Self-confidence and unraveling in matching markets*

Marie-Pierre Dargnies (University of Paris Dauphine)
Rustamdjan Hakimov (WZB Berlin)
Dorothea Kübler (WZB Berlin & TU Berlin)

November 2016

Abstract

We document experimentally how biased self-assessments affect the outcome of matching markets. In the experiments, we exogenously manipulate the self-confidence of participants regarding their relative performance by employing hard and easy real-effort tasks. We give participants the option to accept early offers when information about their performance has not been revealed, or to wait for the assortative matching based on their actual relative performance. Early offers are accepted more often when the task is hard than when it is easy. We show that the treatment effect works through a shift in beliefs, i.e., underconfident agents are more likely to accept early offers than overconfident agents. The experiment identifies a behavioral determinant of unraveling, namely biased self-assessments, which can lead to penalties for underconfident individuals as well as efficiency losses.

JEL Classification: C92, D47, D83

Keywords: Market unraveling; experiment; self-confidence; matching markets

---

* We would like to thank Nina Bonge for helping us with conducting the experiments. The paper has benefited from discussions with seminar audiences at the Universities of Cologne and Kiel, Loyola Marymount University, Los Angeles, Stanford University as well as with conference participants at ASFEE 2016 in Cergy, “Journée incertitude et décision publique” in Nanterre, and the ESA Conference 2016 in Tucson. We want to thank Georg Weizsäcker for detailed comments on our paper, as well as Siqi Pan, Al Roth, and Emanuel Vespa for their suggestions.

1. Introduction

Markets unravel when contracts are formed earlier and earlier in time. Inefficient matchings can be the result if insufficient information about the quality of the match is available at the early contracting stage compared to later stages. Also, when some transactions are made early and others later, the market is thinner at any point in time, and market participants have fewer possible matching partners to choose from. A classic case of unraveling is the market for medical interns in the US in the 1950s. Future doctors and hospitals agreed on matches long before the medical students had graduated. This led to uncertainty about the quality of the candidates at the time of contract formation.[2]

In real-life labor markets, workers have to form beliefs about their relative productivity and their attractiveness for employers in order to decide where to apply, which job to accept etc. We ask how these beliefs affect matching outcomes and unraveling. It can be expected that the less confident a worker is about her own value to prospective employers, the more likely she is to accept an early offer. This is because the candidates with low self-confidence are pessimistic about their job prospects once their true quality will be revealed. Indeed, internet forums are full of requests for advice by jobseekers facing offers that expire soon. These jobseekers fear to reject the offer and to end up worse off and at the same time are afraid of missing a better opportunity by accepting a job too early. Depending on their self-confidence, they may finally reject or accept such offers. Thus, we expect that unraveling of markets becomes more pervasive, the less confident the workers are.

Studying the role of beliefs about one's relative performance on the outcome of markets is notoriously difficult in the field. The main problem is that the level of self-confidence of market participants cannot be observed. Moreover, differentiating the role of self-assessments from other types of information that affect the decisions of market participants is often impossible. A laboratory experiment allows for a clean identification of the effects of relative self-assessments that can then guide field studies to evaluate the external validity of the findings.

In our experimental labor market, the productivity of workers is determined by their performance in a real-effort task. We experimentally shift the workers' self-confidence with the help of two different real-effort tasks causing participants to exhibit underconfidence and overconfidence respectively. This allows us to investigate whether self-confidence has a causal impact on the amount of unraveling in a market. More precisely, we are interested in whether there will be less unraveling the more confident our experimental jobseekers are.

The role of self-confidence in matching markets has not been studied so far, with two notable exceptions. Ma, Wu, and Zhong (2016) study Chinese college admissions data where rank-order lists for the Boston

[2] Other examples of markets with unravelling include the market for federal court clerks (Avery, Jolls, Posner, and Roth, 2001), for gastroenterology fellows (Niederle and Roth, 2003, 2004; McKinney, Niederle, and Roth 2005) and for college football games (Fréchette, Roth, and Ünver, 2007; Roth, 2012).

mechanism have to be submitted before students learn the result from their entrance exam. They find that female students are matched to worse universities than male students for a given exam score, due to female students ranking worse colleges first on their list. In contrast, Pan (2016) studies the role of self-confidence for outcomes of the Boston mechanism directly with the help of experiments. She also looks at a situation where students have to submit rank-order lists prior to learning their result from the entrance exam. She identifies a penalty for underconfident agents when the Boston mechanism is used, as compared to the strategy-proof serial dictatorship. Both studies provide evidence that there is a correlation between the matching outcome of an individual and the person's gender and self-confidence.

Our experimental design builds on the insight that unraveling can occur when there is uncertainty about the quality of the agents such as the productivity of the workers (Roth and Xing 1994, Li and Rosen 1998). Uncertainty about who is on the long and the short side of the market, together with the preferences of firms being similar can make it optimal for both sides to contract early, which was also pointed out by Halaburda (2010) and Niederle, Roth, and Ünver (2013). The idea is the following: consider a situation where firms and workers have similar preferences and there is a shortage of high-quality applicants. Suppose that at an early contracting stage, the quality of the workers is not known by the firms nor by the workers themselves. Then it can be optimal for some firms to make early offers and for some workers to accept these offers. Firms can be lucky and hire a good worker with an early offer while workers can avoid a bad matching outcome if they turn out to be of low productivity.

