Gender differences in reactions to feedback and willingness to compete

1

Gender Differences in Reactions to Feedback and Willingness to Compete

Noémi Berlin[†], Marie-Pierre Dargnies[‡]

July 2016

Abstract

In Western societies, it is generally known that men have a greater taste for competition than women. However, the determinants of the decision to enter competitions are still not fully understood. The aim of this paper is twofold. We first evaluate how participants update their beliefs after receiving feedback informing them of whether their performance is below or above the median performance. Second, we are interested in how men and women react to this information in terms of competitive entry. Our first result is that participants, and women in particular, react more strongly to the feedback they receive than would a Bayesian agent. As far as entry into competition is concerned, below-median participants adjust their entry decision according to the competition they expect to face, while above-median participants do not. However, the behavior behind these results is quite different for men and women: women mainly react to information on their own performance, while men seem to respond more to their beliefs over the competition they will face. Moreover, most of the effect of feedback and the information regarding the level of the competition on the decision to compete seems to operate via beliefs.

Keywords: Experimental economics, beliefs, performance feedback, gender, competition.

JEL classification: C9, D8.

*This work was financially supported by Paris School of Economics and the Social Science Research Centre in Berlin (WZB). We are grateful to Maxim Frolov for programming the experiment. We would also like to thank Muriel Niederle, Louis Lévy-Garboua, Marie-Claire Villeval, Luis Santos-Pinto, Nicolas Jacquemet, Jean-Christophe Vergnaud, Susan Laury, Dorothea Kuebler, Jennifer Rontganger, and the participants to numerous seminars and conferences (WZB and Paris 1 internal seminars, European ESA 2011, SABE 2012, ADRES Doctoral conference 2014, ASFEE 2014). Finally, we would like to thank the referees and editor whose comments allowed us to greatly improve the manuscript.

[†]University of Edinburgh, School of Economics, 31 Buccleuch Place, EH89JT Edinburgh, UK. E-mail: Noemi.Berlin@ed.ac.uk.

[‡]Corresponding author. Univ Paris-Dauphine, PSL Research University, UMR [7088], DRM, [Finance], 75016 PARIS, FRANCE. E-mail: marie-pierre.dargnies@dauphine.fr.

1 Introduction

There are many possible explanations for the continuing gender differences in labor market outcomes. Differences in preferences are frequently cited to explain this phenomenon (Croson and Gneezy, 2009). A rapidly growing literature studies more specifically gender differences in competitiveness (starting with Gneezy et al., 2003, Niederle and Vesterlund, 2007). Understanding beliefs and the way information about relative performance is processed is crucial to explain the surprisingly robust gender gap in self-selection in competitions. Indeed, we base our decisions to enter competitive environment to a great extent on our beliefs about our relative performance and we update these beliefs as we get, mostly noisy, feedback about how we perform in comparison to others. Not only is the belief-updating process worthy of being carefully looked at, the consistency between beliefs and actions is also decisive.

The goal of this experimental paper is twofold. First, we study how men and women update their beliefs following the reception of relative performance feedback. Second, we look at how men and women react in terms of tournament entry decisions to both this feedback and information about the level of competition.

In our set-up, subjects decide in two rounds whether to enter a tournament or to be paid according to a piece rate. We focus on a rather stereotypical-male task for which men are known to self-select into competition more often than women do. In the first decision round, the subject knows the opponent will be randomly-selected amongst all other participants and will therefore be of totally unknown ability. After the participants have made this first decision and performed the task, they receive a binary feedback telling them whether a past performance also based on a tournament was above or below the median in their session.[1] We have two treatments allowing us to manipulate the degree of competition our subjects face. In the *Ability Group* treatment, the second decision round requires subject to decide whether to enter a competition knowing that their opponent will be randomly selected among participants belonging to the same performance group as their own. In the

[1] Contrary to Wozniak et al. (2011), who provide an exact performance feedback based on piece rate, we provide a binary performance feedback based on the tournament.

Repetition treatment, the second round requires subjects to decide once again whether to enter a competition with an opponent of a totally unknown performance level. We elicit beliefs both before and after subjects receive their performance feedback. We can then study how beliefs are updated. We also look at how beliefs and the way they are updated affect the tournament entry decision. While the literature studied the effect of performance feedback on competitive entry (Cason et al., 2010, Wozniak et al., 2011), our paper is, to our knowledge, the first attempt to directly manipulate the level of competition participants are involved in. It allows us to study the combined effect of feedback and information on competition level while carefully monitoring beliefs about relative performance.

Our first result is that subjects update their beliefs following performance feedback more drastically than a Bayesian agent would. Both men and women are more pessimistic than a Bayesian agent following a below-median feedback; we find the opposite effect after an above-median feedback. Below-median women update even more pessimistically than their male counterparts when controlling for their actual performance level. We also show that below-median participants adapt their tournament entry decision to the ability level of the competition, while above-median participants do not.

However, men and women do not react to the feedback in terms of competitive entry in the same way. While women are especially sensitive to information on their own performance, men react stronger to the level of their competitors. Below-median men seem to take into account the possibility that their performance will improve over time, in the *Repetition* treatment, that is when they receive the feedback but there is no change in the level of competition. This is not the case for women. In other words, below-median women consider their performance level *per se* while below-median men think there is room for improvement.

Regarding the efficiency of choices in terms of expected payoffs maximization, men and women do not make the same kind of mistakes. Men enter the tournament too often when they should not while women do not enter enough when they should. Men and women make as many mistakes to start with. After a below-median feedback, men depart slightly more from the payoff-maximizing situation than women. When our participants both receive

a below-median feedback and face an ability group tournament, men made much smaller mistakes than women. Men, furthermore, improve the quality of their decisions in this last situation.

This paper contributes to the literature studying the gender wage gap and the extreme overrepresentation of men in positions perceived as 'prestigious'. Men have often been found to have a greater taste for competition than women (Gneezy et al., 2003, Niederle and Vesterlund, 2007, Datta Gupta et al., 2013, Niederle and Vesterlund, 2011), regardless of whether this taste is measured based on the decision to enter competitions or based on the performance in a competition imposed on all subjects.[2]

A number of papers tackle how subjects update their beliefs about their relative performance following the reception of performance feedback. Wozniak et al. (2011) provide experiment participants with precise ratings on how other participants performed in the piece rate. While there is a significant gender gap in tournament entry without feedback, it disappears when feedback is provided. Indeed, high-ability women choose more competitive compensation schemes and low-ability men choose less competitive compensation schemes with feedback than without it. Möbius et al. (2013) provide their subjects with noisy feedback via a simple binary signal for a performance in the top 50% and look at the belief updating. They find that subjects update their beliefs about their IQ being in the top 50% to a lesser degree than Bayesian agents would in response to both positive and negative signals, and women update less severely than men do. They also show that subjects react more to positive than to negative information (and there is no gender difference in this respect). Kuhnen and Tymula (2012) show that in competitive settings productivity and beliefs are influenced by privately observed information about relative rank. A number of papers (Möbius et al., 2013, Ertac, 2011, Grossman and Owens, 2011) find that individuals deviate from Bayesian beliefs more in self-relevant contexts (i.e., when they have to evaluate their own relative performance) than in self-irrelevant contexts (i.e, when they have

[2]However, in matrilineal societies the gender gap in tournament selection is inverted (i.e., women select the tournament more often than do men Gneezy et al., 2008). It is also affected by whether the task is stereotypical-male or stereotypical-female, and the level of pressure under which it is performed (Shurchkov, 2012, Gunther et al., 2010).

to evaluate somebody else's relative performance or update their beliefs about a neutral event). In our case, we differ from these papers in that our feedback is not noisy (contrary to Möbius et al., 2013, Grossman and Owens, 2011) and is based on tournament performance, and subjects are asked to assess their beliefs on their relative performance in a competitive context (while Ertac (2011) uses a task remunerated under piece rates). Closer to our setting, Murad et al. (2015) analyze how high and low performers of a group in one stage of a tournament change their confidence levels (as measured by the subjective probability they assign to being in the top half) in the next stage when they are re-grouped with other high or low performers. Their main finding is that low performers become more underconfident, while high performers become more overconfident.

Regarding the second aspect of our paper, how performance feedback influences men and women's actions, Camerer and Lovallo (1999) find that subjects considering themselves to be trivia experts were more confident when competing against each other on trivia questions than against subjects who did not consider themselves to be especially knowledgeable on the topic. The authors assume this shows evidence of a tendency to underadjust to changes in the reference group one competes against which they name the reference group neglect. Azmat and Iriberri (2010) find that providing feedback on relative performance to high-school students improves their grades by 5% regardless of where they are in the distribution. In the laboratory, this feedback-performance effect is only found for men (Azmat and Iriberri, 2012). Gill and Prowse (2010) show that men and women react differently following a loss (negative feedback) in a competitive environment. While women tend to reduce effort, independently of the monetary value of the prize, men, on the other hand, reduce effort only after failing to win large prizes. Möbius et al. (2013) first find a causal relationship between confidence (as instrumented by the number of positive feedbacks received) and tournament entry. Second, while positive feedback does increase competitiveness, this effect shrinks and becomes insignificant once they control for reported beliefs.

In this paper, we obtain results on how feedback and information on the level of the competition impacts both beliefs of winning the tournament and tournament entry. Our

experimental design allows us to address these two points simultaneously.

