1 and 2 (€ 1 per correct guess) Performance feedback, own group	Task 1: Five-minute addition task - Piece rate (€ 0.5) Task 2: Five-minute addition task - Tournament (€ 2, winner takes all)
*One advisor randomly matched with one advisee (**gender (not) revealed**)*	
Receive advisee's performance info *Give advice (50% of the advisee's task 3 earnings)* *Choose up to three reasons (preference for competition, confidence, risk of earnings)*	*Send own performance info* *Receive advice* *Receive up to three reasons (preference for competition, confidence, risk of earnings)*
	Task 3: Five-minute addition task - Selection of compensation scheme Task 4: Selection of compensation scheme task 1 Self-evaluation, task 1 and 2 (€ 1 per correct guess) Performance feedback, own group

***Table 1**: Timeline of tasks and compensation scheme in the experiment.*

In the real-effort tasks, 1-3, 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 supplementary material). The three real-effort tasks differ in the circumstances under which the participants perform them. In task 1, subjects first perform the task under the piece payment scheme and in task 2 under the tournament payment scheme. Under the piece rate payment scheme subjects receive €0.50 for each correct sum. For the tournament payment scheme, subjects are matched in groups of four (two women and two men seated in the same computer row), and only the group member with the best performance receives €2 for each correct sum, while the other group members do not earn anything.

In task 3 participants have to first decide whether the piece rate or the tournament payment scheme will be applied to their performance in the addition task. 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 task 3 if she solves more problems correctly than each of her group members in task 2 does. 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.

For advisors the advice stage follows after they have made all their other decisions. For advisees the advice stage follows after they have completed tasks 1 and 2, that is, immediately before they have to choose the payment scheme for task 3. Both advisors and advisees know at the advice stage which decisions the members of the other generation have already made.

In task 4 subjects do not have to do the addition task but only have to make an entry decision. They have to decide whether to apply the competitive or the noncompetitive payment scheme to their (past) task 1 performance. Finally, subjects rank their performance in tasks 1 and 2 relative to the group members' performances on a scale from 1 (best) to 4 (worst), respectively. At the end of the self-evaluation task, each participant receives feedback on the task 1 and task 2 performances of all her group members.

3.2. Advice

The exact sequencing of the advice stage is as follows. Each advisee is randomly matched to exactly one advisor, and each advisor has only one advisee. At this point in time advisors and advisees are – in the gender matching case (GM) - informed about the gender of their counterpart. This was not the case in the gender-blind data set (GB), where advisors and advisees did not receive any information about one another's gender. Revealing the counterpart's gender in the gender-matching case is the only difference compared to the gender-blind situation.[6] The counterpart's gender is revealed in a natural, verbal way on the information and decision screens of the advice stage.[7] Note that, at this stage, advisors also have already received feedback on task 1 and task 2 performances of all their group members.

After the matching, the advisee sends information about his or her task 1 and 2 performances to his or her advisor. We had the advisee actively sending the information to the advisor to create a feeling of interaction between advisor and advisee, instead of having the computer automatically send the information to the corresponding advisor. Upon receiving this information, the advisor sends a message, telling the advisee whether he or she recommends entering the competition. The advisor is then asked to give the advisee reasons for the recommendation. We provide three pre-formulated reasons for each of the two possible recommendations ('tournament' or 'piece rate') from which the advisor can select as many as she wishes to.[8] After having received the advisor's recommendation and (possibly) reasons for this recommendation, the advisee decides whether to enter the competition in task 3.

[6] Also, the instructions are identical in the gender-blind and in the gender-matching case.

[7] Students at Spanish universities commonly call one another "chico" or "chica," approximately equivalent to "guy" or "girl." On the advisor's information screen, it says "Vas a dar consejo a un chico/ una chica." ("You are going to give advice to a guy/ girl.") with "a guy/ girl" shown in bold letters. On the advisee's information screen, it says "Un chico/ Una chica te va a dar consejo." ("A guy/ girl is going to give advice to you.") with "a guy/ girl" shown in bold letters. On the subsequent decision screens, we continue referring to the matched partner as "the guy/ the girl," but without highlighting it in bold letters. We felt that this would be the most natural way of revealing gender in our setup without risking that the gender information would be overseen.

[8] For the advice 'piece rate', the phrases are: (1) "Porque no es divertido competir con otros." / "Because it is not fun to compete with others." (preference for competition), (2) "Porque no deberías tener confianza de que te vaya bien." / "Because you should not be confident that you will succeed." (self-confidence), and (3) "Porque con la

Advisors in our experiment are no experts in the task, but they have experienced the situation once and have some information about it. Since advisors have received (after completing all their tasks) information feedback on task 1 and task 2 performance of all their group members, they have not only made an entry decision, but have also seen how people perform in the addition task in a small sample of four people. We chose a design with these features because we felt that this is a rather natural setup. Usually, a person who has previously participated in a competition task will have some idea about performance levels in that task, but does not have access to a large database on the matter. 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. In natural environments, the reward of advisors can be nonmonetary in form of building a reputation or in form of a good feeling because of giving good advice to somebody.

Each advisee knows that his or her advisor has just completed all tasks and that the advisor has information about task 1 and task 2 performances of the participants in his or her own group. However, the advisee does not know that his or her advisor is compensated for giving advice and the advisor does not know that the advisee does not know. We chose this option to eliminate the influence of social preferences on the advisee's entry decision. Advisees do know that their advisor has some informational advantage, but they need to trust that the advisor will advise them correctly.[9]

3.3. Group composition, procedures, and subject pool

Participants were allocated to fixed groups of four, composed of two women and two men. We made sure that participants were not aware of the fact that we controlled for the gender composition because the salience of this information might change people's behavior (Iriberri and Rey-Biel, 2017). Each group of four shared the same row in the computer laboratory and participants knew that their competitors were seated in the same row as them.

---

remuneración por unidad ganas algo seguro." / "Because with the piece rate you earn something for sure." (risk of earnings). For the advice 'tournament', the three phrases are: (1) "Porque es divertido competir con otros." / "Because it is fun to compete with others." (preference for competition), (2) "Porque deberías tener confianza de que te vaya bien." / "Because you should be confident that you will succeed." (self-confidence), and (3) "Porque en la competición puedes ganar mucho más." / "Because in the competition you can earn much more." (risk of earnings).

[9] In Wozniak et al. (2014) and Ewers (2012), the experimenters inform the participants about their relative performance with respect to the other group members in their own generation (or about the performance distribution in general as in Ewers, 2012) and the participants know that this information is correct. Our focus is somewhat different in that our interest is in studying the advice process as a whole and the effects of *human interaction* between advisors and advisees. We believe that this is an important dimension of the interaction that takes place at the workplace.

Subjects received a show-up fee of €5 plus €4 for completing tasks 1–4. Advisors were paid for giving advice and advisees were paid an additional €2 because they had to wait for approximately 15 minutes at the beginning of the experiment. At the end of the experiment, we chose one of tasks 1–4 at random and paid participants according to their performance in that task. Finally, we paid subjects for the self-evaluation task. On average, our participants earned €18.35.[10] The average duration of a session was 1 hour and 30 minutes, starting with reading aloud the general instructions and finishing after participants filled out a questionnaire and received their payment.

The experiment was conducted in January 2012 and December 2014. Subjects were recruited from a pool of subjects via the online recruitment system ORSEE (Greiner, 2004) and were mainly undergraduate students from UAB. Students in all departments at UAB were invited to subscribe to ORSEE via flyers distributed and posted on campus and through student mailing lists. The experiment was programmed and conducted with the experimental software z-Tree (Fischbacher, 2007). The UAB has a total of 50,000 students and our subject pool contains approximately 2,500 students. Thus, there is a very low likelihood that participants of the same group knew each other because all 2,500 students of the subject pool received the invitations for the sessions at the same time, and we assigned participants randomly to the role (advisor or advisee) and group.

Table 2 summarizes the characteristics of the data we use in this paper and the number of observations; we use two different data sets. In the *Gender Matching (GM*) data set, the advisor's and the advisee's genders are revealed to each other. In the sessions with *same gender,* each male (female) advisor is matched with a male (female) advisee. In the sessions with *opposite gender* each male (female) advisor is matched with a female (male) advisee. In the *Gender Blind (GB*) data, advisors and advisees are matched randomly and their genders are not revealed. We analyzed advice following–but not advice giving–with the GB data in Brandts et al. (2015). In this paper, we use the GM and the GB data to analyze the effect of gender and gender matching on advice giving. To analyze the effect of gender matching on advice following we use only the GM data. The number of observations in the GM data (100+112 advisors/ advisees) intended to match the number of observations in the GB data set (112 advisor/ advisees): Considering the size of the gender gap in competitiveness (80.0% of men and 58.9% of women entered into the tournament) and the effect size of advice on tournament entry of high-performing women (40.0% without advice and 83.0% with advice) in Brandts et

[10] On average, €3.48 was earned for the advice, with payoffs ranging from €0 up to €25.

al. (2015), we aimed at 54 women and 54 men for each role (advisor – advisee) and matching. Taking into account the effect of advice on tournament entry of intermediate- and high-performers (advice causes a gender gap of 36.4pp among intermediate-performing advisees, the gender gap of 40pp among high-performing advisors disappears with advice), we aimed at respectively 33 intermediate-performing women and men as well as 24-28 high-performing women and men for each role (advisor – advisee) and matching (power of 80% and 5% significance level).

Data sets	Session type	Characteristics	Observations
Gender blind (GB)	*No gender*	Gender of paired advisor/advisee not revealed, random pairing of advisor and advisee	112 advisors, 112 advisees
Gender matching (GM)	*Same gender*	Women (men) give advice to a woman (man), random pairing of advisor and advisee within gender	100 advisors, 100 advisees
	Opposite gender	Women (men) give advice to a man (woman), random pairing of advisor and advisee across gender	112 advisors, 112 advisees

***Table 2**: Overview of the data we use*

3.4. Research questions

Motivated by the reviewed literature on advice and gender matching, we focus on the following novel aspects of the advice process: First, we take a close look at how men and women *give* advice with respect to a competitive decision. Second, we take into account the advisor's *and* the advisee's gender in a tightly controlled environment. We also analyze whether the gender matching affects advice following. Our research questions are about the specifics of the advice process, with questions 1 and 2 pertaining to advice *giving* and question 3 to advice *following*. In the following, we formulate research questions as well as corresponding hypotheses.