We create a relatively large experimental market with 16 workers and 13 firms. In the first stage firms have the possibility to make an early offer. Those workers who accept an early offer leave the market. However, the productivity of workers becomes known only after the first stage. In the second stage, there is a clearinghouse that implements the assortative matching among all remaining workers and firms.

In the baseline setup, the productivities of the workers are randomly generated and unraveling is predicted to occur, i.e., offers are made and accepted at the first market stage before the firms' preferences over workers become known. Note that workers and firms have symmetric information in this treatment. However, our main interest is in the case where workers can form beliefs about their relative productivity that is determined by a real-effort task while firms do not learn anything about a worker's productivity at the early market stage. Thus, we create a situation of asymmetric information between workers and firms by determining the workers' productivity rank with the help of a real task in the experiment.

In order to exogenously manipulate the self-confidence of workers, we use two different tasks. The task employed in one treatment is relatively easy while the task in the other treatment is hard. We expect the simple task to generate overplacement, i.e., an overly optimistic belief about one's rank, and the hard task to lead to underplacement. Thus, we chose the tasks so as to generate different levels of self-confidence regarding one's rank in the group of workers. This self-confidence is predicted to affect unraveling. Suppose workers are miscalibrated about their true productivity rank. If they underestimate their type, firms can profit from early offers because there is a chance that a worker will accept an offer which is worse than their matching partner in the assortative match. On the other hand, if workers know their type perfectly or are overconfident, firms would not make early offers as they could only lose vis-a-vis the assortative matching outcome in the stage.

Our study contributes to the question to what extent beliefs that may be biased are causal for market outcomes. The main contribution is twofold: (1) To our knowledge, we are the first to employ the so-called "hard-easy gap" (Lichtenstein et al. 1982) to induce different beliefs. This method of shifting the workers' beliefs about their relative performance with the help of the experimental treatments could be useful in other contexts where the effect of self-confidence is studied. (2) The design allows us to investigate the effect of the beliefs on the market outcome. The market equilibrium depends on the beliefs of the workers about their own ranks (that is, their self-confidence), and on the second-order beliefs of firms and workers. It is an empirical question how market participants form such beliefs in a setting where the ranks of workers are determined by a real task. By demonstrating that market outcomes are influenced by biased beliefs, we speak to the longstanding question of whether biases can affect market outcomes or whether they are wiped out by the mechanics of the market.

Specifically, the experiment was designed to answer the following questions: (i) Can unraveling be observed in the treatment with randomly determined ranks, as predicted by the equilibrium? (ii) Are early offers rejected more often in the treatment with the easy task compared to the treatment with the hard task and is this driven by differences in beliefs? Are the payoffs of workers affected by the treatments? Subjects who are more optimistic about their prospect for the second stage should reject more early offers. Thus, more assortative matches should be formed in the treatment with the easy task than with the hard task. (iii) Is there a gender difference in the acceptance decisions? If women are found to be overall less confident than men, they should accept more early offers and do worse than men with the same productivity rank. (iv) Do firms make early offers and if yes, which firms? Note that in the case where the ranks are determined randomly, some firms have an incentive to make early offers in equilibrium. In the treatments with the real-effort tasks, the firms will only make early offers if they expect to be able to hire a better worker than by waiting for the assortative matching. Thus, firms can profit more from early offers, the more under-confident the workers are.

Our first finding is that in the treatment with randomly determined ranks, subjects reject 92% of offers that should be rejected in equilibrium while they accept only 65% of offers that should be accepted in equilibrium. This behavior can be explained by over-optimism about the outcome of the random draw, and we find support for this explanation in the subjects' beliefs. Ad (ii) we find that subjects are on average overconfident in the treatment with the easy task and on average underconfident in the treatment with the hard task, which leads to more frequent rejections of early offers in the treatment with the easy task compared to the treatment with the hard task (29.2% of early offers received are accepted in the easy-task and 42.3% in the hard-task treatments, $p<0.01$). The treatment difference is driven by a shift of the beliefs: controlling for beliefs, there is no treatment difference in the acceptance decisions. With respect to payoffs, we observe that workers earn more in the easy-task treatment where the average overconfidence is higher than in the treatment with the hard task. Our third finding, ad (iii), is that women are less confident than men, but do not accept early offers more often than men. Thus, women are modest when stating their beliefs, but this modesty does not translate into decisions. Regarding (iv), we observe that the lower the quality of the firm, the more likely it is to make early offers, as predicted. Also, the less self-confident firms believe workers to be, the more likely they are to make early offers. Overall, we find that underconfidence is a source of unraveling, leading to inefficient matchings, more instability, and increased profits of lower-quality firms, compared to overconfidence. Over-optimism and overconfidence limit unraveling.

Our experiment complements the literature studying the causes for unraveling. When markets are congested and market participants fear that they do not have sufficient time to find a good matching partner, they have an incentive to close early contracts (Roth and Xing 1997). The study of UK regional entry-level labor markets for doctors has led to the hypothesis that a stable centralized clearinghouse, operating once all the relevant information has become available, prevents unraveling. A matching is considered stable if (a) no two agents prefer each other to their assigned partner and (b) no agent finds his partner less desirable than being unmatched. On the other hand, if the matchings generated by the clearinghouse are unstable, then unraveling will be a likely response of market participants (Roth 1991; Kagel and Roth 2000). Other reasons for unraveling that have been analyzed are strategic uncertainty of the proposing side about how many other agents go early (Echenique and Pereyra, 2016), search costs that can render it optimal to accept an early offer (Damiano et al. 2005), and the dynamic arrival of new agents, which provides an incentive for contracting early to avoid being surpassed by new agents of higher quality (Du and Livne 2016).