The remainder of the paper is organized as follows. Section 2 describes the experimental design. Section 3 reports the results on changes in performance and beliefs, adaptation to the level of competition, reaction to feedback, and expectations of improvement in performance. We also perform welfare analysis in terms of efficiency. Section 4 discusses our results and concludes.

2 Experimental design

We use a real-effort task consisting of solving as many sums of five two-digit numbers as possible (Niederle and Vesterlund, 2007) within five minutes. There are a total of six sequential steps,[3] one of which is randomly selected at the end of the experiment to determine the payoff, as well as incentivized belief-assessment questions. The *Repetition* and *Ability Group* sessions differ only in steps 4 and 4 prime (see below).[4] The same number of men and women take part in each session, and each of them includes at most 20 participants and lasts about 1 hour and 20 minutes. Participants are aware at the beginning of the session of the number of steps but only know the content of each step prior to its start; instructions are given step by step.

Step 1: Piece rate (PR) remuneration scheme. Subjects have five minutes to solve as many sums as they can, and earn 0.50€ per correct sum.

Step 2: Standard tournament (ST). Subjects have five minutes to solve as many sums as they can. They are randomly paired with another player in their session (whose gender is unknown to them). If step 2 is randomly chosen for remuneration, the winner

[3]The fact that the tasks are completed sequentially may have an effect on subjects' decisions to compete, as learning and fatigue may affect these decisions. However, as this paper focuses on gender effects and differences across treatments, the relevant question is whether men and women or participants randomly assigned to different treatments, are affected in a different way by the sequential format of the tasks; this is, in our opinion, unlikely.

[4]The instructions read to the subjects can be found in the supplementary materials.

in each pair (with the best step 2 performance) earns 1€ per correct sum, the loser receives nothing.

First round of belief elicitation: After the second step, the participants have to evaluate the probabilities that their step 2 performance belong to each of the four performance quartiles, with respect to the other participants in her session.[5] The sum of these four probabilities (in %) is equal to 100.[6] They answer four questions corresponding to the four quartiles. For example, the question for the fourth quartile was: "What is, according to you, the probability in % that your step 2 performance belongs to the fourth quartile (being in the 25% best performers)?"

To incentivize the answers, we use a confidence rule (Möbius et al., 2013, Hollard et al., 2010): for each of the four answers, the computer randomly picks a number y between 0 and 100. Let x_i be the subject's answer for quartile i ($i = 1, 2, 3, 4$).

- If $x_i > y$, the subject earns 1€ if her score belongs to the i^{th} quartile and zero otherwise.
- If $x_i < y$, the subject earns 1€ with y% probability.

Step 3 (hereafter Decision 1): Before solving the sums, subjects have to choose between the Piece Rate (PR) and Standard Tournament (ST) remuneration schemes. Those who choose PR receive 0.50€ per correct sum if step 3 is picked at the end of the experiment. If a subject chooses the tournament, she is randomly paired with another subject and wins the tournament (which pays 1€ per correct sum) if her step-3 performance is greater than her opponent's step 2 performance.[7]

[5] We chose to elicit the beliefs about step 2 performance, which is paid under a tournament remuneration scheme, because we are primarily interested in subjects' beliefs about winning the tournament. They have no idea beforehand that they will be asked to state their beliefs.

[6] The subject is asked to enter four beliefs: one for each quartile. If the sum of these four beliefs does not amount to 100, the participant cannot go on to the next step and has to enter new beliefs adding up to 100.

[7] As the randomly-chosen opponent may not have chosen the tournament, the step 3 performance of the participant is compared to the step 2 performance of the opponent when he was performing in a tournament. This way, the decision to enter the tournament is not affected by beliefs about whether the opponent is going to enter. In addition, it allows us to rule out the possibility that a participant may not enter because she does not want to inflict a loss on her opponent.

Step 3 prime (hereafter Decision 1 prime): Participants have to choose between *submitting* their step 1 performance to PR or ST. They do not have to solve sums at this step. The payoffs depend only on their step 1 performance. If a subject chooses to submit this performance to the tournament, she is randomly paired with another participant and earns 1€ per correct sum if her step 1 performance is greater than her opponent's. If she chooses PR, the remuneration is the same as in step 1. The reason why the step 1 performance is used is because we want to see the role played by confidence in one's relative ability, risk, ambiguity, and feedback aversion in the decision to enter a tournament when no tournament performance is involved. In consequence, one's decision to enter the tournament cannot be explained by the belief that performing under a tournament remuneration scheme will affect one's performance.

Feedback: Each participant receives feedback on their step 2 performance, telling them whether it was above or below the median.[8]

Second round of belief elicitation: This second round allows us to analyze how subjects update their beliefs after receiving an ability signal. They have to re-estimate the probabilities that their step 2 performance was in the two possible quartiles consistent with their feedback (the fourth and third quartiles for performers above the median, the second and first quartiles for performers below the median). We use the same incentive rule as in the first round, for both of the elicited beliefs.

Step 4 *Repetition* or *Ability Group* (hereafter Decision 2): After making their decision, participants have five minutes to solve as many sums as they can.

- In *Repetition* sessions, step 4 is exactly the same as step 3 (decision between PR or ST, the remuneration rule stays the same).
- In *Ability Group* sessions, subjects have to again choose between a piece rate and a tournament. If the piece rate is chosen, the subject earns 0.50€ per correct sum. But if she chooses the tournament, she is randomly matched to another

[8]Participants do not know they will receive a feedback until they reach this step.

participant who belongs to the same ability group. That is, if her step 2 performance was below (above) the median she is paired with someone whose step 2 performance was below (above) the median as well. We call this the "ability group tournament". A subject wins if her step 4 score is greater than her opponent's step 2 score, when the opponent is in the same ability group. In this case she wins 1€ per correct sum and zero otherwise.

Step 4 prime (hereafter Decision 2 prime): Participants have to choose between *submitting* their step 1 performance to a piece rate or a tournament. In *Repetition* sessions, step 4 prime has the same features as step 3 prime. In *Ability Group* sessions, the decision to enter competition leads to an ability group tournament such that both of the step 1 performances compared belong to the same ability group. The remuneration rule is the same as in step 3 prime.

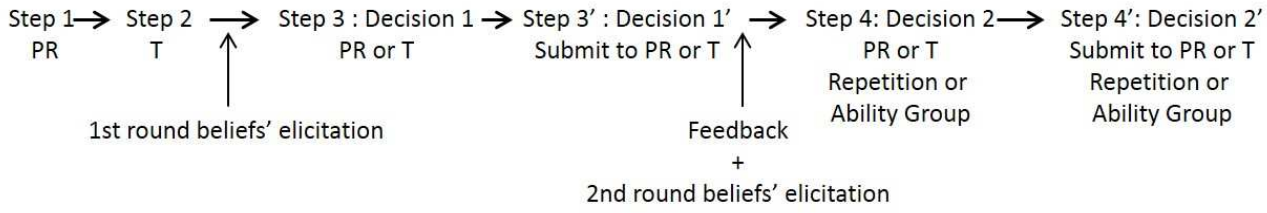
Steps 3 prime and 4 prime are respectively identical to steps 3 and 4[9] except for the fact that they do not involve a future performance. In particular, the participant who chooses to *submit* her past performance to the tournament does not have to perform under the pressure of competition. As a consequence, any changes in behavior between steps 3 and 3 prime (or 4 and 4 prime) will be attributed to the taste for performing in a competitive environment. On the contrary, if participants have the exact same behavior in steps 3 and 3 prime (and 4 and 4 prime), this means that tournament entry decisions are driven by confidence about one's relative performance and aversion to risk, ambiguity, and feedback.

At the end of each step, all participants are told their absolute performance, that is, the number of sums they solved. To the extent that order effects play the same role in both treatments, they should not explain our results. At the end of the experiment, a screen summarizes the earnings from each step and whether a tournament, if chosen by the participant, was won or lost. One step is randomly chosen to determine part of the payoff, in addition to the belief-assessment questions. Figure 1 summarizes all the steps in the

[9]In both cases, the tournament is a riskier decision, implying more ambiguity. Participants only discover whether they have beaten their opponent at the very end of the experiment.

experiment.

Figure 1: **Experiment summary**



The experiment was run at the "*Laboratoire d'Economie Expérimentale de Paris*" (LEEP) between February and April 2011. The same number of men and women took part in each session. Respectively 112 subjects (56 men and 56 women) and 116 subjects (58 women and 58 men) participated in the *Repetition* and *Ability Group* sessions. One step was randomly chosen at the end of the experiment to be paid in addition to the belief-assessment questions and a 7€ show-up fee. Participants earned 15.30€ on average (see Table 9 in the Appendix for detailed sample characteristics).

3 Results

In subsection 3.1, we look at the changes in performance across steps and confidence assessments. Subsection 3.2 focuses on whether our participants are Bayesian updaters. Subsection 3.3 is interested in whether our participants adapt their tournament entry decision to the information received about both their own performance and that of the opponent they will face. In subsection 3.4 we study the pure effect of feedback in the Repetition group and whether this is different across gender. Finally, subsection 3.5 analyzes the welfare implications of the behaviors we observe.

3.1 Changes in performance and confidence assessments

We start by analyzing how performance alters between steps 1 and 2, that is, when the remuneration scheme changes from being piece rate to tournament.