1. How is advice *giving* affected by *gender* and *gender matching*? We take a close look at how men and women *give* advice with respect to a competitive decision and take into account the advisor's *and* the advisee's gender as a potentially important factors of social interaction. Both this research question 1 and research question 2 below have, to our knowledge, not been considered before. We cannot directly justify a hypothesis based on previous results. However, there is some related literature that is suggestive. There is evidence that women are more risk averse than men (with varying degree depending on the elicitation method, e.g.,

Filippin and Crosetto, 2016). Furthermore, when self-assessment is observable in a competitive environment, women seem to feel more shame for overestimating themselves than men (Ludwig, Fellner-Röhling, and Thoma, 2017). Both inclinations could translate into less confident advice giving by women. Hence, our hypothesis is that women will be less likely to give 'tournament' advice than men will. With respect to effects of gender matching, we do not have any a priori relevant information so that our hypothesis is that there will be no difference. We could imagine effects in both directions, i.e., that advisors are more likely to recommend the tournament to men (because advisors anticipate that men are more willing to take risk; e.g., Filippin and Crosetto, 2016) or to women (because advisors anticipate that women enter the tournament too little; e.g., Niederle and Vesterlund, 2007).

2. Does the *'quality'* of advice depend on *gender* and *gender matching*? We think of 'quality' in terms of discouraging low-performing participants from entering the tournament and encouraging high-performing participants to enter. The answer to this question may be important to suggest from whom men and women should seek advice. Assuming that both men and women are on average risk averse, but women more so, and that in addition women may feel more easily ashamed of overestimating themselves, our hypothesis is that women will give less often the 'tournament' advice than men. With respect to effects of gender matching we do not have any a priori relevant information so that our hypothesis is that there will be no difference.

3. How is advice *following* affected by *gender matching*? Again, different scenarios are possible: advisees might follow men's (women's) 'tournament' ('piece rate') advice less because they believe that men (women) recommend too often the 'tournament' ('piece rate'), see arguments for research question 1. Furthermore, receiving advice from someone of the same gender may create a common social identity and trust in the matched individual (see for instance, Tanis and Postmes, 2005; McEvily et al., 2006). We hypothesize that, given the stylized nature of our environment; gender matching will have no effect.

4. RESULTS

Throughout the results section, whenever we mention performance, we mean the number of correct sums. If not otherwise noted, to test for differences in the performance between subjects, we use two-sided Mann–Whitney U tests; to test for differences in advice giving and tournament entry, we use the two-sided Fisher's exact test. Throughout the paper,

the number of observations per test is shown in the percentage calculation in parentheses. For the regression analysis, we use linear probability models with robust standard errors.[11]

In our results presentation we distinguish results on advice giving and advice following. Section 4.1 presents our results on how gender and gender matching affect advice giving in the GM and the GB data sets (research questions 1 and 2). In section 4.2 about advice following we first summarize some general observations about the overall impact of advice following in the GM data comparing it with the impact in the GB data in Brandts et al. (2015). Subsequently we show whether gender matching in GM affects the reaction to the advice received (research question 3).[12]

4.1. Men's and women's advice giving

Advice giving was not studied in Brandts et al. (2015) and, hence, here we will use data from both the GB and the GM data set.[13] As mentioned above, when advising the next generation, advisors hold several pieces of information: the distribution of performance in task 1 and task 2 in their *own* group, the advisee's information on task 1 and task 2 performance as well as the advisee's gender. The first two pieces of information (performance) are provided in both the GB and GM data, the third piece of information (advisee's gender) is provided in the GM data only.

Advisors are likely to use all available pieces of information to formulate their advice. To take into account the diverse information environments of advisors, we separate advisees' performance levels into three *relative* performance intervals based on task 2 performance (see below). The rationale behind this classification is that advisors may condition their advice on advisees' performance in task 2, which is arguably a better predictor of performance in task 3 than task 1. The three intervals are the following: (1) the advisee's task 2 performance is lower than the second best performance in the advisor's *own* group (low). (2) The advisee's task 2 performance is between the 2nd (including) and the 1st best performance in the advisor's *own* group (intermediate). (3) The advisee's task 2 performance is at least as good as the 1st best

[11] Logit and Probit regressions with robust standard errors lead to very similar results.

[12] The analysis of the reaction to advice in the GB data set can be found in Brandts et al. (2015).

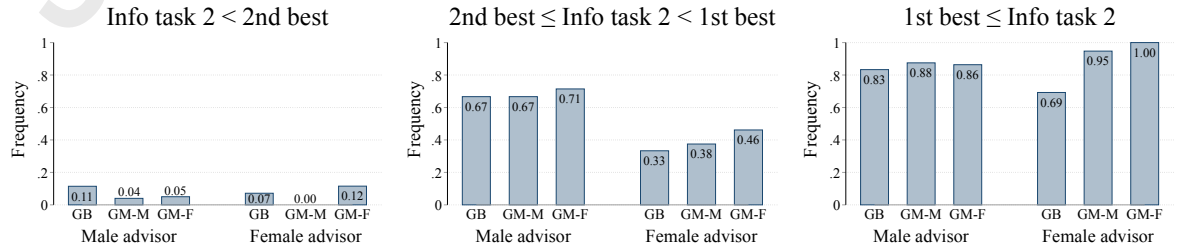
[13] To address the first research question pertaining to advisors, we use exclusively the advisors' observations in the GB and the GM data. To address the second research question pertaining to advisees, we use only the advisees' observations in the GM data Weighing the risks of being too liberal (false positives) and too conservative (false negatives), we think that correcting for multiple testing within each research question suffices: 3 performance groups for the analysis of gender differences in advice giving and following and (3 performance groups x 2 matching conditions) for the impact of gender matching on advice following.

performance in the advisor's *own* group (high).[14] Note that these intervals (low, intermediate, high) are defined by the advisee's task 2 performance *relative* to the 1st and 2nd best performance in the advisor's *own* group. A performance below the 2nd best performance is likely to be considered too weak for tournament entry.

We first present the results of two-sided Fisher exact tests and then the results of regression analysis. Aggregated over all advisors (across the two data sets) only 7.10% (11 of 155) recommend tournament entry if the advisee's performance is relatively low, with the share increasing to 51.32% (39 of 76) in the intermediate relative performance interval, and to 87.1% (81 of 93) in the strong relative performance interval. The better the advisee's performance relative to the performance in the advisor's *own* group the more likely advisors are to recommend tournament entry.

With respect to our research question 1, we find that men's and women's advice does not differ if the advisee's relative performance is either low (lower than 2nd best) or high (at least equal to 1st best). However, the result is different for the intermediate performer group. Women are significantly less likely to recommend tournament entry than men if the advisee's performance is at least equal to the 2nd best but weaker than the 1st best performance in the advisor's own group. The shares of advisors recommending the competition in the weak relative performance interval are 7.50% (6 of 80) for men and 6.67% (5 of 75) for women (p = 1.000). In the strong relative performance interval, 86.00 % (43 of 50) of men and 88.37% (38 of 43) of women recommend the tournament, again an insignificant difference (p = 0.767). For the intermediate relative performance interval 68.75% (22 of 32) of men advise tournament entry whereas only 38.64% (17 of 44) of women do so (p = 0.012).

The next step is to study the effects of gender matching. Figure 1 shows the proportion of male and female advisors who recommend tournament entry disaggregated for the three relative performance intervals, the advisor's gender (male or female) and the advisee's gender (gender blind, male, or female).



[14] The results do not change considerably if we include (or exclude) the limits of any of the three relative performance intervals.

***Figure 1**: Proportion of male and female advisors recommending 'tournament' entry (by advisee's task 2 performance info relative to the performance in the advisor's group, advisor's and advisee's gender; sample is the GB and GM data where GB = advisee's gender blind, GM-M = male advisee, GM-F = female advisee).*

In the left panel of Figure 1 (low relative performance interval), 4% to 11% of men recommend to enter the competition compared to 0% to 12% of women ($p > 0.621$; separately for gender blind, male advisee, and female advisee). In the right panel (high relative performance interval), the corresponding shares range from 83% to 86% for men and from 69% to 100% for women ($p > 0.533$; separately for gender blind, male advisee, and female advisee). In the panel in the middle (intermediate relative performance interval), 67% (6 of 9, gender blind), 67% (6 of 9, male advisee), and 71% (10 of 14, female advisee) of men recommend competition entry while only 33% (5 of 15, gender blind), 38% (6 of 16, male advisee) and 46% (6 of 13, female advisee) of women do so ($p < 0.253$; separately for gender blind, male advisee, and female advisee).

	Info task 2 < 2nd best				2nd best ≤ *Info task 2* < 1st best				1st best ≤ *Info task 2*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: *Advice 'tournament'*												
Female advisor	-0.011	0.008	-0.041	-0.036	-0.345**	-0.303**	-0.449*	-0.373	0.014	0.029	0.084	0.082
	(0.041)	(0.044)	(0.048)	(0.053)	(0.109)	(0.113)	(0.193)	(0.191)	(0.076)	(0.078)	(0.100)	(0.105)
Female advisee			-0.011	-0.013			-0.025	-0.042			0.003	-0.031
			(0.065)	(0.067)			(0.204)	(0.195)			(0.110)	(0.133)
*Female advisor * Female advisee*			0.091	0.093			0.225	0.217			0.038	0.050
			(0.091)	(0.094)			(0.290)	(0.284)			(0.111)	(0.120)
Advisor's entry task 3		0.076		0.027		0.135		0.244		0.144		0.115
		(0.048)		(0.061)		(0.113)		(0.151)		(0.074)		(0.098)
R-squared	0.060	0.079	0.140	0.143	0.213	0.229	0.217	0.267	0.021	0.061	0.049	0.082
Dependent variable: *Correct advice*												
Female advisor	0.044	0.039	0.043	0.045	-0.248*	-0.269*	-0.427	-0.458*	0.006	0.010	-0.173	-0.172
	(0.042)	(0.046)	(0.062)	(0.063)	(0.116)	(0.122)	(0.229)	(0.227)	(0.080)	(0.082)	(0.115)	(0.114)
Female advisee			0.047	0.046			-0.222	-0.217			-0.203	-0.239*
			(0.070)	(0.071)			(0.237)	(0.240)			(0.115)	(0.118)
*Female advisor * Female advisee*			-0.082	-0.080			0.488	0.494			0.398*	0.409*
			(0.105)	(0.108)			(0.331)	(0.336)			(0.160)	(0.159)
Advisor's entry task 3		-0.019		0.012		-0.065		-0.087		0.035		0.128
		(0.051)		(0.064)		(0.122)		(0.157)		(0.081)		(0.092)
R-squared	0.212	0.213	0.237	0.237	0.122	0.126	0.147	0.154	0.299	0.300	0.415	0.433
Control for *Info task 2, Info task 2 - task 1, 1st best and 2nd best performance task 2* (and *(Info task 2)^2* for dependent variable *Correct advice*)	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	155	155	92	92	76	76	52	52	93	93	68	68

Notes. *** p<0.001, ** p<0.01, * p<0.05. OLS regressions with robust standard errors in parentheses. Constant omitted. For the ease of result interpretation, OLS regression results are displayed. Logit and probit regressions lead to similar results. The dependent variable Advice "tournament" takes the value 1 for tournament and 0 for piece rate. The dependent variable Correct advice takes the value 1 for advice 'piece rate' ('tournament') if Info task 2 < 11 (> 10) and 0 otherwise. The sample is advisors of the GM and GB data sets in models (1), (2), (5), (6), (9), and (10) as well as the GM data set in models (3), (4), (7), (8), (11), and (12).