An exogenous variation in beliefs has been employed in a number of previous papers, such as Jensen (2010) in the realm of education, Möbius, Niederle, and Rosenblat (2014) with respect to belief formation about own ability, Costa-Gomes, Huck, and Weizsäcker (2014) in the context of a trust game, as well as in recent work by Van der Weele and Schwardmann (2016) who focus on the optimality of beliefs.

2. Experimental Design

This section describes the market that the subjects participated in, the real-effort tasks employed to determine the workers' rank as well as details about the belief-elicitation task and the three different treatments.

2.1 The market game

We implemented a labor market with workers searching for a job and firms searching for a worker. Each market (and session) has 24 subjects, and we invited an equal number of women and men to each session.[3] Of them, 16 participants were in the role of workers, and the remaining eight in the role of firms. Five additional firms (leading to a total of 13 firms) were played by the computer We designed a relatively large market in order to provide a setting where the workers can form beliefs about their rank and where over- or underconfidence are less often truncated for the best and the worst workers compared to a small market. The five computerized firms took on the role of the five best firms and were programmed not to make early offers. Note that in RANDOM, the best five firms do not make early offers in equilibrium. For treatments OVER and UNDER, creating common knowledge about the strategy of the best five firms simplifies the game considerably, as detailed when we derive the predictions for the game.

The value of a worker for the firms depends on her relative productivity. All firms have identical preferences over workers, i.e., earn a higher profit, the higher the productivity of the worker they are matched with. The payoff of a firm only depends on the productivity of the worker it is matched to. The exact payoffs implemented in the experiment can be found in Table 1, For example, a firm receives 50 points if it is matched to one of the five most productive workers.

	Most productive workers 1-5	Workers 6-9	Workers 10-13	Least productive workers 14-16
Payoff of firm (points)	50	25	15	10

Table 1: Payoff of firms depending on the productivity of the worker.

The payoff of a worker depends on the firm that hires her and, for workers hired by one of the top-five firms, on her own productivity. All workers want to be matched with the best possible firm. There are five high-quality firms that are computerized in the experiment. Furthermore, there are four firms of intermediate or middle quality (firms 6 to 9) and four of low quality (firms 10 to 13). Each of the five most productive workers earns 50

[3] Sometimes the sessions were not fully balanced with respect to gender due to participants not showing up. The most unbalanced session in OVER or UNDER had 14 men and 10 women, and in RANDOM 16 men and 8 women. In total we have 175 men and 161 women.

points when being matched to the five best firms while the other workers only earn 32 points at these firms.[4] At all firms except the best five, workers earn the same number of points, independent of their productivity.

We chose this payoff structure for three reasons. First, the absence of complementarities between worker and firm quality (except for the top five firms) simplifies the payoff structure for the participants. Tables 1 and 2 were presented in the instructions for the participants and contain all relevant information. Second, in our framework each firm is only interested in the productivity group of the worker they hire and not on his/her specific productivity rank. This makes the decision of firms whether to make an early offer or not easier. Third, since our primary interest concerns the impact of biased beliefs of the workers on market outcomes, we attempt to limit the potential effect of efficiency concerns on unraveling. Therefore, the possible welfare losses caused by unravelling are small in the market we designed. Most departures from the assortative matching do not cause a welfare loss but merely transfer payoffs from one agent to another.[5]

	Five most productive/all other workers
Five best firms	50/ 32
Sixth-best firm	32
Seventh-best firm	31
Eight-best firm	30
Ninth-best firm	29
Tenth-best firm	17
Eleventh-best firm	16
Twelfth-best firm	15
Thirteenth-best firm	14
No job	0

Table 2: Payoffs of workers depending on the quality of the firm and their own productivity.

[4] We chose to differentiate the payoffs at the best five firms for different worker types in order to make the equilibrium as simple and clear-cut as possible. In particular, it provides incentives to workers of intermediate productivity to accept offers of similar firms and not hope for a match with a top firm in stage two.

[5] There are two ways in which a welfare loss occurs: (1) each time that one of the best five workers is not employed by one of the five best firms, this worker gets 32 instead of 50, corresponding to a loss of 18; (2) whenever the unemployed workers are not workers 14-16, welfare is lost. The worst-case scenario in this respect is when workers 14, 15 and 16 receive and accept an early offer and when workers 6, 7 and 8 end up unmatched, resulting in a welfare loss of 3*15=45 (employing a bad worker instead of a good one provides a payoff of 10 instead of 25, hence the opportunity cost of 15).

Each round of the experiments consists of two stages:

First stage: At the beginning of the first stage, the quality of the firms is revealed to all participants. The productivity of the workers is not known to anybody, neither to the firms nor to the workers themselves. Middle- and low-quality firms can make early offers to the workers while the five computerized firms never make early offers. The offer of a firm goes to a random worker. Workers are free to accept or reject an offer and have 30 seconds to make this decision. If an offer of a firm is accepted, both the firm and the worker leave the market. If an offer of a firm is rejected, it is automatically sent to another worker. The first stage consists of a maximum of nine rounds, thus an offer can be rejected at maximum nine times. The same offer is never given to a worker who has already rejected it once. This sequential ordering was implemented in order to avoid that workers try to be faster than others in accepting an offer.