Table 1 shows the average number of correctly-solved sums by gender in step 1 (piece rate) and step 2 (tournament). Participants perform significantly better in the tournament than in the piece rate. This suggests that the remuneration scheme affects performance, even though learning may also play a role here. We do not find any performance differences between men and women. This is in line with Niederle and Vesterlund (2007) and Ertac (2011).

Table 1: **Average number of correctly-solved sums at step 1 and step 2 (the p-values correspond to two-tailed t-tests).**

	Step 1 Performance (PR)	Step 2 Performance (T)	Diff
Men	7.8	9.2	p=0.00
Women	7.5	8.9	p=0.00
Diff	p=0.48	p=0.46	

We pick up participants' confidence in their chances of tournament success via a question on their beliefs about their relative performance; this is asked both before and after they receive feedback telling them whether their performance was above or below the median. Before receiving feedback, participants had to state their beliefs that their step 2 performance fell in each of the four performance quartiles; after receiving feedback, they had to assess their beliefs that their performance would fall in each of the two forthcoming quartiles (as they now knew that their performance was either above or below the median).

Before receiving feedback, below-median[10] men were not significantly more confident than below-median women. The beliefs of belonging to the second as opposed to the first (worst) quartile of their session represent respectively 70.2% and 69.4% of the beliefs of

[10]Below-median participants are those whose step 2 performance was below the median step 2 performance in their session, and who therefore received the "below-median" feedback after the first round of confidence-assessment questions.

being below the median: the two-sided Mann-Whitney test yields $p = 0.88$. However, after receiving feedback that their performance was below the median, below-median men were more confident than below-median women regarding their chances of belonging to the second quartile (below-median men and women respectively thought they were 67.3% and 57.5% likely to belong to the second quartile: the two-sided Mann-Whitney test yields $p < 0.01$).

Before receiving feedback, above-median men were more confident than above-median women.The beliefs of belonging to the fourth (best) as opposed to the third quartile of their session represent respectively 53.0% and 40.0% of the beliefs of being above the median (this difference is significant in a two-sided Mann-Whitney test with $p < 0.01$). Once they learned that their performance was above the median, men were still more confident than women, but to a lesser degree: above-median men and women believed their performance had respectively a 60.6% and 51.4% ($p = 0.03$) chance of belonging to the fourth (top) quartile.

The above results suggest that men and women do not react in the same way to performance feedback, with women adjusting more strongly to both below-median and above-median feedbacks than men. In particular, women become less confident after below-median feedback and more confident after above-median feedback.

3.2 Are the participants bayesian updaters?

To further our investigation, we calculate for each subject the beliefs she would have held during the second round of confidence-assessment questions (i.e, after performance feedback) had she updated her first-round beliefs in a Bayesian way. We later refer to these beliefs as "Bayesian beliefs". They are calculated as follows.

We denote by b_{1_j} the first-round elicited beliefs about belonging to the j^{th} quartile ($j = 1, .., 4$). We differentiate Bayesian beliefs for below-miedan performers ($Bayes_{below}$) and above-median performers ($Bayes_{above}$) such that:

- $Bayes_{below} = 100 * \frac{b_{1\_2}}{b_{1\_2}+b_{1\_1}}$
- $Bayes_{above} = 100 * \frac{b_{1\_4}}{b_{1\_4}+b_{1\_3}}$

For instance, consider a subject who holds the following beliefs of her performance belonging respectively to the 4th, 3rd, 2nd, and 1st quartiles: 30%, 20%, 25% and 25%. This subject receives the feedback that her performance is above the median, meaning it belongs either to the fourth or third quartile. If this subject updates her beliefs in a Bayesian way, she should now believe that there is a 60% chance (respectively a 40% chance) that her performance belongs to the 4th quartile (respectively the 3rd).

We then compare actual second-round beliefs to Bayesian beliefs. To continue our previous example, if during the second round of belief-elicitation questions, our subject answers that she thinks there is an 80% chance of her performance belonging to the best quartile (and therefore a 20% chance that it belongs to the third quartile), she will be considered more optimistic than a Bayesian updater.

Figure 2: **Beliefs update compared to Bayesian beliefs.**

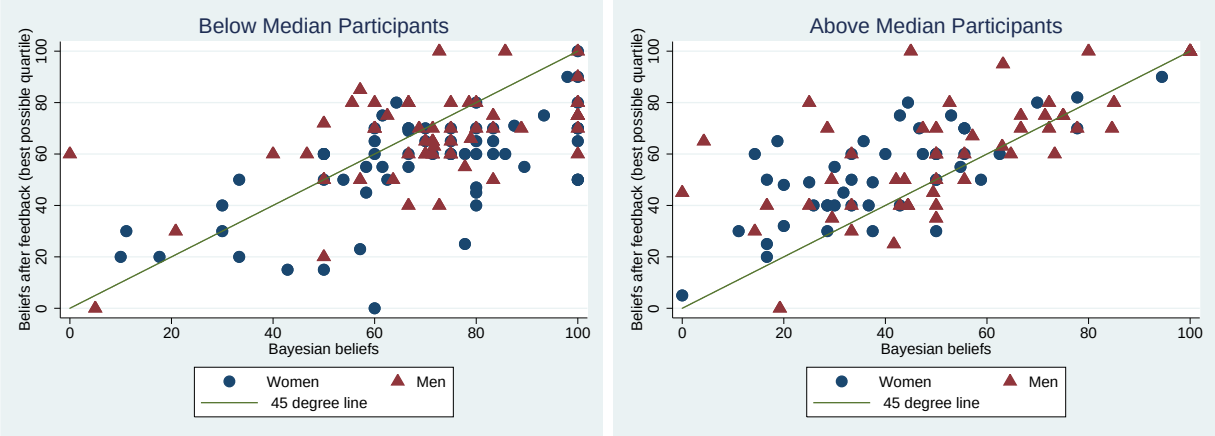


Figure 2 displays the Bayesian beliefs compared to the actual updated beliefs after receiving the feedback, for below- and above-median participants, broken down by gender. If our subjects were Bayesian updaters, beliefs should be situated on the 45 ° line (where Bayesian beliefs are equal to the elicited beliefs). However, we can see that overall, most of the below-median participants' beliefs are situated below the 45 ° line, meaning that they update more pessimistically than a Bayesian agent would. The opposite result is found for above-median participants.

More precisely, both men and women overreact to the feedback they receive. Above-median women's beliefs are significantly more optimistic than Bayesian beliefs (a Wilcoxon signed-rank test yields $p < 0.01$); this is also the case for above-median men ($p < 0.01$). Concerning below-median participants, women's beliefs are significantly more pessimistic than Bayesian beliefs (a Wilcoxon signed-rank test yields $p < 0.01$) while this holds to a lesser extent for men (p=0.04).[11]

Table 2: **OLS estimation on the second-round elicited beliefs**

VARIABLES	Below-median	Above-median
	(1)	(2)
Female	-9.393***	-0.939
	(2.750)	(3.362)
Performance in step 2	1.373*	0.759*
	(0.758)	(0.451)
Bayes	0.467***	0.629***
	(0.070)	(0.073)
Constant	23.978***	18.995**
	(6.783)	(7.896)
Observations	242	194
R-squared	0.408	0.521

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Female is a dummy variable equal to 1 for female participants. Perf_step2 is the performance during step 2, Bayes is the beliefs participants should hold if they were bayesian updaters. Standard errors are clustered at the individual level.

But do these results continue to hold after controlling for the original level of step 2 performance? We ran regressions of the second-round elicited beliefs on the Bayesian benchmark, a "Female" dummy and the step 2 performance, separately for below-median and

[11] Murad et al. (2015) find, in line with our results, that high performers (respectively low performers) become more overconfident (respectively underconfident) when they are re-grouped with other high performers (respectively low performers). Their findings cannot exactly be compared to ours though as they use a different measure of confidence (the subjective probability a subject assigns to being in the top half) and do not compute a bayesian benchmark. Furthermore, they do not find any gender difference in terms of confidence.

above-median participants in Table 2. The coefficient of "Female" is negative and significant ($p < 0.01$) for below-median participants, showing that when controlling for the performance level, below-median women update their beliefs more pessimistically than men following the reception of their feedback. After being informed that their performance is below the median, below-median women are more likely to think, as compared to a Bayesian updater, that their performance is in the weakest quartile. This same coefficient is not significant for above-median participants, showing that women do not update their beliefs differently than men following the reception of an above-median feedback when controlling for the performance level.

Result 1: *Both men and women overreact to the feedback they receive. Below-median women update even more pessimistically than their male counterparts when controlling for their actual performance level.*

3.3 Do participants adapt to the competition level?

3.3.1 Rough treatment effect

The answer to our main question can be seen by comparing the change in the tournament entry decision between decisions 1 and 2 in our two treatments (*Repetition* and *Ability Group*). Figures 3 and 4 illustrate these results. In both treatments, participants receive feedback between decisions 1 and 2 telling them whether their step 2 performance was above or below the median. However, only in the *Ability Group* does the opponent belong to the same performance group as the subject. If below-median participants adjust their tournament entry decisions to the level of the competition, we expect the decrease in tournament entry rate between decisions 1 and 2 to be lower in the *Ability Group* (where the level of the competition becomes lower) than under *Repetition* (where it remains the same). For above-median participants the increase in tournament entry rate should be greater under *Repetition* than in the *Ability Group* in order to prove that they adpat to the level of the competition.