Table 3*: OLS regressions of advice giving by advisors and correctness of advice (sample is the GB and GM data).*

We use regression analysis to look closer at advice giving. We therefore aggregate observations across data sets to account for the relatively small number of observations in the subgroups. Table 3 shows the results of Linear Probability Model regressions (with robust standard errors) for advice giving. Logit and Probit regressions lead to similar results. In the upper part of Table 3, the dependent variable *Advice 'tournament'* takes the value 1 if the advisor's recommendation is 'tournament' and 0 if the advice is 'piece rate'. Regression models (1) – (4), (5) – (8), and (9) – (12) correspond to advisors in the low, intermediate, and high relative performance interval, respectively. The sample is advisors in data sets GB and GM in models (1), (2), (5), (6), (9), and (10) and advisors in the GM data set only in models (3), (4), (7), (8), (11), and (12).

As explanatory variables, we include the dummy variables for *Female advisor* and for *Female advisee*, which take the value 1 if the advisor (advisee) is a woman and 0 if he is a man, as well as some controls.[15] The interaction term of *Female advisor* and *Female advisee* captures whether women and men react differently to the advisee's gender. The variable *Female advisee* and the interaction term are thus included in models (3), (4), (7), (8), (11), and (12) with the GM data set only. We also add the advisor's own tournament entry decision (*Advisor's entry task 3*) to test whether the advisor's own choice (which is likely to be influenced by risk and competition preferences as well as confidence) is correlated with the advice they give.

The regression analysis shows significant gender differences if the task 2 performance information falls between the first- and second-best performance in the advisor's *own* group (models (5) and (6), $p < 0.01$), and no gender differences else (models (1), (2), (9), and (10)). In the intermediate performance group, women are on average more than 30 percentage points less likely to recommend tournament entry than men are. Regression analysis with GM data only include a dummy variable for the advisee's gender and an interaction term of the advisor's and the advisee's gender and confirm the same pattern.[16]

We do not find significant effects of gender matching on advice giving. In regression models (3), (4), (7), (8), (11), and (12) in Table 3, the coefficient estimates of *Female advisee*,

[15] We include as controls the information that advisees send to their advisor about their task 2 performance (*Info task 2*) as well as the difference between the task 2 and the task 1 information (*Info task 2 – task 1*) reflecting the improvement from task 1 to task 2. Finally, the information about the best as well as the second-best performance in task 2 in the advisor's *own* group of four (*1st best performance task 2* as well as *2nd best performance task 2*) are incorporated to check whether advisors take this information into account in a sensible way. The explanatory variables *Info task 2* and *Info task 2 – task 1* are positively correlated. However, *Info task 2* is a strong predictor of the advice "tournament" even though the two variables are correlated.

[16] The results for the GB and the GM data sets, models (5) and (6), remain significant after applying the Holm-Bonferroni correction for multiple testing (correcting for testing in three relative performance intervals). The gender difference in model (7) becomes insignificant after applying the Holm-Bonferroni correction.

(gender-matching effect on male advisors; $p > 0.817$) and the post-estimation tests of *Female advisee* and the interaction term (gender-matching effect on female advisors; $p > 0.160$) are all insignificant. The only exceptions are the coefficient estimates of *Female advisor* in models (7) and (8) which correspond to the intermediate relative performance interval. We highlight the answer to our research question 1:

Result 1: *Women are less likely than men to recommend tournament entry to advisees that are intermediate performers, regardless of the advisee's gender.*

This result qualifies our hypothesis above, since the gender difference we predicted pertains only to the group of intermediate performers. Again, for the case of intermediate performers, advisors have more leeway than in the two extreme cases to let themselves be affected by their own circumstances or biases.[17]

The second research question pertaining to advice giving is who gives better advice, men or women? Figure 2 complements Figure 1 in that, in addition to the share of advisors who recommend 'tournament', it also shows the share of advisors who recommend 'piece rate'. Furthermore, the darker colors represent advisors who choose the 'correct' piece of advice and the lighter colors the share of advisors who recommend the 'wrong' piece of advice. The advice 'piece rate' is considered to be correct if the advisee's task 2 performance information has less than 11 correct answers, otherwise it is considered wrong. In parallel, the advice 'tournament' is considered to be correct if the advisee's task 2 performance information has more than 10 correct answers and is otherwise considered wrong.[18] We use this threshold because the probability of winning the tournament is about 25% with 10 correct answers; a detailed explanation of the calculation can be found in Appendix B. This is when the expected payoff from choosing the piece rate is about the same as the expected payoff from choosing the tournament. The bars thus indicate the quality of advice in the sense of maximizing the advisee's expected payoffs.[19] The figure suggests that, as before, advice giving to the intermediate group needs to be analyzed more in detail.

[17] Given this qualification it is not clear whether the result is consistent with risk aversion or with issues of self-assessment. A more detailed analysis of the mechanisms behind this result is beyond the scope of this paper.
[18] Slight modifications (e.g., considering any piece of advice correct in case of ten correct answers) do not change the overall results.
[19] Note that this classification assumes expected earnings maximization which is also the case for the optimal rank and expected forgone earnings calculations.

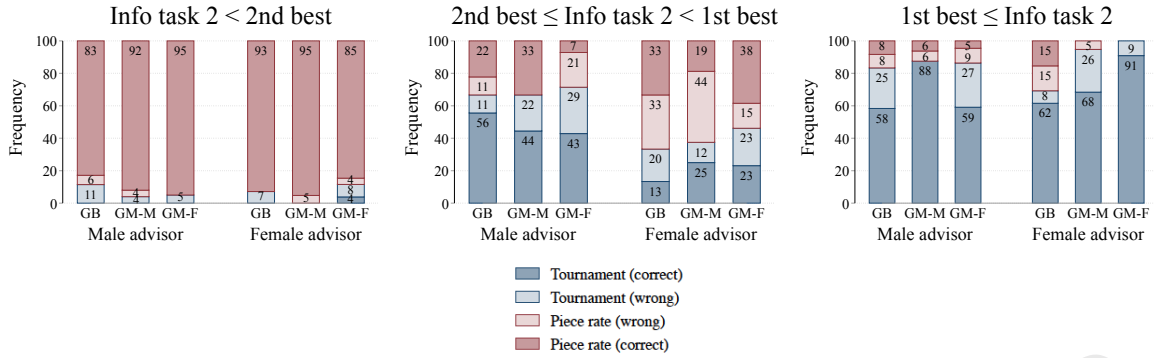


Figure 2: *Correctness of advisor's piece of advice (in per cent). Advice 'piece rate' ('tournament') is considered correct if the advisee's task 2 performance info is less (more) than 11 (10) correct answers, otherwise wrong (by advisee's task 2 performance info relative to the performance in the advisor's group, advisor's and advisee's gender; sample is the GB and GM data where GB = advisee's gender blind, GM-M = male advisee, GM-F = female advisee).*

In the lower part of Table 3, the dependent variable *Correct advice* takes the value 1 if the advisor's recommendation ('tournament' or 'piece rate') is correct as defined above and 0 otherwise. Regression models (1) – (4), (5) – (8), and (9) – (12) correspond to advisors in the low, intermediate, and high relative performance interval, respectively. The explanatory variables and the samples are the same as in the upper part of Table 3 (regression output with dependent variable 'tournament' advice).

From the light shaded parts in the bars in Figure 2, it can be seen that women tend to give more often the wrong piece of advice in the intermediate performance group (middle panel). Using the just described definition of quality, the regression analysis confirms that women's advice in the intermediate relative performance interval is at least 25 percentage points less likely to be correct (models (5) and (6)). In the GM data, women are even more than 40 percentage points less likely to maximize male advisees' expected earnings with their advice (negative coefficient estimate for *Female advisor*, models (7) and (8)). Assuming that the advice is followed perfectly, one could alternatively compare an advisee's expected forgone earnings with the given advice. Replacing the dummy variable *Correct advice* by the expected foregone earnings with the given advice leads to even more significant results reinforcing their validity.[20] The calculation of expected forgone earnings is explained in the subsequent section

[20] The coefficient estimate of the variable *Female advisor* becomes positive and significant at the 1% level in the medium-relative performance interval (models (5) – (8)). The estimates remain insignificant in the low and high relative performance intervals (models (1) – (4) and (9) – (12)). Regression results are available upon request. The gender difference in models (5) – (8) with the dependent variable *Correct advice* becomes insignificant after applying the Holm-Bonferroni correction (correcting for testing in three relative performance intervals). When using the advisee's expected forgone earnings as dependent variable measuring the quality of advice in regressions with the same exogenous variables as in Table 3, we find that the gender difference in the intermediate performance

4.2 and in more detail in Appendix B. Expected earnings are maximized by entering the tournament (choosing the piece rate) if an individual gives eleven or more (nine or less) correct answers. There is some evidence that men and women give better advice to advisees with the same gender,[21] but we would need more data to draw reliable conclusions. With respect to research question 2, the following result summarizes the way in which women's advice is worse than that of men:

Result 2: *Women maximize less often the advisee's expected earnings if the advisee's performance is not clearly low or high.*

For intermediate-performance advisees, women's advice is more cautious than men's advice.[22] This caution hurts women's advisees economically. These findings add to the evidence that it is in the intermediate-performance group where gender matters. When it is very clear what to do or advise – as in the case of low and high performance – men and women are very similar. However, when it is less clear then women are more reluctant than men to give the advice to enter the tournament. As for result 1, this difference is not affected by the gender of the advisee, i.e. there is no gender-matching effect.