We employ a variant of the strategy method introduced by Bardsley (2000). Any worker who receives an offer from a firm gets two fictitious offers in addition. Each worker has to decide whether he accepts each of the three offers or not. This means that he can accept all three offers, only two of them, only one or none of the offers. He does not know which offer is real and which offers are fictitious. If he refuses an offer, he will never receive another offer from this firm regardless of whether the offer was real or fictitious. Compared to a full strategy method where a worker would be asked to respond to potential offers from all firms, our procedure preserves the real-world feature that job offers create a "hot" state, which can potentially affect decisions. Moreover, the random fake offers complicate any potential inferences about other workers' rejection behavior from the set of offers received.

Second stage: All workers and firms who remain unmatched at the end of the first stage move on to the second stage. At the beginning of the second stage, the productivity of all workers is revealed. Moreover, it is announced which firms and workers have already left the market in the first stage. Then, the assortative matching of all remaining firms and workers is implemented: the five best unmatched workers are assigned to the five high-quality firms, the sixth-best unmatched worker is assigned to the best among the medium and low-quality firms which are still in the market and so on. The three workers of the lowest productivity among all workers at the second stage remain unmatched and receive a payoff of 0. Notice that workers and firms only have to make decisions in the first stage. The second stage is executed by the computer, according to the above description.

Three practice rounds of the market with randomly drawn productivities were conducted before the payoff-relevant market was played.

2.2 Treatments, belief elicitation, and experimental procedures

We implemented three treatments between subjects: RANDOM, OVER and UNDER. The only difference between the treatments is how the relative productivity of workers is determined in the payoff-relevant market. In treatment RANDOM, the computer randomly attributed productivity ranks to workers. In treatments OVER and UNDER, subjects had to solve a real-effort task and were ranked according to how well they did in the task in comparison to other subjects. In the OVER treatment, subjects had to solve as many additions as they could in the course of five minutes. Every correct answer granted one point to the participant and there was no penalty for an incorrect answer. This task is easy in the sense that subjects know what they have to do and that they can solve a certain number of the problems in the given time. In the UNDER treatment, subjects had to solve as many logic questions (taken from IQ tests) as they could in ten minutes. The participants gained one point per correct answer and one point was subtracted from the participant's score for each incorrect answer. Participants were not allowed to skip questions. This task is hard in the sense that we expect subjects to correctly answer only a small number of questions, and in particular fewer questions than they expect to solve correctly, and because there is no clear technique that can be applied to find the answers. The two tasks were chosen to induce relative overconfidence in the case of the easy additions task (see Niederle and Vesterlund 2007) and relative underconfidence in the hard task with IQ questions, since we expect subjects to place too much weight on their own absolute performance and neglect the difficulty of the task for others (Kruger 1999, Moore and Healy 2008).

In both treatments, the worker who solved most tasks correctly received rank 1, etc. The participants in the role of firms knew that workers had to solve a real-effort task but were not informed about the nature of the task. Thus, the treatments OVER and UNDER were exactly identical for the participants in the role of firms. This allows us to interpret all differences between the treatments as differences in the workers' decisions and beliefs.

After the first stage and before the start of the second stage, subjects in the role of workers were asked to guess their productivity rank. They were incentivized using a quadratic scoring rule. The following payment rule was implemented:

$5 - 0{,}25 * (\text{guessed rank} - \text{true rank})^2$ points.

At the same time, subjects in the role of firms were asked to guess the beliefs of every worker about their productivity. They were also incentivized using a quadratic scoring rule, for guesses of every worker separately. They were told that only one of the 16 workers for whom they had made a guess would be drawn randomly to determine their payoff from this task. The following payment rule was implemented for each guess:

$5 - 0.25 * (\text{guessed rank for worker i by a firm} - \text{own guessed rank of worker i})^2$.

Note that the maximum payoff from the belief elicitation task was only 5 points, compared to earnings of up to 50 points from the market. We chose this relatively small remuneration for the stated beliefs in order to limit the incentive to hedge.

We conducted the experiments between April 2014 and September 2015 in the experimental economics lab at Technical University Berlin. Overall, 14 sessions were run, five for the treatments OVER and UNDER, respectively, and four for RANDOM. In total 336 subjects participated in the experiments, mostly students from various fields and universities of Berlin. We ran our experiments with the help of computers using z-Tree (Fischbacher, 2007). Subjects were recruited from a database with the help of ORSEE (Greiner, 2004). Each point earned was exchanged for 0.40 euro at the end of the experiment. The average session lasted 65 minutes, and the average payoff of participants was 11.31 euros plus 5 euros show-up fee.

3. Equilibrium Predictions

In all three treatments of the experiment, the workers had not received any information about their productivity rank when they had to accept or reject early offers by the firms. According to how the ranks of the workers were determined, by a known random process or by a real-effort task, the equilibrium predictions differ. In all treatments, a worker has an incentive to accept an early offer if the payoff from accepting it is higher than the expected payoff from the matching implemented in stage two. A worker has to take into account that some firms and workers will have left the market before stage two, which affects the expected return.

3.1 Predictions in treatment RANDOM

In treatment RANDOM the workers' ranks are determined randomly by the computer. There is a unique equilibrium outcome in which all middle-quality firms make early offers which are accepted by the workers who receive these offers. All other market participants are matched in stage two.[6] To see this, we show (by elimination of dominated strategies) that firms 1 to 3 never make early offers and that workers always reject early offers from low-quality firms regardless of their beliefs about what other firms and workers are doing. Knowing this, firms 4 and 5 will never make early offers. Given this, middle-quality firms (firms 6 to 9) will make early offers, and workers will accept these offers. More specifically, in equilibrium workers accept the first offer they receive from

[6] The proof is relegated to the appendix.

a middle firm. Finally, firms 14 to 16 are indifferent between making an early offer or not, since their early offers are always rejected by the workers.