We compute the diff-in-diff estimators of the choice to enter the tournament for Decision 1 vs. Decision 2 and *Repetition* vs. *Ability Group* by estimating the following equation:

Figure 3: **Proportion (%) of below-median women (left) and men (right) choosing tournament entry in Decision 1 and Decision 2.**

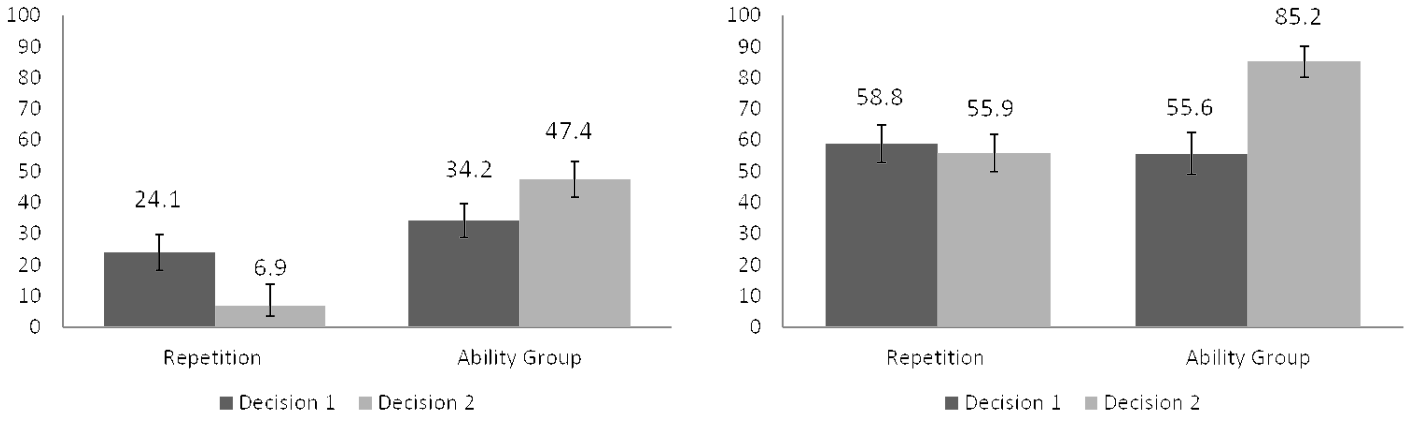
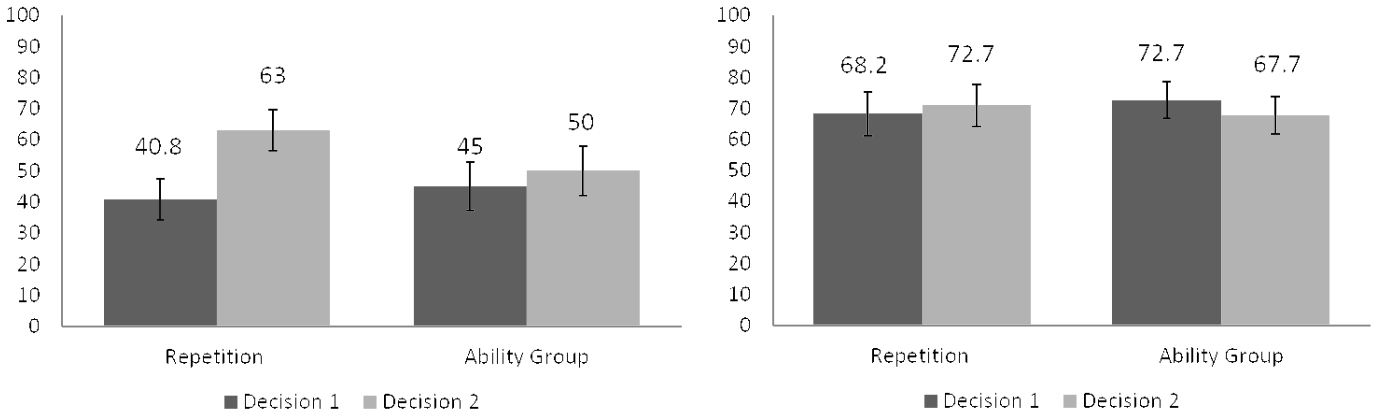


Figure 4: **Proportion (%) of above-median women (left) and men (right) choosing tournament entry in Decision 1 and Decision 2.**



$$Tourn_{it} = \beta_0 + \beta_1 AbGr_{it} + \beta_2 Decision2_{it} + \beta_3 AbGr * Decision2_{it} + \epsilon_{it}$$

Here $Tourn$ is a dummy variable equal to 1 if subject i entered the tournament. $AbGr$ is the dummy treatment variable for the subject taking part in one of the *Ability Group* treatment's

sessions. *Decision*2 is a dummy variable equal to 1 if this is the second decision a participant makes, i.e, after the feedback. The diff-in-diff estimate is thus given by the coefficient β_3 on the interaction term $AbGr * Decision2$. These estimates are positive and significant for both below-median women and men: the coefficients are respectively 0.30 (p=0.04) and 0.33 ($p = 0.02$), indicating a treatment effect for those subjects in that they adapt their entry decision to the competition level (see columns (1) of Table 3). Regarding above-median participants, the diff-in-diff estimators are negative (respectively -0.17 and -0.08 for above-median women and men) but not significant (respectively $p = 0.23$ and $p = 0.54$) (see columns (1) of Table 4). Above-median participants do not, therefore, increase tournament entry significantly more under *Repetition* (where the level of the competition remains the same) than in the *Ability Group* (where it is higher).

Result 2: *Below-median participants adapt their decision to enter a tournament to the level of the competition. Above-median participants do not.*

3.3.2 How beliefs explain competitive behavior

In order to further our investigation, we want to assess the role of beliefs in explaining competitive behavior. We begin by studying whether the reaction to feedback and tournament entry could depend on how confident a participant is to start with. Columns (2) of tables 3 and 4 show the impact of initial beliefs on participants' tournament entry decisions. *BeliefAbove* corresponds to the belief, before receiving feedback, of being above the median. In Table 3 (which corresponds to the sample of below-median participants), the coefficient of the interaction term BeliefAbove*Decision2 is negative and significant at the 10% level for below-median women (not for men), suggesting that if a participant had a high belief of being above the median but receives a below-median feedback, she has a lower probability of entering the tournament when having to decide for the second time. Moreover, the addition of BeliefAbove and BeliefAbove*Decision2 to the explanatory variables leaves the coefficient of ability group*Decision2 basically unchanged for both below-median men and women. We will study this effect separately in our treatment groups in the following sections. We now

consider the case of above-median subjects. In columns (2) of Table 4, the coefficient on BeliefAbove*Decision2 is negative and significant at the 1% level for women but not for men. If a participant starts with a low confidence level, the reception of the above median feedback increases her probability of entering the tournament the second time. Together with our previous result about below-median women in Table 3, this suggests that women are prone to what we call a "surprise effect". Women are more likely to change their competitive behaviour if they receive a feedback they did not expect: an above-median (below-median) feedback tends to increase (decrease) more the competitiveness of women who did not expect it. This effect is not found for men for whom the beliefs held before feedback have no effect on their tournament entry.

Result 3: *Women react more strongly to feedback when they did not expect it. We call this the "surprise effect".*

Second, we create the variable "beliefWin" as a proxy for the subject's beliefs concerning tournament success (section 5.1 in the Appendix explains in more details the construction of this variable). Remember that we elicited beliefs twice, both before and after the feedback was provided. We denote by b_{i_j} the beliefs elicited at round $i = \{1, 2\}$[12] of their performance being in the $j = \{1, 2, 3, 4\}$th quartile. We make the additional assumption here that whenever a subject thinks her performance belongs to a given quartile, she actually thinks it lies at the exact midpoint of this quartile[13]. During the first round of beliefs elicitation, we get the subjects' beliefs about belonging to each of the four quartiles and define the belief of winning the Decision 1 (step 3) tournament as follows: $beliefWin = 0.875 * b_{1\_4} + 0.625 * b_{1\_3} + 0.375 * b_{1\_2} + 0.125 * b_{1\_1}$, where b_{1_4} is the subject's belief, before feedback, that her performance belongs to the best quartile. If she holds such beliefs, we consider that she thinks she will beat 87.5% of her potential opponents, who are all the subjects from her session, that is, the 75% of subjects whose performance belongs to the 1st, 2nd, and 3rd quartiles plus half of the subjects whose performance belongs to

[12] $i = 1$ before receiving the performance feedback, $i = 2$ after.

[13] E.g, if I think my performance belongs to the best quartile, we make the assumption that it means that I think 12.5% of Step 2 performances were better than mine and 87.5% of step 2 performances were lower than mine

the fourth quartile. The same reasoning is used to find the coefficients of b_{1_3}, b_{1_2} and b_{1_1}. In order to compute "beliefWin" for the Decision 2 tournament, we use the second-round beliefs elicited after the feedback is provided ($i = 2$). Because the level of the competition changes in *Ability Group* and hence the probability for a participant of winning a tournament, $beliefWin = 0.75 * b_{2\_2} + 0.25 * b_{2\_1}$ for both below-median and above-median participants. In *Repetition*, $beliefWin = 0.375 * b_{2\_2} + 0.125 * b_{2\_1}$ for below-median subjects. Following the same reasoning, $beliefWin = 0.875 * b_{2\_2} + 0.625 * b_{2\_1}$ for above-median subjects. This measure is not perfect as it does not take into account the possibility that the subject might think he will improve between steps 2 and 3 but we still consider it a reasonable proxy.