4.2. Men's and women's advice following

As anticipated above in this section we first present some general observations with respect to women's and men's performance and tournament entry decisions. We then turn our attention to research question 3; that is, how the piece of advice affects advisees' competitive choices and how this depends on gender matching. As mentioned earlier, the data set for the analysis of the reaction to the received advice is GM only.

We start by verifying whether in the GM data set there are performance differences between men and women in tasks 1 and 2, i.e., in the absence of advice.[23] To do this we aggregate the data from the GM data set (424 observations from 212 advisors and advisees, respectively) because at this point there is no difference between the types of gender matching

group remains significant at the 5% level after applying the Holm-Bonferroni correction for simultaneously testing for three hypotheses.

[21] Merging the intermediate and high relative performance intervals and running the same regressions as in (7) and (8)/ (11) and (12), give highly significant coefficient estimates for the variables *Female advisor*, *Female advisee* (gender-matching effect on male advisors), and their interaction term as well as significant post-estimation tests of *Female advisee* and the interaction term (gender-matching effect on female advisors).

[22] This cautiousness may be related to risk-aversion, but we do not have independent information to back up this claim.

[23] Figure A1 in Appendix A shows men's and women's performance distributions in task 1 and task 2.

(same or mixed) and the randomly assigned role (advisor or advisee). Men solve on average 7.8 problems in task 1 and 9.8 in task 2. Women solve on average 7.3 problems in task 1 and 9.4 in task 2. The gender difference in performance is not statistically significant both for task 1 ($p = 0.603$) and task 2 ($p = 0.921$).[24] The absence of a significant gender difference in performance replicates the results reported in Brandts et al. (2015). We thus confirm a consistent subject pool across the GB and the GM data sets and find no difference in women's and men's performance in the math task – contrary to the widespread stereotype.[25]

The first main outcome variable is the average expected forgone earnings without and with advice from male (female) advisors. Expected foregone earnings are defined as the difference between expected earnings under the payment scheme the participant did not choose and expected earnings under the payment scheme she chose if a participant chose the (for her) inferior payment scheme. Otherwise, they are zero. The overall impact of advice as well as the calculation of the probability of winning and expected forgone earnings are described in detail in Appendix B. With advice, expected forgone earnings go significantly down, in particular for women.

However, the overall gender gap in tournament entry does not diminish with advice. Similarly we also replicate the gender gap without advice: 54.7% of men (58 of 106) and only 36.8% of women (39 of 106) enter the tournament ($p = 0.013$). Upon receiving advice from a male advisor, 64.0% (32 of 50) of men and 41.1% (23 of 56) of women enter the tournament ($p = 0.021$). After receiving advice from a female advisor, the rates are 60.7% (34 of 56) of men versus 36.8% (18 of 50) of women ($p = 0.012$). This is about the same as the gender gap without advice (54.7% of men versus 36.8% of women, $p = 0.013$). Both these results are consistent with those for the gender-blind data in Brandts et al. (2015).

[24] Performances in task 1 and task 2 are highly correlated and the same accounts for task 2 performance and the performance change from task 1 to task 2 (overall and separately for men and women). In task 3 men solve on average 10.0 problems correctly and women 9.6. Performance does not differ across gender ($p = 0.790$).

[25] Comparing those who will later be advisors with those who will later be advisees, we find insignificant differences for task 1 (and task 3), but not for task 2. In task 1, advisors solve 7.4 problems while advisees solve 7.8 (7.6) who will receive advice from a male (female) advisor ($p = 0.250$ and $p = 0.628$ for advisors vs. advisees matched with male and female advisors, respectively; $p = 0.505$ for advisees matched with male vs. female advisors). However for task 2, the performance levels are 9.0 correct answers for advisors vs. 10.1 for advisees with male advisors and 10.2 for advisees with female advisors ($p = 0.012$ and $p = 0.016$ for advisors vs. advisees matched with male (female) advisors, respectively; $p = 0.868$ for advisees matched with male vs. female advisors). Performance in task 3 is again insignificant ($p > 0.488$ for all three comparisons) which leads us to believe that the difference in task 2 performance is a somewhat random event. For the analysis presented in this paper, the difference is not of major importance because we will control for performance.

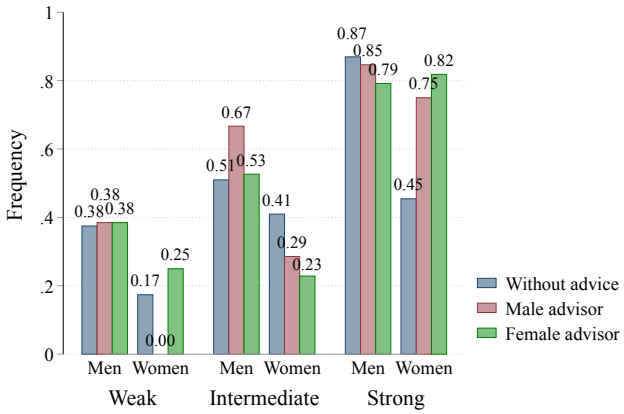


***Figure 3**: Proportion of men and women who enter the competition for a given performance group (task 2 performance, by advisor's and advisee's gender; sample is the GM data).*

The entry behavior at different performance levels explains the apparent contradiction between the reduction in forgone earnings and the persistence of the gender gap in tournament entry. To do this we divide participants into three groups, according to their performance in task 2: weak (26%), intermediate (52%), and strong performance (22%), see Table B1 in the Appendix B. Note that these performance groups are different from the relative performance groups in the previous section. Here, the intervals (weak, intermediate, strong) are defined by an individual's most likely task 2 performance rank. While participants in the weak and strong performance groups are most likely to be ranked fourth and first, respectively, and thus the optimal tournament choice is unambiguous, it is less clear for the intermediate performance group (ranked second or third). Figure 3 depicts the tournament entry in the different treatments and performance groups. Without advice, the gender gap is substantial among strong performers: whereas 87% (20 of 23) of strong-performing men enter the tournament only 45% (10 of 22) of women do so ($p = 0.005$). This changes with advice: More strong-performing women enter the competition after having received advice both from a man (75%, 15 of 20) or from a woman (82%, 9 of 11), thereby closing the gender gap among strong performers (compared to 85%, 11 of 13, of men with male advice, $p = 0.676$; and 79%, 19 of 24, of men with female advice; $p = 1.000$).

The reason why the overall gender gap does not disappear with advice seems to be due to the effect of advice on the group with intermediate performance. Women in the intermediate group are less likely to enter the tournament with advice, whereas men are more likely to enter, leading to a gender gap in tournament entry among intermediate performers ($p < 0.038$),

consistent with the findings in Brandts et al. (2015).[26] We believe that it points to an interesting phenomenon. Compared to both weak and strong performers, intermediate performers have more leeway to interpret advice in a biased way. Intermediate-performing women tend to interpret advice in a more pessimistic way than men and this leads to the gender gap in tournament entry. Undoubtedly, high-performers are of utmost interest for the labor market. However, if one accepts that in natural environments most people are probably intermediate performers, then the phenomenon we identify here may be of broader importance. Below we will report on other differences that pertain particularly to the group of intermediate performers.

	Advisees' tournament entry rates					
	Weak		*Intermediate*		*Strong*	
Performance group	Men	Women	Men	Women	Men	Women
Male Advisor						
Advice "piece rate"	38%	0%	**60%**	**14%***	100%	50%
	(5 of 13)	(0 of 8)	**(9 of 15)**	**(2 of 14)**	(1 of 1)	(2 of 4)
Advice "tournament"	---	---	78%	43%	83%	81%
	---	---	(7 of 9)	(6 of 14)	(10 of 12)	(13 of 16)
Female Advisor						
Advice "piece rate"	38%	25%	**40%**	**8%***	67%	50%
	(5 of 13)	(1 of 4)	**(4 of 10)**	**(2 of 24)**	(6 of 9)	(1 of 2)
Advice "tournament"	---	---	67%	55%	87%	89%
	---	---	(6 of 9)	(6 of 11)	(13 of 15)	(8 of 9)

Note. Bold men–women value pairs are statistically significantly different.
*** p < 0.001, ** p < 0.01, * p < 0.05 (two-sided Fisher's exact test for entry rates).

***Table 4**: Tournament entry rates of advisees (sample is the GM data).*

Why do intermediate-performing men and women enter so differently with advice? The advisor's gender does not seem to play a role. The key to the persistence of the overall gender gap is the fact that men and women differ substantially in how they react to the particular advice that they receive. Table 4 shows the tournament entry rates for men and women after receiving the 'tournament' or 'piece rate' advice from a male or female advisor and for all three performance intervals.

Consider first the results aggregated over all three intervals. Whereas women react strongly to the advice 'piece rate' and enter the tournament few times independent of whether it comes from a male advisor (15%, 4 of 26) or female advisor (13%, 4 of 30), men still enter

[26] With a Holm-Bonferroni correction (two matching conditions x three performance groups), the test results remain significant for one gender matching condition applying the chi2 test instead of the Fisher's exact test.

the tournament frequently with 52% (15 of 29, male advisor; p = 0.010) and 47% (15 of 32, female advisor; p = 0.009).

	Expected forgone earnings	
	(1)	(2)
Female advisee	-1.002**	0.188
	(0.311)	(0.520)
Female advisor	-0.185	0.319
	(0.379)	(0.588)
*Female advisee * Female advisor*	0.327	-0.485
	(0.460)	(0.774)
Task 2 performance	0.070	-0.241
	(0.058)	(0.189)
Task 2 - task 1 performance	-0.014	0.035
	(0.043)	(0.072)
Constant	0.747	2.884
	(0.466)	(1.914)
Observations	101	43
R-squared	0.129	0.081

Notes. *** p<0.001, ** p<0.01, * p<0.05. OLS regressions with robust standard errors in parentheses. The dependent variable *Expected foregone earnings* is the difference between expected earnings under the payment scheme the participant did not choose and expected earnings under the payment scheme she chose if a participant chose the (for her) inferior payment scheme. The sample (GM data set) is weak- and intermediate-performing advisees receiving the advice 'Piece rate' in model (1) and intermediate- and strong-performing advisees receiving the advice 'Tournament' in model (2).