3.2 Predictions in treatments OVER and UNDER

The outcome of the markets where a real-effort task determines the workers' ranks cannot be predicted without making assumptions about how well the workers can predict their rank. Moreover, the market outcome depends on the second-order beliefs of workers and firms about the productivity of other workers. But note that subjects in the role of firms do not know the nature of the tasks the workers have to solve. Therefore, the beliefs and the behavior of the firms should be the same in treatments UNDER and OVER. Now, consider two extreme cases: if it is common knowledge that the workers know their own rank, there is no unraveling in equilibrium as workers will only accept offers coming from firms better than their assortative matching partner and firms will therefore refrain from making early offers that can only make them worse off than waiting for the second stage. In the other extreme case where it is common knowledge that every rank is equally likely for each worker, the game is as in RANDOM, and unravelling will take place.

The common knowledge assumption regarding the workers' beliefs about their ranks appears sensible in the game where players know that ranks are determined randomly, but is less convincing in the game where the ranks are based on the real performance in a task. The question is then which beliefs about their own rank and about the other participants' beliefs and choices the players actually hold, since this affects their choices and the resulting market outcome. [7] ,We hypothesize that more early offers are accepted in treatment UNDER than in treatment OVER. The following beliefs and strategies of workers support this hypothesis: [8]

The workers who believe to be among the top-five most productive workers have a dominant strategy to reject any early offers, as they will be matched to a top-five firm in the second stage for sure, and those who believe to be among the three least productive workers will accept any offer (for simplicity, we assume that they will accept all first round offers they receive), as they are unmatched in the second stage. Finally, assume that the

[7] For example, the decision of a worker believing to be of productivity rank 7 regarding an early offer from firm 6 may not be the same whether she believes the top-five workers to be underconfident or not. In the first situation, the worker expects to be matched to a top-five firm in the second stage (because some top-five workers may have already left the market by then). In the second situation, the worker believes he can never be matched to a top-five firm and will therefore accept the offer from firm 6.

[8] For simplicity, we assume here that workers only receive one round of offers. This means that a worker never rejects an early offer in the hope of receiving a better early offer in the next round. Rather, a worker will only reject an early offer if he expects a better payoff from waiting for the assortative matching.

probability that workers who believe to be of ranks 6 to 13 accept an early offer weakly increases in their guessed rank, for a given type of firm, and is constant between treatments.[9]

We expect that our treatment effect is such that the subjects in OVER assign a higher subjective probability to being weakly better than a certain rank compared to the subjects in UNDER. This corresponds to the guessed ranks in UNDER first-order stochastically dominating the guessed ranks in OVER. If such is the case, then more early offers are accepted in UNDER, according to the strategies of the workers specified above.

No general statements can be made regarding the comparison of OVER and UNDER with RANDOM regarding the amount of unraveling. Too see this, note that in OVER and UNDER there might be workers who believe that they are among the least productive (14-16), and thus they should accept any early offer, even offers from bad firms, which is never the case in RANDOM. In the extreme case, all workers have such beliefs, leading to all early offers being accepted in stage one. On the other hand, there can be workers in OVER and UNDER who believe to be among the best five workers (1-5), and thus they reject offers from middle firms. In the extreme case, all workers believe to be among the top five, and no early offers are accepted. Thus, depending on the beliefs of workers in OVER and UNDER, there can be more or less acceptances of early offers compared to RANDOM.

4. Results

This section first presents the results concerning the productivity of workers in the tasks employed in OVER and UNDER as well as the workers' beliefs regarding their performance. Then we proceed to our main question, namely whether the treatments affect the acceptance of early offers. All tests that we perform on the data are two-sided unless indicated otherwise. For the regressions, we cluster at the level of the markets (that are equivalent to sessions).

[9] This assumption of a monotone relationship between guessed rank and acceptance decisions excludes the following cases, among many others: It can be optimal for worker 6 to accept the offer from firm 13 if she believes that firms 6-13 make early offers, that the top five workers are well calibrated (and therefore reject any early offer), and that all other workers accept any offer. Thus she believes that she will be unassigned in the second stage if she rejects an early offer. Similarly, it can be optimal for a worker with productivity 13 to reject an offer from firm 6 if she believes that all firms 6-13 make an early offer, and workers 14, 15 and 16 reject all offers, while all other workers accept any offers. Thus, worker 13 believes that she will be matched to the fifth firm in the second round is therefore indifferent between accepting and rejecting the offer of firm 6 which yields the same payoff.

4.1 Productivity in OVER and UNDER

We start by studying the productivity of workers, i.e., the rank obtained by each worker after working on the task in the treatments UNDER and OVER. This allows us to check whether the performance in the tasks differs between treatments and for men and women respectively.

To determine the workers' performance, we count the number of correct answers in OVER and the number of correct answers minus the number of incorrect answers in UNDER. The average performance score in OVER where additions had to be solved was 9.1 while it was -0.7 in UNDER with IQ quiz questions. Furthermore, performance levels are less dispersed in UNDER than in OVER. The standard deviations of performance scores are 2.9 in UNDER and 4.72 in OVER, and there were 21 unique values of scores in OVER versus 15 in UNDER. Whenever two workers from the same session obtain the same score, the tie is broken randomly to determine productivity ranks. Effectively, there were more ties in UNDER than in OVER. If participants are aware of this, they will be less certain of their productivity rank in UNDER than in OVER. It could also be that it is more difficult for someone to assess his number of correct answers for the IQ logic questions than for additions which is a more common task. This would also lead to more uncertainty about one's performance level in UNDER than in OVER. We will be careful to address this issue when analyzing the treatment effect on the decision to accept or reject early offers.