Column (3) of Table 3 shows that when beliefWin is added to the regressors, the coefficient on Ability Group*Decision2 becomes lower and insignificant for below-median participants of both genders. The adjustment of competitive entry to the level of their opponents is thus mainly driven by beliefs. The lower confidence of below-median participants following a below-median feedback explains their diminished desire to enter the tournament thereafter. The introduction of "beliefWin" into the regressors in column (3) of Table 4 does not change the coefficient on Ability Group*Decision2 which remains insignificant in both the above-median male and female regressions. For both genders and ability levels, the coefficient of the estimate of "beliefWin" is positive and significant showing that an increase in the beliefs of winning the tournament will result in a higher probability of entering the tournament.

Finally, we intend to see if the observed action of entering tournament is consistent with the reaction in terms of the belief of winning the tournament. We aim at answering questions such as: entry results show us that below-median participants enter more in the Decision 2 tournament when they know their opponent will also be of low ability than when she will be randomly chosen among all participants in the session. Is it also the case that below-median participants realize that they have a higher chance of winning the Decision 2 tournament when they know their opponent will also be of low ability than when she will be randomly chosen among all participants in the session? Formally, columns (1prime) of tables 3 and 4 correspond to the "mirror" regressions of the column (1) i.e we regress "beliefWin" on

"Ability Group", Decision 2 and the interaction term AbilityGroup*Decision 2, separately for men and women and for below-median and above-median subjects (similar to what we do for the entry decisions). We are mostly interested in the sign and significance of the interaction terms. It turns out to be positive and highly significant ($p < 0.01$) for both below-median men and women. It shows that, in line with their entry decision, below-median participants realize that they are more likely to win the Decision 2 tournament when they compete against another below-median participant than when their opponent is randomly chosen among all the participants from their session.

As far as above-median participants are concerned, the coefficient of the interaction term is negative and highly significant for both men and women ($p < 0.01$ both times). Above-median participants think that they are less likely to win the tournament, despite what their entry decisions suggest.

Table 3: **LPM for the tournament entry decision of below-median women and men**

	Below-median women				Below-median men			
VARIABLES	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)
Ability group	0.101	-1.976	0.106	0.100	-0.033	-7.531**	0.059	0.064
	(0.113)	(3.475)	(0.105)	(0.107)	(0.130)	(3.590)	(0.130)	(0.128)
Decision 2	-0.172*	-24.871***	0.113	0.040	-0.029	-26.618***	0.119	0.311**
	(0.100)	(2.207)	(0.149)	(0.106)	(0.079)	(2.144)	(0.145)	(0.126)
Ability group*Decision 2	0.304**	26.869***	0.300**	0.093	0.326**	34.831***	0.295**	-0.120
	(0.148)	(2.832)	(0.144)	(0.162)	(0.131)	(3.029)	(0.132)	(0.189)
BeliefAbove			0.007***				0.007***	
			(0.002)				(0.002)	
BeliefAbove*Decision 2			-0.005*				-0.002	
			(0.003)				(0.002)	
beliefWin				0.009***				0.013***
				(0.003)				(0.004)
Constant	0.241***	52.543***	-0.119	-0.207	0.588***	56.522***	0.140	-0.135
	(0.081)	(2.546)	(0.098)	(0.152)	(0.086)	(2.537)	(0.184)	(0.237)
Observations	134	134	134	134	122	122	122	122
R-squared	0.102	0.441	0.170	0.147	0.060	0.526	0.157	0.143

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Ability group equals 1 for an *Ability Group* session, 0 for a *Repetition* session. Decision 2 is a dummy equal to 1 for the second decision to enter the tournament i.e, after feedback. BeliefAbove is a continuous variable corresponding to the stated belief of being above the median before receiving feedback. Tourn. women and Tourn. men refer to the regression of the decision to enter the tournament of respectively men and women. beliefWin women and beliefWin men refer to the regressions of beliefWin (the proxy for one's belief in her chances of winning the tournament). Standard errors, in parentheses, are clustered at the individual level.

Table 4: **LPM for the tournament entry decision of above-median women and men.**

	Above Median women				Above Median men			
VARIABLES	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)
Ability group	0.043	-1.737	0.054	0.058	0.028	2.803	0.019	0.001
	(0.150)	(3.923)	(0.142)	(0.141)	(0.132)	(4.718)	(0.130)	(0.127)
Decision 2	0.222*	17.926***	0.880***	0.062	0.045	14.716***	0.135	-0.097
	(0.112)	(2.388)	(0.201)	(0.129)	(0.104)	(2.904)	(0.244)	(0.108)
Ability group*Decision 2	-0.172	-23.488***	-0.187	0.038	-0.078	-25.893***	-0.076	0.173
	(0.143)	(3.033)	(0.126)	(0.166)	(0.127)	(3.811)	(0.127)	(0.152)
BeliefAbove			0.008***				0.005	
			(0.002)				(0.003)	
BeliefAbove*Decision2			-0.010***				-0.001	
			(0.002)				(0.002)	
beliefWin				0.009**				0.010**
				(0.004)				(0.004)
Constant	0.407***	57.574***	-0.091	-0.108	0.682***	63.261***	0.327	0.068
	(0.097)	(3.048)	(0.140)	(0.243)	(0.102)	(3.849)	(0.264)	(0.313)
Observations	94	94	94	94	106	106	106	106
R-squared	0.031	0.443	0.119	0.069	0.002	0.272	0.053	0.080

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Ability group equals 1 for an *Ability Group* session, 0 for a *Repetition* session. Decision 2 is a dummy equal to 1 for the second decision to enter the tournament after feedback. BeliefAbove is a continuous variable corresponding to the stated belief of being above the median before receiving feedback. Tourn. women and Tourn. men refer to the regression of the decision to enter the tournament of respectively men and women. beliefWin women and beliefWin men refer to the regressions of beliefWin (the proxy for one's belief in her chances of winning the tournament). Standard errors, in parentheses, are clustered at the individual level.

3.4 Reaction to feedback

We could be tempted to conclude from these first results that men and women react similarly to the level of opponent they face. However, finding that below-median participants adjust their tournament entry decision to the level of competition could reflect two different phenomena. Subjects could be reacting either to the feedback only (*Repetition*) or also to the level of their opponent (*Ability Group*). Participants would be reacting to feedback if, for instance, the reception of a below-median feedback reduced their willingness to compete in *Repetition* but there was no difference in tournament entry between decisions 1 and 2 in the *Ability Group*, where they know that their opponent is also of low ability. On the other hand, if the reception of a below-median feedback did not change the tournament entry decision in *Repetition*, but did so in the *Ability Group* and participants enter more in the Decision 2 than in the Decision 1 tournament, then we would conclude that participants react more to the level of the competition *per se*.

We compute the diff-in-diff estimators for Decision 1 vs. Decision 2 tournaments and below-median vs. above-median for each treatment, by gender:

$$Tourn_{it} = \gamma_0 + \gamma_1 Above\_Median_{it} + \gamma_2 Decision2_{it} + \gamma_3 Above\_Median * Decision2_{it} + \epsilon_{it}$$

Above_Median is a dummy variable that corresponds to the feedback they receive and is equal to 1 for above-median participants and zero for the below-median ones. Results of these regressions appear in the columns (1) of Tables 5 and 6. For the *Repetition* group, the estimate of the coefficient on the interaction term of *Above_Median* and *Decision2* which corresponds to the diff-in-diff estimate for women, is positive and significant (the coefficient is 0.39, with p=0.01). It means that women increase tournament entry between Decision 1 (step 3) and Decision 2 (step 4) more following an above-median than a below-median feedback. The nature of the feedback therefore seems to have considerable impact on women's competitive decisions, which is in line with *result 3*. The coefficient of Above_median*Decision2 loses significance when "beliefWin" is added to the regressors in column (3) showing that

the effect of the nature of the feedback is driven by beliefs. As far as men are concerned, tournament entry decisions are not affected by the nature of the feedback received (the coefficient is 0.07, p=0.57).

For the *Ability Group* (results reported in Table 6), men respond differently according to whether they are above or below the median (the coefficient is -0.33, p=0.01), so that they react more to what the feedback implies for the level of the opponent they will face rather than the information on their own performance level. Again, this effect is mainly driven by beliefs. In this same group, the insignificant coefficient (p=0.56) for women underlines that they place more importance on the personal information than on the implication of their opponents' ability. When running the three regressions from Table 5 for *Repetition* women and men together, adding the triple interaction Decision2*Above_Median*Female, the coefficient of this variable is close to being significant (p=0.11, 0.12 and 0.14 respectively for regressions (1), (2), and (3)). It confirms that the effect of Decision2*Above_Median is somewhat different for *Repetition* men and women, i.e, *Repetition* men and women react differently to the type of feedback they receive.