Table 5: *Expected foregone earnings of advisees (sample is the GM data)*

While women overall follow the received advice – in particular the advice 'piece rate' – from both male and female advisors, men refuse to follow the advice 'piece rate'. The difference is particularly large in the intermediate-performance interval and more marked than with gender-blind advice (Brandts et al., 2015). For the advice 'tournament', the differences are much less pronounced (p = 0.221 and p = 0.509, respectively), which is different than in Brandts et al. (2015) where men react very strongly to the gender-blind advice 'tournament'. This result is at odds with our hypothesis above.[27]

[27] Note that the overall findings that are significant at the 1% level are also significant at the 5% level after applying the Holm-Bonferroni correction and taking into account a family of 4 null hypotheses (two matching conditions x two pieces of advice). Applying the Holm-Bonferroni correction to the analysis in the three performance groups leads to insignificant gender differences. We therefore interpret the findings in the intermediate performance interval cautiously.

In line with our hypothesis above, we do not find a gender-matching effect in the intermediate performance group. Table 5 shows OLS regressions with robust standard errors. The dependent variable is advisees' expected foregone earnings and explanatory variables are dummies for the advisee's and the advisor's gender, their interaction term, task 2 performance and the difference in performance between task 1 and task 2. The sample is advisees receiving the 'piece rate' advice in the weak and intermediate performance group, model (1), and advisees receiving the 'tournament' advice in the intermediate and strong performance group, model (2). The gender difference in the reaction to the 'piece rate' advice is confirmed by the significant coefficient estimate of the *Female advisee* dummy in model (1). However, there is no effect of gender matching: The coefficient estimates of *Female advisor* (gender-matching effect on male advisees) and a post-estimation test of *Female advisor* and the interaction term (gender-matching effect on female advisees) are insignificant in both models ($p > 0.574$). We summarize the findings pertaining to our research question 3 in the following result:

Result 3: Men follow the 'piece rate' advice less often than women independently of the advisor's gender, but there is no difference for the advice 'tournament'. The gender difference in following the 'piece rate' advice seems to be particularly pronounced in the intermediate-performance group.

Note that for advice following we find a similar result to those for advice giving. It is in the intermediate performance group that women's behavior is most different from that of men.

A separate issue that can be studied with our data is communication style. A detailed analysis of the reasons given for a particular advice and how advisees react to the reasons can be found in appendix C. [28] The main findings are that advisors choose more reasons in a positive sense (when giving the advice 'tournament') than in a negative sense (when giving the advice 'piece rate'). Furthermore, men are more reluctant to discourage women's than men's confidence in their success (when giving the advice 'piece rate'), but in general advisors who enter the tournament themselves are more likely to discourage the advisee's confidence. Advisees' reaction to the discouragement is not strongly pronounced.

[28] There is some scientific evidence that suggests that men and women have different communication styles (Basow and Rubenfield, 2003; Chodorow, 1999; Lakoff, 2004; Wood, 2012) and pursue different communication goals (Leaper, 1991; Maltz and Borker, 1982; Wood, 2012). Giving advice is inseparably linked with communicating. In our setup, communication is not free form, but also the choice of preformulated statement allows us to look into gender differences in communication styles.

5. SUMMARY AND CONCLUSIONS

In this paper we study the impact of the advice process on entry into competition. We focused on two issues. The first is how advice *giving* is affected by gender and gender matching and the second is how advice *following* is affected by gender matching.

Here we summarize our three main results, all of which refer to intermediate performers. With respect to advice giving, we find, first, a gender gap for the intermediate relative performance interval. We find that women are less likely to recommend 'tournament' here and the gap is independent of the advisee's gender. Second, for the correctness of advice our results show that, in the intermediate relative performance interval, women's advice is less often correct, in the sense of maximizing the advisee's expected earnings. Third, with respect to advice following our results show that men and women do not follow advice differently depending on the advisor's gender, but there is a gender gap among advisees because men follow the advice 'piece rate' less often. This latter finding seems to be particularly pronounced in the intermediate performance group.

We think that our results for the intermediate-performance interval are particularly interesting, since in natural environments most people are arguably intermediate performers Women on both sides of the advice process show more reluctance to tournament entry: Female advisor are more reluctant to recommend competition than male advisors and a gender gap in tournament entry emerges among advisees in the intermediate group. In situations in which very good (though not top) performers are important for an organization, the design of mentoring programs for this particular group might require special attention. These results resonate with finding in the bargaining literature that document that gender differences are largest when ambiguity as to whether bargaining is possible is largest (see Hernandez-Arenaz and Iriberri, 2017, for a recent survey and the papers cited therein).

Naturally, all our results have to be evaluated in the context of our very stylized environment. It is, of course, possible that with more contact between partners, as with face-to-face communication, gender and gender matching would have stronger or different effects than the ones we find. Women's reluctance to advising tournament entry in the intermediate relative performance interval and men's reluctance to discouraging women's compared to men's self-confidence give some indication that this could indeed be the case. However, we do feel that our results are of interest, since they capture the pure effects of gender matching. In a very broad sense, our results are in line with those of Card et al (2020). There, simply knowing the gender of authors affects reviewers and editors.

We end with an assessment of changes in priors, given our new evidence. For gender matching, we cannot support a change in future priors for two reasons: one is statistical (we do not find gender-matching effects), the other is design-related (the interaction between advisor and advisee is limited). For the impact of advice on competitiveness – and more concretely, gender differences in advice giving and following – our study sets new priors in that we find significant and substantial effects for both. These are robust across gender matching conditions.

REFERENCES

Basow, S. A., & Rubenfeld, K. (2003). "Troubles talk": Effects of gender and gender typing. *Sex roles*, *48*(3-4), 183-187.

Bateman, A.J. (1948). Intra-sexual selection in Drosophila. *Heredity* 2, 349-368.

Bettinger, E. P., & Baker, R. B. (2014). The effects of student coaching an evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, *36*(1), 3-19.

Bettinger, E., & Long, B. T. (2005). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review: Papers and Proceedings, 95*(2), 152-157.

Beugnot, J. & Peterlé, E. (2020). Gender bias in job referrals: An experimental test. *Journal of Economic Psychology*, 76, Article 102209.

Bischof, D., & Zurich, C. (2017). New graphic schemes for Stata: plotplain and plottig. *The Stata Journal*, *17*(3), 748-759.

Blau, F. D., Currie, J. M., Croson, R. T., & Ginther, D. K. (2010). Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial. *The American Economic Review: Papers & Proceedings*, *100*(2), 348-352.

Brandts, J., & Garofalo, O. (2012). Gender pairings and accountability effects. *Journal of Economic Behavior & Organization*, *83*(1), 31-41.

Brandts, J., V. Groenert and C. Rott (2015). The impact of advice on women's and men's selection into competition. *Management Science*, *61*(5), 1018 - 1035.

Brandts, J., K. Gërxhani and A. Schram (2020). Are women status-ranking averse?. *Journal of Behavioral and Experimental Economics*, 84, article 101485.

Brandts, J., Cooper, D. and C. Rott (2019). Communication in Laboratory Experiments. *Handbook of Research Methods and Applications in Experimental Economics*, edited by Arthur Schram and Aljaz Ule, Edward Elgar Publishing Ltd.

Buser, T., Niederle, M. and Oosterbeek, H. (2014). Gender, Competitiveness and Career Choices. *Quarterly Journal of Economics*, 129(3), 1409-1447.

Buser, T., Peter, N., & Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review*, *107*(5), 125-130.

Card, D., Della Vigna, S., Funk, P. and Iriberri, N. (2020). Are Referees and Editors in Economics Gender Neutral? *Quarterly Journal of Economics, 135(1),* 269-327.

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics, 125*(3), 1101-1144.

Carlsson, F., Lampi, E., Martinsson, P. & Yang, X. (2020). Replication: Do Women Shy Away from Competition? Experimental Evidence from China. *Journal of Economic Psychology*. Article 102312.

Carrington, B., Tymms, P. & Merrel, C. (2008). Role models, school improvement and the "gender gap" – Do men bring out the best in boys and women the best in girls? *British Educational Research Journal, 34*(3), 315-327.

Chodorow, N. J. (1999). *The reproduction of mothering: Psychoanalysis and the sociology of gender*. Univ of California Press.

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

Dasgupta, U., Mani, S., Sharma, S. & Singhal, S. (2019). Can gender differences in distributional preferences explain gender gaps in competition? *Journal of Economic Psychology*, *70*, 1-11.

Datta Gupta, N., A. Poulsen, and M.C. Villeval (2013). Gender and competitiveness: Experimental Evidence. *Economic Inquiry*, 51: 816-835.

Dee, T. S. (2004). Teachers, race and student achievement in a randomized experiment. *The Review of Economics and Statistics, 86*(1), 195-210.

Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity or Gender Matter? *American Economic Review: Papers and Proceedings, 95*(2), 158-165.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources 42*(3), 528-554.

De Sousa, J. & Hollard, G. (2016). Gender differences: Evidence from field tournaments. Working paper.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review, 48*(3), 547-561.

Eskreis-Winkler, L., Milkman, K. L., Gromet, D. M., & Duckworth, A. L. (2019). A large-scale field experiment shows giving advice improves academic outcomes for the advisor. *Proceedings of the National Academy of Sciences*, *116*(30), 14808-14810.

Ewers, M. (2012). Information and competition entry. *Discussion Paper 6411*, Institute for the Study of Labor (IZA), Bonn, Germany.

Filippin, A., & Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11), 3138-3160.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.

Friedl, A., Pondorfer, A. & Schmidt, U. (2020). Gender differences in social risk taking. *Journal of Economic Psychology*, *77*, March 2020, 102182.

Gerdes, C. & Gränsmark, P. (2010). Strategic behavior across gender: A comparison of female and male chess players. *Labour Economics*, 17(5), 766-775.

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.

Hernandez-Arenaz, I. & Iriberri, N. (forthcoming). A Review of Gender Differences in Negotiation. *Oxford Research Encyclopedia of Economics and Finance*.

Hoffmann, F., & Oreopoulos, P. (2009). A Professor Like Me: The Influence of Instructor Gender on University Achievement. *Journal of Human Resources, 44*(2), 479-494.

Holmlund, H., & Sund, K. (2008). Is the Gender Gap in School Performance Affected by the Sex of the Teacher? *Labour Economics, 15*(1), 37-53.

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

Jung, S. & Vranceau, R. (2019). Competitive compensation and subjective well-being: The effect of culture and gender. *Journal of Economic Psychology*, *70*, 90-108.