We observe that in both treatments, the relative productivity of men and women is similar. In treatment UNDER, the average ranks of men and women are 8.59 and 8.39, respectively (Mann Whitney test $p=0.85$). In treatment OVER, the average ranks of men and women are 8.03 and 8.95, respectively (Mann Whitney test $p=0.38$). Thus, women perform slightly better in the IQ quiz in UNDER while men perform slightly better in the additions task in OVER.

4.2 Beliefs of workers

In all three treatments, beliefs were elicited by asking subjects in the role of workers to guess their productivity rank. The elicited beliefs are crucial to determine whether the treatments OVER and UNDER had the desired effects of shifting the beliefs about the relative rank. We start by constructing the variable Overconfidence equal to the difference between the productivity and the belief (a positive value means that the subject's actual rank is higher than her guessed rank, indicating overconfidence).

We find that the subjects are on average underconfident in UNDER and overconfident in OVER, as shown by Figure 1, left panel, and the difference between OVER and UNDER is significant (Wilcoxon rank-sum two-sided $p<0.01$). Moreover, the level of overconfidence is not significantly different in RANDOM and OVER ($p=0.98$), but it is significantly higher in RANDOM than in UNDER ($p=0.02$).

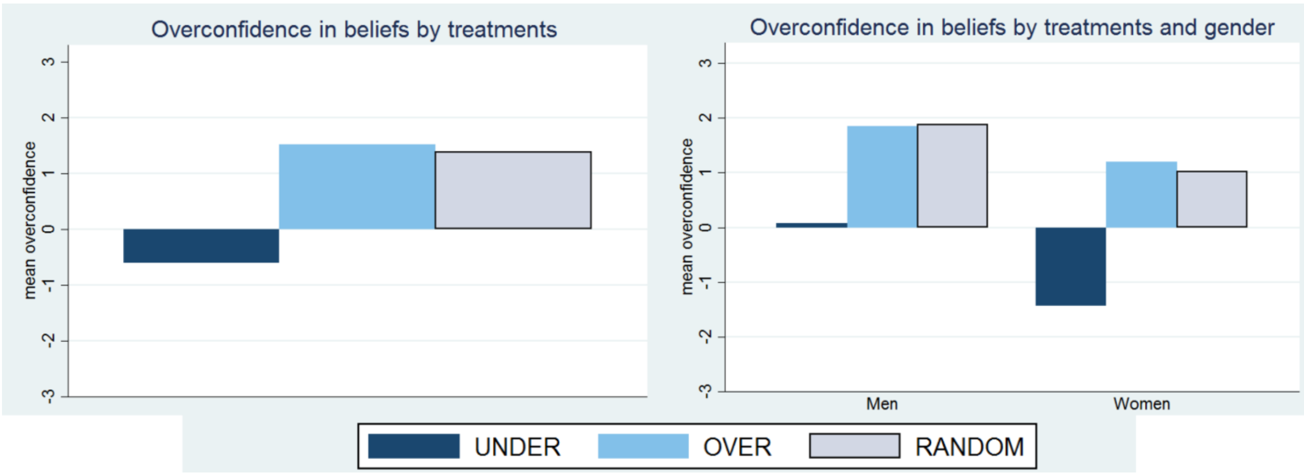

Figure 1: Average overconfidence of men and women in treatments UNDER, OVER, and RANDOM.

Figure 2 presents the cumulative distribution functions of the workers' belief for each of the three treatments. The distributions of the stated beliefs are significantly different between UNDER and OVER (Wilcoxon rank-sum two-sided p<0.01). Moreover, the beliefs in UNDER first-order stochastically dominate the beliefs in OVER, except for beliefs to be of rank 15 or higher. Thus our assumption regarding the relationship between beliefs in OVER and UNDER to support our prediction regarding the amount of unravelling in OVER and UNDER finds support in the data. Finally, the distribution of stated beliefs is not significantly different between OVER and RANDOM (p=0.41), but significantly different between UNDER and RANDOM (p<0.01).[10]

Turning to possible gender differences depicted in the right panel of Figure 1, it is apparent that the confidence level of women is lower than for men in all three treatments. However, we do not find a significant gender difference in OVER nor in UNDER (Wilcoxon rank-sum test, two-sided, p-values are 0.56 and 0.26, respectively). Men are more optimistic than women in RANDOM (p=0.08). When pooling OVER and UNDER, the gender difference remains insignificant (p=0.34). Only for all the three treatments together there is a significant gender difference (p=0.05). Finally, there is no difference between women and men regarding how correct their beliefs are, i.e., with respect to the difference in the absolute value between Belief and Productivity, for either treatment.

[10] Despite the absence of a difference between OVER and RANDOM, note that the calibration of subjects in the two treatments is significantly different (p<0.05). Thus despite on average being similarly overoptimistic about the placement in RANDOM and OVER, subjects in RANDOM are significantly less calibrated, which is to be expected, since their ranks are determined randomly.