Result 4: *While women react mainly to feedback on their own performance level, men respond more to the competitive environment and hence the level of their opponent.*

Table 5: **LPM for the tournament entry decision among *Repetition* participants by gender.**

	Repetition women				Repetition men			
VARIABLES	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)
Above_Median	0.166	5.031	0.097	0.093	0.094	6.739	0.031	0.028
	(0.126)	(3.967)	(0.114)	(0.113)	(0.133)	(4.608)	(0.137)	(0.140)
Decision 2	-0.172*	-24.871***	0.321**	0.186*	-0.029	-26.618***	0.029	0.230
	(0.101)	(2.215)	(0.137)	(0.110)	(0.079)	(2.148)	(0.156)	(0.145)
Decision2*Above_Median	0.395**	42.797***	0.466***	-0.223	0.075	41.334***	0.086	-0.328
	(0.150)	(3.249)	(0.148)	(0.182)	(0.130)	(3.609)	(0.142)	(0.200)
BeliefAbove			0.009***				0.005*	
			(0.002)				(0.003)	
BeliefAbove*Decision2			-0.009***				-0.001	
			(0.003)				(0.002)	
beliefWin				0.014***				0.010**
				(0.003)				(0.005)
Constant	0.241***	52.543***	-0.237**	-0.516***	0.588***	56.522***	0.267	0.037
	(0.081)	(2.556)	(0.099)	(0.175)	(0.086)	(2.542)	(0.191)	(0.280)
Observations	112	112	112	112	112	112	112	112
R-squared	0.193	0.716	0.307	0.302	0.019	0.689	0.086	0.075

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Above_Median is a dummy equal to 1 for subjects being above the median. Decision 2 is a dummy equal to 1 for the second decision to enter the tournament, after feedback. BeliefAbove is a continuous variable corresponding to the stated belief of being above the median before receiving feedback. Tourn. women and Tourn. men refer to the regression of the decision to enter the tournament of respectively men and women. beliefWin women and beliefWin men refer to the regressions of beliefWin (the proxy for one's belief in her chances of winning the tournament). Standard errors, in parentheses, are clustered at the individual level.

Table 6: **LPM for the tournament entry decision among *Ability Group* participants by gender.**

	Ability Group women				Ability Group men			
VARIABLES	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)	Tourn. (1)	beliefWin (1prime)	Tourn. (2)	Tourn. (3)
Above_Median	0.108	5.270	0.068	0.106	0.154	17.074***	-0.040	-0.054
	(0.138)	(3.415)	(0.138)	(0.141)	(0.128)	(3.726)	(0.160)	(0.148)
Decision 2	0.132	1.998	0.437**	0.142	0.296***	8.213***	0.445***	0.196*
	(0.109)	(1.781)	(0.210)	(0.112)	(0.105)	(2.142)	(0.150)	(0.112)
Above_Median*Decision2	-0.082	-7.561***	-0.040	-0.071	-0.329**	-19.390***	-0.248*	-0.092
	(0.140)	(2.575)	(0.134)	(0.145)	(0.128)	(3.264)	(0.143)	(0.156)
Belief_above			0.005**				0.007**	
			(0.002)				(0.003)	
Belief_above*Decision2			-0.006*				-0.003	
			(0.003)				(0.002)	
BeliefWin				0.004				0.012***
				(0.003)				(0.004)
Constant	0.342***	50.568***	0.047	0.135	0.556***	48.991***	0.198	-0.042
	(0.079)	(2.371)	(0.128)	(0.170)	(0.098)	(2.543)	(0.180)	(0.205)
Observations	116	115	116	115	116	116	116	116
R-squared	0.017	0.028	0.053	0.029	0.049	0.202	0.134	0.155

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Above_Median is a dummy for subjects being above the median. Decision 2 is a dummy equal to 1 for the second decision to enter the tournament after feedback (step 5). BeliefAbove is a continuous variable corresponding to the stated belief of being above the median before receiving feedback. Tourn. women and Tourn. men refer to the regression of the decision to enter the tournament of respectively men and women. beliefWin women and beliefWin men refer to the regressions of beliefWin (the proxy for one's belief in her chances of winning the tournament). Standard errors, in parentheses, are clustered at the individual level.

It should be remembered that those entering a tournament at a given step will win if their performance at this step is greater than the step 2 performance of their opponent. As such, if we expect performance to increase from one step to another, the probability of tournament success should be greater in later steps. The fact that men do not react to feedback in the *Repetition* treatment (i.e, when it only provides information about their own relative performance at step 2) could mean that the below-median men expect their performance to improve over time and more so than above-median men.

We have two ways of checking this hypothesis. We can first use the fact that in Decisions 1 prime and 2 prime participants decide whether or not to submit a past performance, which therefore cannot be changed, to a tournament. If a subject expects an improvement in her performance overtime, she should be more willing to enter the tournament than to submit a past performance to it. Second, we can use our "beliefWin" variable which was built with one's step 2 performance. "BeliefWin" therefore proxies one's confidence in her chance of winning a given tournament if her performance remains at its step 2 level. We therefore compare the results of the regressions of three dependent variables (decision to enter, decision to submit and beliefWin) on Above_Median, Decision 2 and Above_Median*Decision2. The results for men from *Repetition* are reported in Table 7. You can see that the coefficient of Decision 2 is negative and significant for the submission decision and beliefWin but is not significant for the entry decision. This means that men from *Repetition* whose step 2 performance was below the median understand that they are less likely to win the tournament after receiving a below-median feedback if their performance cannot be changed but they are not less likely to enter the tournament after receiving such a feedback. This indicates that they expect their performance to improve. Furthermore, Above_Median*Decision2 is positive and significant for the submission decision and beliefWin but is not significant for the entry decision. This suggests that below-median participants expect a bigger performance improvement than above-median participants. As far as women are concerned the results of the regressions of entry, submission and beliefWin are in line.[14]

[14]We chose to only report the one result such that the results of the "beliefWin" and submission decision are both consistent with one another and different from the result of the regression of the entry decision.

Table 7: **LPM for the tournament entry and submission decisions and beliefWin among *Repetition* men.**

	Repetition men		
VARIABLES	(Entry (1))	(Submission (1prime))	(Beliewfin)
Above_Median	0.094	-0.048	6.739
	(0.133)	(0.136)	(4.608)
Decision 2	-0.029	-0.176*	-26.618***
	(0.079)	(0.010)	(2.148)
Above_Median*Decision2	0.075	0.313**	41.334***
	(0.130)	(0.155)	(3.609)
Constant	0.588***	0.412***	56.522***
	(0.086)	(0.086)	(2.542)
Observations	112	112	112
R-squared	0.019	0.04	0.689

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Above_Median is a dummy for subjects being above the median. Decision 2 is a dummy equal to 1 for the second decision to enter the tournament after feedback (steps 4 and 4 prime). beliefWin refers to the regressions of beliefWin (the proxy for one's belief in her chances of winning the tournament). Standard errors, in parentheses, are clustered at the individual level.

Result 5: *Below-median men expect their performance to increase over time when deciding to enter a tournament. Women do not.*

3.5 Welfare analysis

We now turn to the consequences of competitive behavior on welfare. More precisely, we are interested in whether decisions maximized expected payoffs.

To calculate the expected payoffs from entering the standard tournament (i.e, the tournament where the opponent is randomly drawn from among all other participants in the session), 100,000 performances were drawn by sampling with replacement from the step 2

Only then do we have a clear pattern allowing us to think our participants anticipate an evolution of their performance.

performances of our 228 participants. For each level of performance, the probability of tournament success was computed by calculating the number of times out of 100,000 this given performance exceeded the opponent's performance. Similarly, for ability-grouping tournament success for below-median participants, 100,000 performances were drawn from the step 2 performances of the potential opponents, i.e., participants whose step 2 performance was also below the median. We then calculate, for each performance level, the number of times out of 100,000 this given performance exceeded the opponent's performance. The same method is used to compute ability-grouping tournament success for above-median participants. We then compare, for each performance level, the payoff from choosing the piece rate to that from entering the tournament. This tells us which participants would have maximized their payoffs by entering the tournament, and we compare this to participants' actual decisions.

Given the distribution of step 2 performances, all participants with a performance of nine or more have a higher expected payoff from standard tournament entry than from the piece rate. For instance, if participants expect their Decision 1 (step 3) performance to be the same as their step 2 performance, 53.5% of participants should enter the Decision 1 tournament. However, taking into account the actual Step 3 performances, which are slightly better than Step 2 performances, 61% of participants would have gained from choosing the tournament. In the same way, all below-median (above-median) participants with a performance at least equal to seven (twelve) should enter the Decision 2 (step 4) tournament in the *Ability Group* treatment. Table 8 shows the optimal choices (i.e. the choice between PR or Tourn. that would have maximized expected payoffs) of participants given their performance at every step for each tournament. Note that these are the optimal choices under the assumption that the performance of any participants would have been the same under the two possible remuneration schemes.