Lahelma, E. (2000). Lack of male teachers: A problem for students or teachers? *Pedagogy, Culture and Society, 8*(2), 173-86.

Lakoff, R. T. (2004). *Language and woman's place: Text and commentaries* (Vol. 3). Oxford University Press, USA.

Leaper, C. (1991). Influence and involvement in children's discourse: Age, gender, and partner effects. *Child development*, *62*(4), 797-811.

Ludwig, S., Fellner-Röhling, G., & Thoma, C. (2017). Do women have more shame than men? An experiment on self-assessment and the shame of overestimating oneself. *European Economic Review*, *92*, 31-46.

Maltz, D. N., & Borker, R. (1982). A cultural approach to male-female miscommunication. In J. J. Gumpertz (Ed.): *Language and social identity*. Cambridge; Cambridge University Press.

McCarty, P. A. (1986). Effects of feedback on the self-confidence of men and women. *Academy of Management Journal*, *29*(4), 840-847.

McEvily, B., Weber, R. A., Bicchieri, C., & Ho, V. T. (2006). Can groups be trusted? An experimental study of trust in collective entities. *Handbook of trust research*, 52-67.

Mobius, M. M., M. Niederle, P. Niehaus and T. Rosenblat (2014). Managing Self-Confidence. *Working paper*.

Niederle, M. (2016). Gender. In: *Handbook of Experimental Economics*, second edition, by John Kagel and Alvin E. Roth (eds). Princeton University Press: 481-553.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067-1101.

Nixon, L. A., & Robinson, M. D. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography, 36*(2), 185-194.

Noe, R. A. (1988). Women and mentoring: A review and research agenda. *Academy of management review*, *13*(1), 65-78.

Price, C.R. (2012). Gender, competition, and managerial decisions (2012), *Management Science*, *58(1)*, 114-122.

Reuben, E., P. Sapienza, and L. Zingales (2017). Taste for competition and the gender gap among young business professionals. *Working paper*.

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

Shurchkov, O. & van Geen, A.V.M. (2019). Why female decision-makers shy away from promoting competition. *Kyklos, 72(2)*, 297-331.

Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining—Beware the same sex! *Experimental Economics*, *12*(3), 318-331.

Tanis, M., & Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, *35*(3), 413-424.

Trivers, R.L. (1972). Parental investment and sexual selection. In *Sexual Selection and the Descent of Man*, ed. B. Campbell. Chicago: Aldine.

Tungodden, J. (2019). When parents decide: Gender differences in competitiveness. *Working Paper 03/19*. Center for Applied Research at NHH.

Wood, J. T. (2012). *Gendered Lives*. Cengage Learning.

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, *32*(1), 161-198.

APPENDICES

APPENDIX A. PERFORMANCE DISTRIBUTION

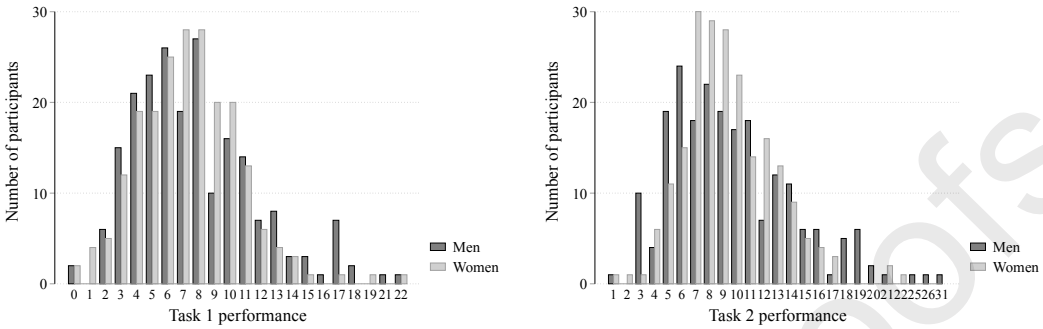
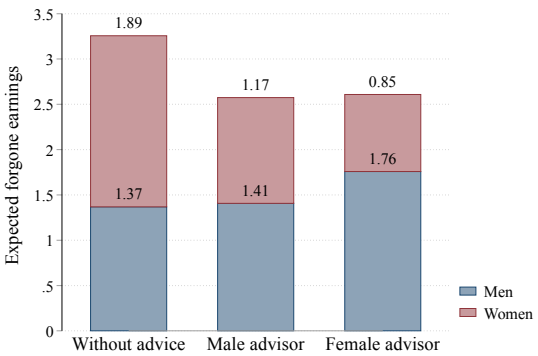


Figure A1: Women's and men's performance in tasks 1 and 2 (sample is the GM data)

APPENDIX B. OVERALL IMPACT OF ADVICE

Figure B1 presents the average expected forgone earnings without and with advice from male (female) advisors.[29] Forgone earnings are zero for a participant who chooses the payment scheme in task 3 that maximizes the expected earnings for her performance. For a participant who chooses the sub-optimal payment scheme (in terms of expected payoff), forgone earnings are defined as the difference between the expected earnings had the participant made the opposite choice minus the actually expected earnings under the chosen payment scheme for task 3. Calculations are ex-ante, i.e., they are based on task 2 performance. Table B1 summarizes the results of the probability calculation. Expected earnings are the same in the tournament and with the piece rate if the probability of winning is 25%. This is about the case with 10 correct answers.



[29] The figures were created using the plot scheme developed by Bischof and Zurich (2017).

Figure B1*: Average expected forgone earnings with and without advice (by gender and advisor's gender, sample is the GM data).*

While the expected forgone earnings without advisors are 1.62 (1.37 for men and 1.89 for women), they are only 1.28 for advisees who receive advice from a male advisor (1.41 for men and 1.17 for women) and 1.33 for advisees matched with a female advisor (1.76 for men and 0.85 for women).[30] Though the magnitude of the drop is slightly larger with male advisors ($p = 0.134$, one-sided Mann-Whitney U test), it is not statistically significant, whereas it is significant with female advisors ($p = 0.034$, one-sided Mann-Whitney U test).

However, the overall gender gap in tournament entry does not diminish with advice. With advice, 62.3% (66 of 106) of men and 38.7% (41/106) of women enter the tournament. This gap is quite similar to the one without advice, where 54.7% of men and 36.8% of women enter ($p = 0.013$). As we will see in more detail below, women close the gap among high-performers by doubling entry rates with advice, but a gender gap emerges among intermediate performers. We can now formulate the following first result:

Result B1: Advice reduces average expected forgone earnings, both with male and female advisors, but only significantly for the latter case. However, the overall gender gap in tournament entry persists, both with male and female advisors.

We thus confirm the finding of gender-blind advice in Brandts et al. (2015), where advice also reduced overall forgone earnings and the gender gap did not disappear with advice.

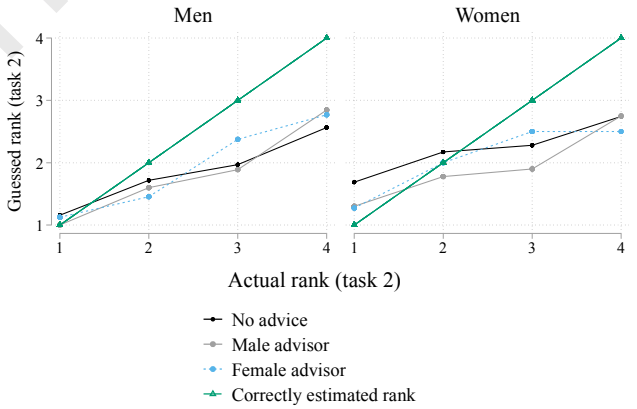


Figure B2*: Guessed rank relative to actual rank (task 2 performance, by advisor's and advisee's gender; sample is the GM data).*

[30] Average foregone earnings of men with female advisors go up from 1.37 to 1.76, but this increase is not statistically significant (MW U-test: $z = -0.104$, $p = 0.9168$).

We also replicate women's increase in self-confidence with gender-blind advice reported in Brandts et al. (2015). Figure B2 depicts the average guessed rank (task 2) of men and women given their actual rank. The actual rank is the rank that a participant is most likely to obtain (using the probability of winning from the expected earnings calculation as described above). Strong-performing women (corresponding to actual rank 1 in task 2) become significantly more confident with advice from men ($p = 0.035$), but not from women ($p = 0.074$).[31]

We now move to the analysis of the effects of gender matching on forgone earnings and the gender gap. We compare the advisor with the advisee generation to evaluate the effects of advice. The differences between advisors' and advisees' forgone earnings are insignificant for men ($p > 0.347$, one-sided Mann-Whitney U test, for both advisor genders), but significant for women ($p = 0.032$ for female advisors vs. female advisees matched with male advisors; $p = 0.005$ for female advisors vs. female advisees matched with female advisors; one-sided Mann-Whitney U tests). With respect to gender matching, we find no differences in the changes in foregone earnings and in the reduction of the gender gap for both men and women consistent with our hypothesis that gender matching should not matter in this case. Note also that with respect to the pure gender effect, men's foregone earnings do not change significantly while those of women go down significantly, both in contrast to what was observed in Brandts et al. (2015).

The changes in the gender gaps depending on gender matchings in advice are the following: Upon receiving advice from a male advisor, 64.0% (32 of 50) of men and 41.1% (23 of 56) of women enter the tournament ($p = 0.021$). After receiving advice from a female advisor, the rates are 60.7% (34 of 56) of men versus 36.8% (18 of 50) of women ($p = 0.012$). Observe that in both cases the gender gap is similar to the one without advice (54.7% of men versus 36.8% of women, $p = 0.013$). The conclusion is that gender matchings do not affect the change in the gender gap.

Result B2: Women's forgone earnings are significantly reduced by advice from both women and men. By contrast, men's forgone earnings do not significantly change after advice. Gender matchings make no difference for the lack of change in the gender gap.

[31] Except for women in the intermediate group (corresponding to actual rank 2 and 3) who become more confident with advice from a man ($p = 0.021$), there are no other significant changes in the self-assessment with advice.

To better understand Result B2 we investigate further the apparent contradiction between the reduction in forgone earnings and the persistence of the gender gap. It can be understood by taking a closer look at the performance levels of those who change their tournament entry decisions. To do this we divide participants into three groups, according to their performance in task 2: weak (26%), intermediate (52%), and strong performance (22%), see Table B1.