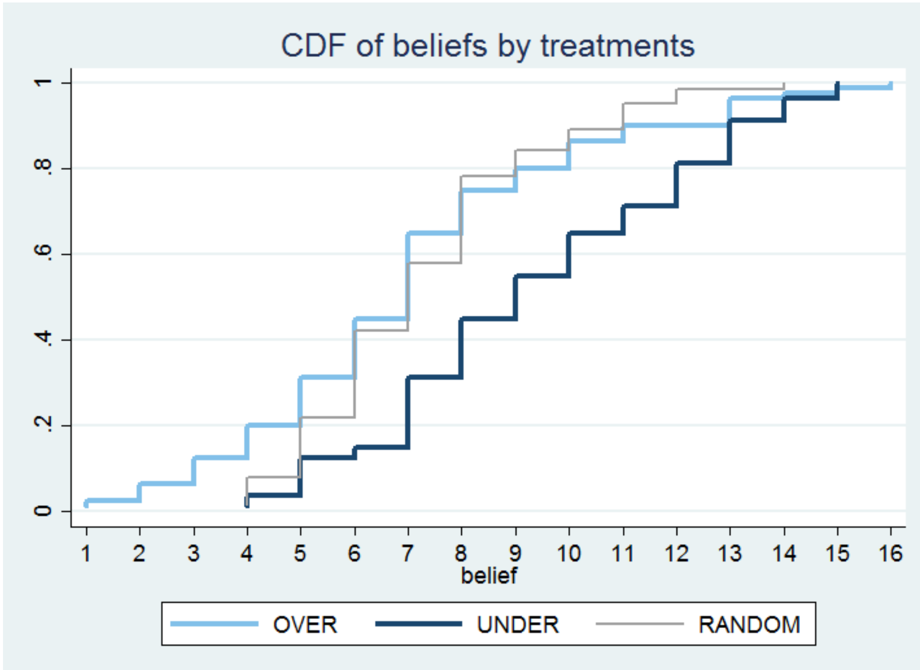

Figure 2: Cumulative distribution functions of beliefs in treatments UNDER, OVER and RANDOM.

Since we observe that women perform better in the hard task while men perform better in the easy task, we study the workers' confidence levels controlling for productivity. We run a linear regression of the dependent variable Belief, i.e., the rank stated by the worker, on Productivity expressed as the performance rank, on the gender dummy Female, and on a treatment dummy using the data from treatments UNDER and OVER. In these treatments the actual performance of participants determines their ranks and the participants can form beliefs about their rank. In contrast, the beliefs in RANDOM are not determined by the productivity in the task but rather by the perception of the random draw, and we therefore do not include this treatment in the regressions. The results are presented in Table 3. Model (1) shows that, controlling for productivity, women are less confident than men and subjects are more confident in treatment OVER than in UNDER.

In order to understand further how the subjects form beliefs and whether beliefs are closer to the correct beliefs in OVER or in UNDER, we compute the payoff-maximizing or optimal guess for each performance level separately in UNDER and OVER. The optimal guess is the belief which is most likely to be correct for each given performance level. In order to do so, we generate 10,000 groups of 15 workers randomly. We take a virtual worker and assume that he faces a group of another 15 workers in the experiment. Then we calculate his most likely productivity rank given each of his possible performance levels. For instance, in UNDER if a worker's score is -6, she is most likely to be the least productive of his group of 16 workers.

Dep var : Belief	Model (1)	Model (2)	Model (3)

Intercept	6.64*** (0.67)	6.35*** (0.55)	7.49*** (0.69)
Productivity	0.24*** (0.06)	-0.02 (0.12)	-0.03 (0.11)
Female	1.47*** (0.38)	1.46*** (0.44)	1.40*** (0.43)
OVER dummy	-2.38*** (0.57)	-2.33*** (0.44)	-4.51*** (0.92)
Optimal Guess		0.29** (0.12)	0.17 (0.13)
OVER*OptimalGuess			0.26*** (0.10)

Table 3: Linear regressions of Belief on explanatory variables for treatments OVER and UNDER, with standard errors clustered at the sessions level. Notes: Values in parentheses represent standard errors. *p<0.1,**p<0.05,***p<0.01

From Model (2) in Table 3, we can take that the guessed productivity rank (Belief) is higher the higher the optimal guess is. This is all the more true in OVER as can be taken from Model (3). The significance of OVER*OptimalGuess shows that it is easier for the subjects to make a good guess in OVER than in UNDER. Thus, the uncertainty of workers about their rank is higher in UNDER, which is in line with the observation of more ties in UNDER than in OVER.

4.3 Acceptances and rejections by workers

Next we study the main question addressed by the experiment, namely whether the treatment affects the decisions of workers to accept early offers. The solid lines of Figure 3 present the average acceptance rates of early offers in OVER and UNDER depending on the rank of the firm that is making the offer. The average acceptance rate is higher in UNDER for all middle firms. For the low-quality firms 10 to 13, acceptance rates are small and similar in both treatments. The graph presents the raw data, without controlling for the performance of the workers. However, it gives a first impression of the strength of the exogenous variation of beliefs on the decisions of workers.