We now compare the observed choices to the optimal ones. More precisely, for each decision step we split our participants into four categories: (1) those whose optimal decision was to enter the tournament and who did choose the tournament, (2) those whose optimal

Table 8: **Optimal choice between piece-rate (PR) or tournament (Tourn.) for each decision step depending on the performance level**

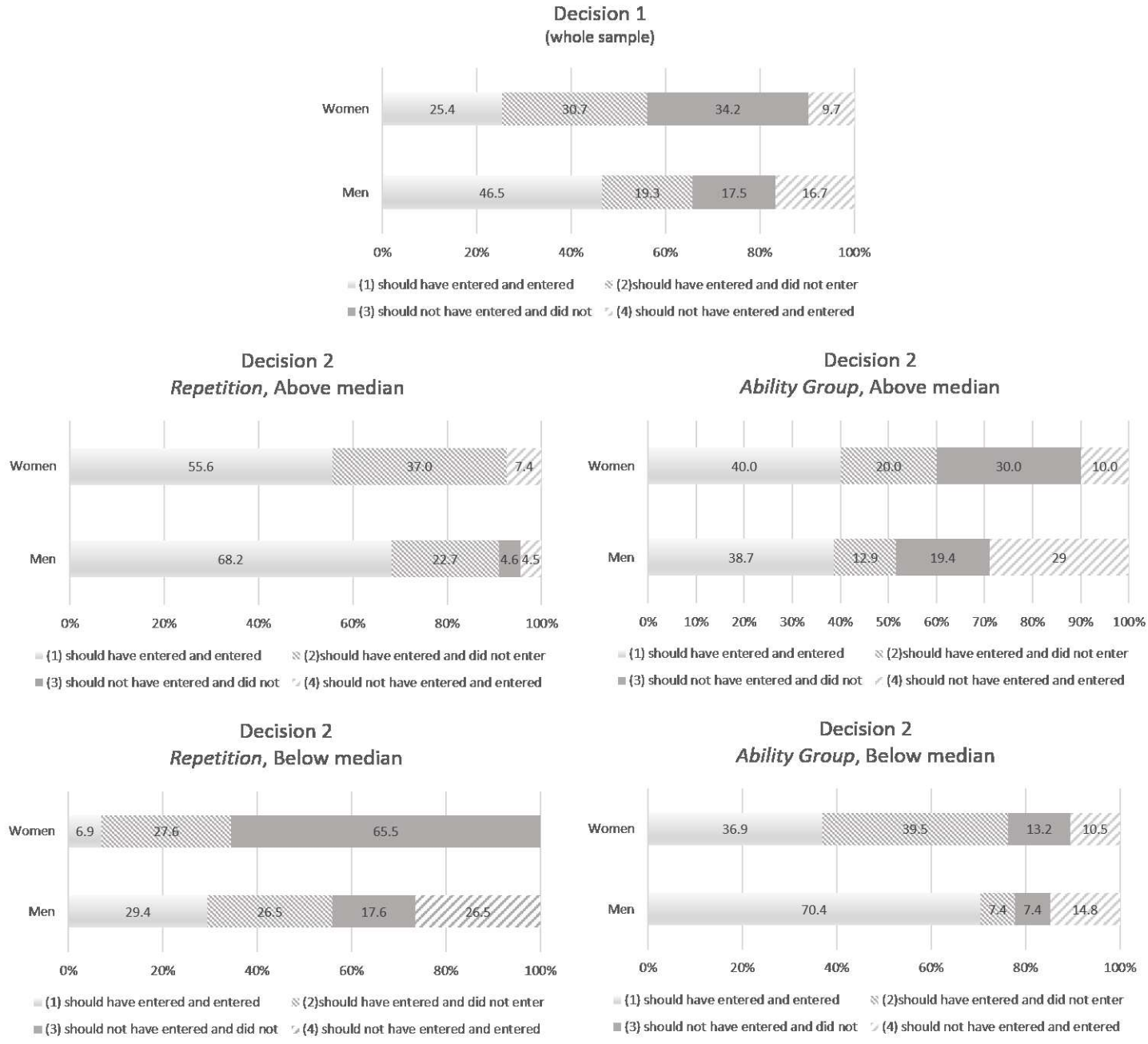
		PR	Tourn.
Decision 1 (Step 3)		if perf<9	if perf $\geqslant$ 9
Decision 2 (Step 4)	*Repetition*	if perf<9	if perf $\geqslant$ 9
	Ability group, below-median	if perf<7	if perf $\geqslant$ 7
	Ability group, above-median	if perf<12	if perf $\geqslant$ 12

decision was to enter the tournament and who chose the piece rate, (3) those whose optimal decision was the piece rate and who did chose the piece rate and (4) those whose optimal decision was the piece rate and who chose the tournament. Hence, both categories (1) ans (3) correspond to right decisions whereas (2) and (4) correspond to mistakes. Figure 5 shows the proportions of right decisions and mistakes from entering piece rate and tournament, before the feedback is received (Decision 1) and after it has been received in both *Repetition* and *Ability group* treatments. The plain parts of the graph correspond to the right decisions (light gradient grey for type (1) and dark grey for type (3)), the diagonally striped parts correspond to mistakes (thin diagonals for mistakes of type (2), thicker diagonals for mistakes of type (4)).

It appears that men and women do not make the same kind of mistakes. Men are more prone to entering the tournament when their optimal choice would be to stay out while women tend to choose the piece rate when they would maximize their expected payoffs by entering the tournament. In Decision 1, there is no difference in the rate of mistakes between men's and women's ($p = 0.50$). However, among the participants whose optimal choice is to enter the tournament, women are more likely to stay out ($p < 0.01$) and among the participants whose optimal choice is the piece rate, men are more likely to enter the tournament ($p < 0.01$).

Regarding Decision 2, the result holds qualitatively: that is, women who should enter the tournament stay out more often than men and men who should not enter the tournament do enter more often than women. However, significance is harder to reach as our sample is then

Figure 5: **Rates of right and wrong decisions according to Decisions 1 and 2 to enter tournament**



split in several small subsamples. We obtain two significant results though. In *Repetition*, among the below-median participants who should not have entered the tournament, more

men than women make the mistake of entering ($p < 0.01$). In *Ability group*, among the below-median participants who should have entered the tournament, more women than men make the mistake of choosing the piece rate ($p < 0.01$). Overall, after the feedback and among below-median participants, men make more mistakes than women (52.9% vs. 27.6%; p=0.04) in *Repetition* while women make more mistakes than men (50% vs. 22.2%; p=0.03) in *Ability Group*. None of those differences were significant before the feedback.

So far, we have considered only rates of mistakes (in the sense of non-payoff-maximizing behaviors) but some mistakes are more costly than others. One can think for instance of two participants whose respective Decision 1 performances are nine and 16, and who both chose the piece rate at Decision 1. Since both of their decision 1 performances are above or equal to nine, we know that their expected payoffs would have been higher if they had chosen the tournament (under the assumption that their Decision 1 performance would have been the same under both remuneration schemes) but the mistake of choosing the piece rate is larger for the second one (whose Decision 1 performance is 16) as her expected payoff from choosing the tournament would have been very high. In order to take into account the size of mistakes, we compute for each participant and Decisions 1 and 2 the difference between the expected payoff of the choice she made and the expected payoff of the alternative choice. For Decision 1, there is no significant difference between men and women. For Decision 2, we run tests separately by treatment and performance level (below-median and above-median) and two out of the four tests yield significant results. In *Repetition*, below-median men depart marginally more than women from the maximum expected payoffs they could have obtained ($p = 0.10$) because they enter the tournament too much. In *Ability group*, below-median women make larger mistakes than men ($p < 0.01$) because they do not enter enough. It therefore seems that the below-median feedback alone (in *Repetition*) has a more negative effect on men than women: women never enter when they should not following a below-median feedback though some of them still stay out when they should compete hence making only one kind of mistake. Men still make both kind of mistakes. However, adding the information that one will compete against another below-median participant (in *Ability*

group) is more detrimental to women than men.

We now look at the evolution of mistakes' sizes, between Decision 1 and Decision 2. We find only one significant result, namely that there is a significant improvement in choices of below-median men in *Ability group* ($p < 0.01$).[15]

Result 6: *Men and women make different mistakes: while men enter too often the tournament when they should not, women do not enter enough when they should. After a below-median feedback men make (slightly) bigger mistakes than women. However, the combination of a below-median feedback and ability group tournament leads men to make much smaller mistakes than women. Furthermore, in this last situation, men improve the quality of their decisions between Decision 1 and Decision 2.*

4 Discussion and conclusion

This paper investigates how beliefs are updated and what role they play in tournament entry decisions. Subjects are provided with relative performance feedback and given the option to participate in a tournament with or without information on the level of their opponent. We find that, following performance feedback, subjects update their beliefs more strongly than a Bayesian agent would. Both men and women are more pessimistic than a Bayesian agent following a below-median feedback; we find the opposite effect after an above-median feedback. Below-median women update even more pessimistically than their male counterparts when controlling for their actual performance level. We also show that below-median participants adapt their tournament entry decision to the level of the competition while above-median participants do not.

Concerning feedback reaction in terms of competitive entry, men and women do not react in the same way. While women are especially sensitive to information on their own performance level, men react more strongly to the level of their competitors. Feedback does not therefore seem to be processed in the same way by men and women. Below-median men, seem to expect their performance to improve over time. This is not the case for women.

[15] We do not report the proportions of mistakes in decision 1 for each subsample.

In other words, below-median women see their performance level at one point in time as indicative of their overall ability while men think there is room for improvement. As a result, women react very strongly to feedback received on relative performance, while men respond more to information on the performance they will have to beat, which is fixed by design (therefore one does not have to take into account the possibility that her opponent may improve). One important point is that the effects of feedback and information on the level of one's opponent on tournament entry decisions are transmitted mostly through the subjective belief of winning the tournament. We furthermore find what we call a "surprise effect" for women: they react more strongly to feedback when they did not expect it.

Our welfare analysis shows whether participants maximize their payoffs by their tournament entry decisions. Did they lose money by making the wrong choice? We show that men and women do not make the same kind of mistakes: men enter the tournament too often when they should choose the piece rate, women do not enter enough when they should. After a below-median feedback, men make slightly bigger mistakes than women. However the combination of a below-median feedback and the information that one's opponent will also be below-median leads to men making much smaller mistakes than women. It is because men improve the quality of their decisions in this situation.