Task 2 performance	<5	5	6	7	8	9	**10**	11	12	13	>13
Men (in %)	<0.1	0.1	0.4	1.8	5.9	13.3	**23.6**	34.7	45.6	57.2	>68.7
Women (in %)	<0.1	0.1	0.6	2.4	6.4	13.2	**22.1**	32.7	42.4	52.0	>63.0
Performance group (% of participants)	**Weak (26%)**			**Intermediate (52%)**					**Strong (22%)**		
Most likely (=optimal guessed) rank	4 (worst)			3 or 2					1 (best)		

Note. Participants in the weak performance group solve six or fewer additions correctly, their probability of winning in the tournament is less than 1%, and their most likely rank is fourth (the lowest rank). Participants in the intermediate group solve between seven and eleven problems and win the tournament with a probability larger than 1% and smaller than 35%. Note that the threshold performance level (ten correct answers), where the expected earnings from entering the tournament more or less equal the earnings from the piece rate, lies within this performance group. Participants in the strong performance group solve twelve or more problems, their probability of winning the tournament is at least 42%, and the most likely rank in a randomly composed group of four is first (the highest rank).

Table B1*: Probability of winning given a certain performance level in task 2 (sample is the GM data).*

Figure 3 depicts the share of men and women who choose the tournament payment scheme for each of the three performance groups. For each performance group (weak, intermediate, and strong) there are three bars: entry decisions of participants without advice, after having received advice from a male advisor, and after having received advice from a female advisor. The sample is composed of advisors and advisees in the GM data set.[32]

We look separately at the gender gap in each of the three performance groups. We start with strong performers. Without advice, the gender gap is substantial: whereas 87% (20 of 23) of strong-performing men enter the tournament only 45% (10 of 22) of women do so ($p = 0.005$). This changes with advice. More strong-performing women enter the competition after having received advice either from a man (75%, 15 of 20) or from a woman (82%, 9 of 11) than without advice (45%, 10 of 22) ($p < 0.068$), thereby closing the gender gap among strong-

[32] In contrast to Brandts et al. (2015), there is no significant improvement in the entry rates of low-performing men. This can be attributed to the high entry rate without device in the gender-blind data.

performers. The difference in entry rates between men and women is now statistically insignificant ($p > 0.676$).

Weak-performing men without advice enter the tournament with 37.5% (12 of 32) rather frequently given that their probability of winning the tournament is less than 1%. By contrast, weak-performing women without advice enter the tournament in 20% (7 of 35) of cases.[33] Neither weak-performing men nor weak-performing women change entry decisions significantly with advice, independent of the advisor's gender.[34] There is no significant gender gap among weak performers with or without advice ($p = 0.138$ without advice, $p = 0.111$ with male advisor and $p = 1.000$ with female advisors).

The reason why the overall gender gap does not disappear with advice is due to the effect of advice on the group with intermediate performance. Women in the intermediate group become less likely to enter the tournament with advice (though not significantly, $p > 0.147$), whereas men are more likely to enter, leading to a gender gap in tournament entry among intermediate performers ($p < 0.038$). This remarkable pattern of changes for intermediate performers is the same as in the gender-blind case, another element of robustness with respect to the results reported in Brandts et al. (2015).

APPENDIX C. REASON GIVING – PREFERENCE FOR COMPETITION, CONFIDENCE, AND EARNINGS RISK

Recall that advisors can choose up to three reasons for their chosen piece of advice. For the advice 'piece rate', the phrases are: (1) "Porque no es divertido competir con otros." / "Because it is not fun to compete with others." (preference for competition), (2) "Porque no deberías tener confianza de que te vaya bien." / "Because you should not be confident that you will succeed." (self-confidence), and (3) "Porque con la remuneración por unidad ganas algo seguro." / "Because with the piece rate you earn something for sure." (risk of earnings). For the advice 'tournament', the three phrases are: (1) "Porque es divertido competir con otros." / "Because it is fun to compete with others." (preference for competition), (2) "Porque deberías tener confianza de que te vaya bien." / "Because you should be confident that you will succeed." (self-confidence), and (3) "Porque en la competición puedes ganar mucho más." / "Because in the competition you can earn much more." (risk of earnings).

[33] Observe that both low-performing men and women are over-confident, but men much more so.

[34] Advice does not correct the entry decisions of weak-performing men, neither when it comes from a male advisor (38%, 5 of 13) nor when it comes from a female advisor (38%, 5 of 13). Though the entry rate of weak-performing women goes down to 0% (0 of 8) with a male advisor, the change is not statistically significant ($p = 0.315$), neither is it with a female advisor (25%, 1 of 4).

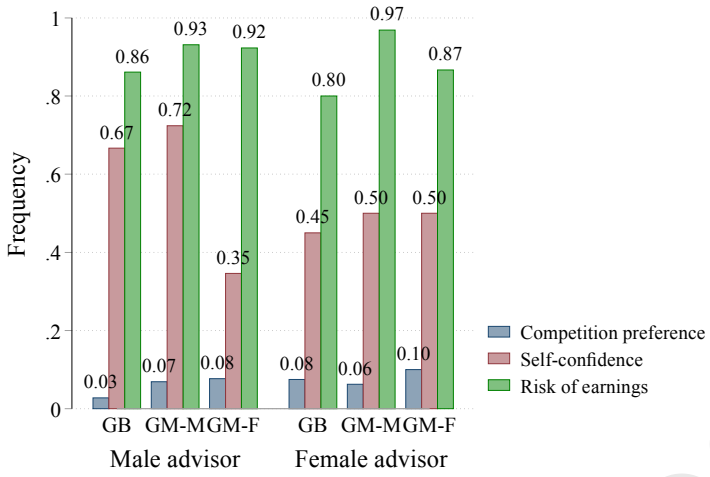


***Figure C1**: Advisor's reasons for the advice 'piece rate' (by advisor's and advisee's gender; sample is the GB and GM data where GB = advisee's gender blind, GM-M = male advisee, GM-F = female advisee).*

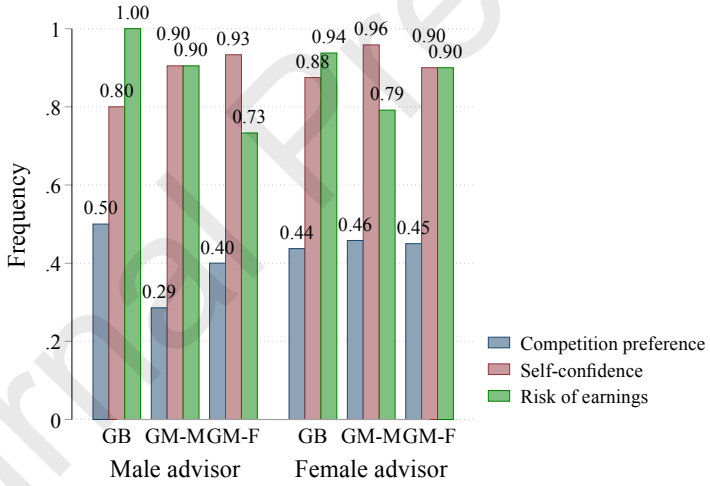


***Figure C2**: Advisor's reasons for the advice 'tournament' (by advisor's and advisee's gender; sample is the GB and GM data where GB = advisee's gender blind, GM-M = male advisee, GM-F = female advisee).*

Figure C1 (C2) shows the share of male and female advisors who choose the phrase referring to competition preference, confidence, and/or risk of earnings for the recommended payment scheme 'piece rate' ('tournament'). The sample includes the GB and GM data sets and the shares are shown separately for the advisor's gender (male or female) as well as the advisee's gender (revelation) (gender blind, male, female). Note first that independently of the recommended payment scheme, advisors almost always mention risk of earnings to underline their advice (between 73% and 100%). That is, advisors refer with similar frequencies to risk

and earnings in a positive sense than in a negative one. Overall, advisors provide more reasons if they recommend the competitive payment scheme 'tournament' compared to the non-competitive payment scheme 'piece rate'. This is mainly driven by underlining more often the preference for competition ($p < 0.057$ for each gender of advisor and advisee) and encouraging more often self-confidence ($p = 0.160$ for male advisors matched with male advisees, $p = 0.365$ for male advisors in gender blind; $p < 0.006$ for all other matches). In other words, advisors refer more to the preference for competition and self-confidence in a positive sense than in a negative sense.

Result C1: Advisors underline more often the preference for competition and self-confidence in a positive sense (when giving the advice 'tournament') than in a negative sense (when giving the advice 'piece rate').

To investigate gender differences in the use of reasons we use regression analysis. Since advisors choose among the three reasons simultaneously, we run Seemingly Unrelated Regression models for all three reasons and separately for the advice 'piece rate' (SUR 2 and SUR 3 in Table C1) and 'tournament' (SUR 4 and SUR 5 in Table C1). Logit and probit regressions with seemingly unrelated estimation lead to similar results. The dependent variable *Reason* takes the value 1 if chosen by the advisor and 0 otherwise. Reason refers to Preference for competition, Self-confidence, and Risk of earnings in regression models (a), (b), and (c), respectively. The explanatory variables are the same as in Table 3 (see description in the main text) and SUR 2 and 4 include in addition the variable *Advisor's entry task 3*, which takes the value 1 if the advisor choses the tournament herself and 0 otherwise. The sample is advisors, who advise 'piece rate' ('tournament') for SUR 2 and 3 (SUR 4 and 5) and from the GM data set only because we are interested in the effect of gender. In SUR 2 and 3, the advisors refer to competition, self-confidence, and in a discouraging way and in SUR 4 and 5 in an encouraging way. For the GB data, the same regression models (without the variable *Female advisee* and without the interaction term) are presented in Table C2 and confirm similar patterns.

	Advice 'piece rate'						Advice 'tournament'					
	SUR 1			SUR 2			SUR 3			SUR 4		
VARIABLES	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
Female advisor	-0.019	-0.211	0.033	-0.015	-0.153	0.018	0.162	0.048	-0.145	0.163	0.049	-0.144
	(0.070)	(0.123)	(0.069)	(0.070)	(0.118)	(0.069)	(0.144)	(0.077)	(0.105)	(0.143)	(0.077)	(0.104)
Female advisee	-0.002	-0.361**	-0.009	-0.005	-0.400**	0.001	0.113	0.029	-0.155	0.077	0.015	-0.174
	(0.073)	(0.128)	(0.072)	(0.073)	(0.122)	(0.072)	(0.137)	(0.073)	(0.099)	(0.138)	(0.074)	(0.100)
*Female advisor * Female advisee*	0.046	0.352*	-0.089	0.049	0.395*	-0.100	-0.103	-0.096	0.340*	-0.092	-0.092	0.346*
	(0.101)	(0.178)	(0.100)	(0.101)	(0.169)	(0.099)	(0.202)	(0.108)	(0.147)	(0.200)	(0.108)	(0.146)
Advisor's entry task 3				0.022	0.326**	-0.083				0.151	0.057	0.077
				(0.054)	(0.090)	(0.053)				(0.103)	(0.056)	(0.075)
Info task 2	0.003	-0.045**	-0.000	0.004	-0.034*	-0.003	0.010	0.001	0.028**	0.009	0.001	0.027**
	(0.010)	(0.017)	(0.010)	(0.010)	(0.017)	(0.010)	(0.014)	(0.008)	(0.010)	(0.014)	(0.008)	(0.010)
Info task 2 - task 1	0.000	0.024	-0.006	-0.000	0.015	-0.004	0.007	0.012	0.018	0.005	0.011	0.017
	(0.012)	(0.021)	(0.012)	(0.012)	(0.020)	(0.011)	(0.020)	(0.011)	(0.014)	(0.020)	(0.011)	(0.014)
1st best performance task 2	0.005	-0.004	0.008	0.005	-0.003	0.008	0.012	-0.004	0.015	0.009	-0.005	0.013
	(0.007)	(0.013)	(0.007)	(0.007)	(0.012)	(0.007)	(0.015)	(0.008)	(0.011)	(0.015)	(0.008)	(0.011)
2nd best performance task 2	-0.010	0.012	-0.011	-0.012	-0.008	-0.006	0.006	0.021	-0.008	0.004	0.020	-0.009
	(0.012)	(0.022)	(0.012)	(0.013)	(0.022)	(0.013)	(0.034)	(0.018)	(0.025)	(0.034)	(0.018)	(0.025)
Constant	0.086	0.979**	0.944**	0.082	0.926**	0.958**	-0.056	0.720**	0.377*	-0.042	0.726**	0.384
	(0.139)	(0.246)	(0.138)	(0.140)	(0.234)	(0.137)	(0.283)	(0.151)	(0.205)	(0.280)	(0.150)	(0.204)
Observations	117	117	117	117	117	117	95	95	95	95	95	95
R-squared	0.012	0.125	0.034	0.014	0.213	0.054	0.043	0.040	0.178	0.064	0.050	0.187

Notes. *** p<0.001, ** p<0.01, * p<0.05. Seemingly Unrelated Regressions (SUR) with standard errors in parentheses. For the ease of result interpretation, SURE results are displayed. Logit and probit regressions with seemingly unrelated estimation lead to similar results. The dependent variable Reason takes the value 1 if chosen by the advisor and 0 otherwise. Reason refers to *Preference for competition*, *Self-confidence*, and *Risk of earnings* in regressions (a), (b), and (c), respectively. Controls are *Info task 2, Info task 2 - task 1, 1st best performance task 2*, and *2nd best performance task 2*. The sample is the GM data set and advice "piece rate" for SUR 1 and 2, and the GM data and advice "tournament" for SUR 3 and 4.

Table C1*: SUR regressions of reasons given by advisor (sample is the GM data).*

Interestingly, there are no gender differences for the reason referring to Preference for competition, models (1a), (2a), (3a), and (4a). However, when recommending the non-competitive payment scheme 'piece rate', we find three interesting gender results: First, women are less likely than men to discourage (male) advisees' self-confidence. The effect size is with 21 percentage points substantial, see model (1b). Second, men are more reluctant to discourage a female advisee than a male advisee to be self-confident about her performance, models (1b) and (2b). The effect size is even more striking and significant at the 1% level. Compared to a male advisee, men are 36 - 40 percentage points less likely to mention to a female advisee that she should not be confident that she would succeed. Third, women do not differentiate between male and female advisees and discourage self-confidence equally: the coefficient estimate for the interaction term *Female advisor * Female advisee* has about the same size as the coefficient estimate of the variable *Female advisee*. The effects are particularly strong for the weak relative performance interval.[35] A similar pattern is observed for Risk of earnings after the advice 'tournament' (models (3c) and (4c) in Table C1).

The advisor's own tournament entry is a strong predictor of her likelihood to discourage a weakly performing advisee's confidence: Men and women who enter the tournament are more likely to discourage confidence after the advice 'piece rate' and explain to a large extend the gender differences in discouragement (model (2b) in Table C1). Put differently, women who entered the competition themselves are not less likely than men to discourage the advisee's confidence. Just like for the piece of advice given, men and women who enter the competition do not differ substantially.

An indication that advisors take advice giving seriously is that the variable *Info task 2* is negative and significant for the reason *Self-confidence* under advice 'piece rate', which is strongly related to ability. Furthermore, the variable *Info task 2* is positive and significant for the reason *Risk of earnings* under advice 'tournament', which underlines the possibility of higher earnings with the competition. These findings parallel the result from Table 3 in the main text.[36]

[35] Regression results not reported, available upon request.

[36] In Table A2, we show the selection of reasons with our data from Brandts et al. (2015) where gender of advisor and advisee was not revealed. The SUR regressions are the same as in Table A1 except that the variable *Female advisee* and the interaction term between *Female advisor* and *Female advisee* are not included since the paired advisee's gender is not revealed. Regression (1b) in Table A2 confirms that women are significantly less likely to mention self-confidence when recommending the piece rate and regression (2b) confirms that advisors who enter themselves the competition are more likely to discourage the advisee's confidence.

Since advisors vary considerably in the discouragement of advisees' self-confidence, it would be interesting to know if it affects advisees' self-confidence. We use the advisees' guessed task 2 rank as an inverse measure of self-confidence. Overall, the discouragement of confidence combined with the advice 'piece rate' has no effect on self-confidence: The average guessed rank of advisees who receive the advice 'piece rate' in the GM data is 2.1 if the advisor does not discourage confidence and 2.3 if she does ($p = 0.278$). A small exception is male advisors who discourage confidence. They seem to lower confidence among male and female advisees slightly ($p = 0.075$) and in particular among female advisees ($p = 0.106$) compared to male advisors who do not discourage confidence. We do not want to oversell these results though because it is difficult to pinpoint causality (the displayed reasons are endogenously chose by advisors) and we would need more data to draw reliable conclusions.

Result C2: Men are more reluctant to discourage women's than men's confidence in their success, but in general, advisors who enter the tournament themselves are more likely to discourage the advisee's confidence. Advisees' reaction to the discouragement is however not strongly pronounced.

Strong-performing women update their beliefs about their relative performance in a positive way (see Appendix B) and low- to intermediate-performing individuals do not update their self-assessment. This finding is (partially) in line with the results in Mobius et al. (2014) who show theoretically and experimentally that individuals over-weight positive feedback relative to negative feedback and update self-confidence after receiving feedback about their performance too little. Interestingly, men's advice following of the advice 'piece rate' and 'tournament' is somewhat weaker when gender is revealed compared to gender-blind advice. The revelation of gender and gender seem to be important factors in the advice process both on the giver's and the receiver's side. With face-to-face or free-form communication, gender could potentially have a stronger or different effect on advice giving and advice following.

	Advice 'piece rate'						Advice 'tournament'					
	SUR 1			SUR 2			SUR 3			SUR 4		
VARIABLES	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
Female advisor	0.045	-0.209	-0.066	0.040	-0.151	-0.102	-0.101	0.064	-0.019	-0.023	0.079	-0.053
	(0.052)	(0.109)	(0.083)	(0.053)	(0.109)	(0.084)	(0.164)	(0.127)	(0.047)	(0.167)	(0.133)	(0.046)
Advisor's entry task 3				-0.021	0.249*	-0.153				0.269	0.055	-0.119*
				(0.056)	(0.114)	(0.088)				(0.176)	(0.141)	(0.048)
Info task 2	0.001	0.001	0.010	0.001	0.003	0.009	0.028	-0.002	-0.014*	0.037	-0.001	-0.018**
	(0.011)	(0.024)	(0.018)	(0.011)	(0.023)	(0.018)	(0.025)	(0.019)	(0.007)	(0.025)	(0.020)	(0.007)
Info task 2 - task 1	-0.008	-0.038	-0.042*	-0.007	-0.049	-0.035	-0.080	0.021	-0.010	-0.103	0.016	-0.000
	(0.013)	(0.027)	(0.020)	(0.013)	(0.026)	(0.020)	(0.056)	(0.043)	(0.016)	(0.056)	(0.045)	(0.015)
1st best performance task 2	0.006	-0.008	-0.018	0.006	-0.014	-0.014	0.046	0.000	0.007	0.044	-0.000	0.008
	(0.008)	(0.018)	(0.014)	(0.009)	(0.018)	(0.014)	(0.034)	(0.027)	(0.010)	(0.033)	(0.027)	(0.009)
2nd best performance task 2	-0.001	0.061*	0.045*	-0.001	0.060*	0.046*	-0.004	0.017	-0.022	-0.007	0.016	-0.021
	(0.014)	(0.029)	(0.022)	(0.014)	(0.028)	(0.022)	(0.054)	(0.042)	(0.016)	(0.053)	(0.042)	(0.014)
Constant	-0.029	0.234	0.654**	-0.026	0.198	0.677**	-0.177	0.610*	1.302**	-0.347	0.575	1.377**
	(0.151)	(0.318)	(0.243)	(0.151)	(0.309)	(0.239)	(0.471)	(0.366)	(0.136)	(0.469)	(0.376)	(0.129)
Observations	76	76	76	76	76	76	36	36	36	36	36	36
R-squared	0.024	0.126	0.113	0.025	0.178	0.147	0.102	0.026	0.313	0.157	0.030	0.412

Notes. *** p<0.001, ** p<0.01, * p<0.05. Seemingly Unrelated Regressions (SUR) with standard errors in parentheses. For the ease of result interpretation, SURE results are displayed. Logit and probit regressions with seemingly unrelated estimation lead to similar results. The dependent variable Reason takes the value 1 if chosen by the advisor and 0 otherwise. Reason refers to Preference for competition, Self-confidence, and Risk of earnings in regressions (a), (b), and (c), respectively. The sample is the GB data set and advice "piece rate" for SUR 1 and 2, and the GB data set and advice "tournament" for SUR 3 and 4.

Table C2*: SUR Regressions of Reasons Given by Advisors (sample is the GB data).*