When the level of competition is low, it might be desirable to emphasize it to low-performing men. In some contexts where feedback can have an effect on subsequent effort, one may also want to be careful not to send too rough negative feedbacks to low-performing women. They might indeed perceive this information as meaning they will always underperform.

References

Azmat, G., and N. Iriberri (2010): "The importance of relative performance feedback information: Evidence from a natural experiment using high school students," *Journal of Public Economics*, 94(7-8), 435–452.

Azmat, G., and N. Iriberri (2012): "The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness," CEP Discussion Paper 1116, Centre for Economic Performance, LSE. 6

Cason, Timothy N., William A. Masters, and Roman M. Sheremeta. 2010. "Entry into winner-take-all and proportional-prize contests: An experimental study." Journal of Public Economics, 94(9): 604 – 611. 6

Camerer, C., and D. Lovallo (1999): "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89(1), 306–318. 4

Croson, R., and U. Gneezy (2009): "Gender Differences in Preferences," *Journal of Economic Literature*, 47(2), 448–74. 6

Datta Gupta, N., A. Poulsen, and M.-C. Villeval (2013): "Gender Matching And Competitiveness: Experimental Evidence," *Economic Inquiry*, 51(1), 816 – 835. 3

Ertac, S. (2011): "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback," *Journal of Economic Behavior & Organization*, 80(3), 532–545. 5

Gill, David, and Victoria Prowse. 2010. "Gender differences and dynamics in competition: The role of luck." 5, 6, 12

Gneezy, U., K. L. Leonard, and J. A. List (2008): "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *NBER Working Paper* (13727). 6

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, CXVIII, 1049–1074. 5

Grossman, Z., and D. Owens (2011): "An Unlucky Feeling: Persistent Overestimation of Absolute Performance with Noisy Feedback," University of California at Santa Barbara, Economics Working Paper Series 1937484, Department of Economics, UC Santa Barbara. 3, 5

Gunther, C., N. A. Ekinci, C. Schwieren, and M. Strobel (2010): "Women can't jump? An experiment on competitive attitudes and stereotype threat," *Journal of Economic Behavior & Organization*, 75(3), 395–401. 5, 6

Hollard, G. I., S. Massoni, and J.-C. Vergnaud (2010): "Subjective beliefs formation and elicitation rules: experimental evidence," Documents de travail du centre d'economie de la sorbonne, Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne. 5

Kuhnen, Camelia M., and Agnieszka Tymula. 2012. "Feedback, self-esteem, and performance in organizations." *M*anagement Science, 58(1): 94–113. 8

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2013. "Managing Self-Confidence: Theory and Experimental Evidence." National Bureau of Economic Research, Inc NBER Working Papers 17014. 5

Murad, Z., M. Sefton, and C. Starmer. 2015. "Snowballing confidence in hierarchical tournaments." Mimeo. 5, 6, 8
6, 15

Niederle, M., and L. Vesterlund (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?," *Quarterly Journal of Economics*, 122(3), 1067–1101. 3, 5, 7, 12

Niederle, M., and L. Vesterlund (2011): "Gender and Competition," *Annual Review of Economics*, 3(1), 601–630. 5

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

Wozniak, D., W. T. Harbaugh, and U. Mayr (2011): "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices," Discussion paper, Department of Economics, UC Santa Barbara. 3, 4, 5

5 Appendix

5.1 Creation of the variable "beliefWin"

The variable "beliefWin" (bw) is a proxy for the subject's beliefs concerning tournament success. Remember that we elicit beliefs twice, both before and after the feedback is provided. We denote by b_{i_j} the beliefs elicited at round i of their performance being in the jth quartile. We have to make the additional assumption here that whenever a subject thinks her performance belongs to a given quartile, she actually thinks it lies at the exact midpoint of this quartile (e.g, if I think my performance belongs to the best quartile, we make the assumption that it means that I think 12.5% of step 2 performances were better than mine and 87.5% of step 2 performances were lower than mine). During the first round of belief elicitation, we get the subjects' beliefs about belonging to each of the four quartiles and define the belief of winning the Decision 1 (step 3) tournament as follows: $bw = 0.875 * b_{1\_4} + 0.625 * b_{1\_3} + 0.375 * b_{1\_2} + 0.125 * b_{1\_1}$. b_{1_4} is the subject's belief, before the reception of the feedback, that her performance belongs to the best quartile (i.e the fourth quartile). If she holds such beliefs, we consider that she thinks she will beat 87.5% of her potential opponents, who are all the subjects from her session, that is, the 75% of subjects whose performance belongs to the 1st, 2nd, and 3rd quartiles plus half of the subjects (12.5% of all subjects in her session) whose performance belongs to the fourth quartile. b_{1_3} is the subject's belief, prior to feedback, that her performance belongs to the third quartile. We consider that, if this is the case, she thinks she will beat 62.5% of her potential opponents, that is the 50% of subjects whose performance belongs to the first and second quartiles plus half of the subjects (12.5% of all subjects in her session) whose performance belongs to the third quartile. b_{1_2} is the subject's belief, prior to the feedback, that her performance belongs to the second quartile. We consider that, if this is the case, she thinks she will beat 37.5% of her potential opponents, that is the 25% of subjects whose performance belongs to the first quartile plus half of the subjects (12.5% of all subjects in her session) whose performance belongs to the second quartile. b_{1_1} is the subject's belief, before the reception

of the feedback, that her performance belongs to the first quartile. We consider that, in such a case, she thinks she will beat 12.5% of her potential opponents, that is, half of the subjects (12.5% of all subjects in her session) whose performance also belongs to the first quartile. This measure is not perfect as it does not take into account the possibility that the subject might think he will improve between steps 2 and 3, but we still consider it as a reasonable proxy.

In order to compute "beliefWin" for the Decision 2 tournament, we use the second-round beliefs elicited after the feedback is provided. These are the beliefs of one's performance that belonging to each of the two quartiles is still possible given the feedback received (e.g, If I was told my step 2 performance was above the median step 2 performance of my session, I know that it belongs either to the fourth (best) quartile or the third (second best) quartile). Again, if we denote by b_{i_j} the beliefs elicited at round i of their performance being in the jth quartile, we get that, in *Ability Group*, $beliefWin = 0.75 * b_{2\_2} + 0.25 * b_{2\_1}$ for both below-median and above-median participants.[16] In *Repetition*, $beliefWin = 0.375 * b_{2_2} + 0.125 * b_{2_1}$for below-median subjects.[17] Following the same reasoning, $beliefWin = 0.875 * b_{2\_2} + 0.625 * b_{2\_1}$ for above-median subjects in *Repetition*.[18]

[16] A below-median subject thinks there is a b_{2_2} probability that her performance belongs to the second quartile, in which case, since her potential opponents' step 2 performance was also below the median, she thinks she will beat all of those whose performance belongs to the first quartile (50% of subjects whose performance is below the median) and half of those whose performance belongs to the second quartile. She furthermore thinks there is a b_{2_1} probability that her performance belongs to the first (worst) quartile, in which case, she thinks she will beat half of her potential opponents from the first quartile (25% of her potential opponents). The same reasoning can be made for above-median subjects in *Ability Group*.

[17] b_{2_2} is the subject's belief, after the reception of the feedback, that her performance belongs to the best possible quartile given that she is below the median (i.e the second quartile). If she holds such beliefs, we consider that she thinks she will beat 37.5% of her potential opponents, who are all the subjects from her session, that is the 25% of subjects whose performance belongs to the first (worst) quartile plus half of the subjects (12.5% of all subjects in her session) whose performance belongs to the second quartile. b_{2_1} is the subject's belief, after the reception of the feedback, that her performance belongs to the worst possible quartile given that she is below the median (i.e the first quartile). If she holds such beliefs, we consider that she thinks she will beat 12.5% of her potential opponents, that is half of the subjects whose performance belongs to the worst quartile, while all other opponents, if randomly chosen to compete against her, would beat her.

[18] Such a subject thinks there is a b_{2_2} probability that her performance belongs to the fourth (best) quartile, in which case, she thinks she will beat all of her potential opponents from quartiles 1, 2, and 3 and half of those from quartile 4. She furthermore thinks there is a b_{2_1} probability that her performance belongs to the third (second best) quartile, in which case, she thinks she will beat all her potential opponents from quartiles 1 and 2 and half of those from quartile 3.

5.2 Tables

Table 9: **Descriptive statistics.**

Variable	Modality	
Age		25.8
Discipline	Economics	31.1%
	Science	3.1%
	Mathematics	2.2%
	Others	63.6%
Study level	Bac or less	3.95%
	Bac+1 to Bac+2	40.8%
	Bac+3	18.86%
	Bac+4 to Bac+5	34.21%
	More	2.19%
Father's education	Bac or less	35.53%
	Bac+1 to Bac+3	28.95%
	Bac+4 and more	35.52%
Mother's education	Bac or less	39.04%
	Bac+1 to Bac+3	34.21%
	Bac+4 and more	26.75%
Already participated in an experiment	Yes	73.25%

The Baccalauréat or "Bac" is an academic exam (and diploma) that French pupils take at the end of high school. Passing the Bac allows students to continue to higher education. Therefore, for instance, Bac+1 refers to a level of education of one year following the acquisition of the Baccalauréat.