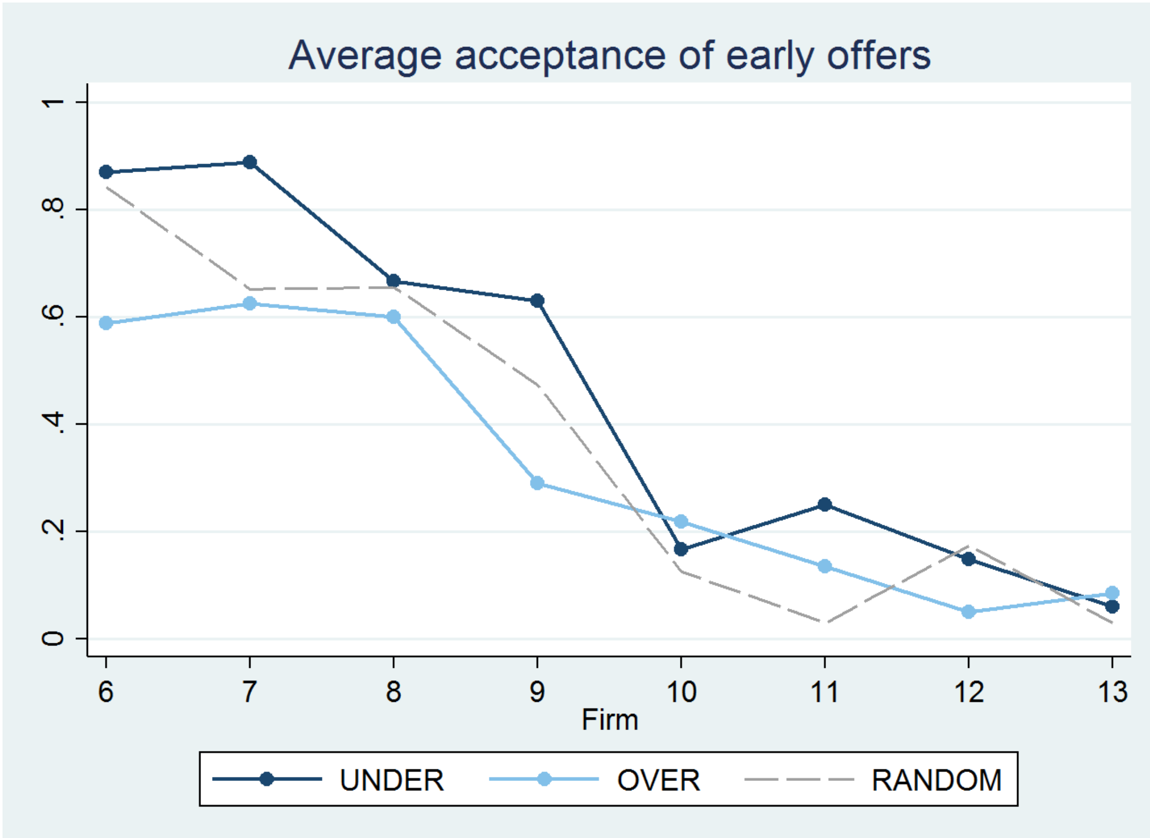

Figure 3: Average rate of acceptance of firms' early offers by treatments.

In order to study the acceptance and rejection decisions of workers in treatments UNDER and OVER controlling for productivity, we regress the decision to accept an early offer on a number of explanatory variables. Model (1) of Table 5 shows that early offers are less likely to be accepted in treatment OVER than in UNDER. Moreover, the higher (i.e. worse) a worker's productivity rank, the more likely he is to accept an early offer. Finally, the variable Offer corresponds to the rank of the firm making the offer (the higher Offer is, the less attractive the firm), and the regression displays that offers from better firms are more likely to be accepted.

Our prediction that more offers are accepted in UNDER than in OVER finds support in the data. However, it is unclear whether the treatment difference arises due to differences in the confidence level. We hypothesize that subjects in OVER accept early offers less often than in UNDER, since more confident subjects have stronger incentives to wait for the assortative matching in the second stage compared to less confident subjects. To test for this channel, we include the beliefs in the regression.

Dep var : Accept offer	Model (1)	Model (2)	Model (3)
OVER_dummy	-0.16**	-0.03	0.02

	(0.07)	(0.06)	(0.09)
Female	0.03	-0.03	0.03
	(0.08)	(0.07)	(0.05)
Productivity	0.02***	0.01	0.01
	(0.01)	(0.01)	(0.01)
Offer	-0.12***	-0.13***	-0.13***
	(0.02)	(0.02)	(0.02)
Belief		0.06***	0.06***
		(0.01)	(0.01)
OVER*female			-0.1
			(0.11)

Table 4: Propensity of workers to accept early offers. Based on data from UNDER and OVER. Marginal effects after probit regression with errors clustered at the session level. Values in parentheses represent standard errors. *p<0.1,**p<0.05,***p<0.01

Model (2) in Table 4 shows that when Belief is added as a regressor, the dummy OVER is not significant anymore. This indicates that the treatment indeed shifts the beliefs: subjects are less likely to accept an early offer in treatment OVER because they are more confident about their productivity rank which determines which firm they will be matched with in the second step. Importantly, we do not observe a treatment effect other than through the beliefs. Thus, the alternative explanation that more early offers are accepted in UNDER because subjects in UNDER are more uncertain about their productivity rank, does not find support. A higher uncertainty about the beliefs cannot explain the finding that the treatment difference disappears when controlling for beliefs. Thus, in line with the prediction, the exogenous shift of subjects' beliefs results in significantly different acceptance decisions.[11]

The treatment effect is mainly driven by workers of high and middle productivity for whom the propensity to accept an early offer in OVER is 27% lower than in UNDER.[12] We also note that the coefficient of

[11] We can also control for whether acceptances are influenced by the period in which an offer is received in stage 1. We include a variable that corresponds to the number of times an offer was rejected, which strongly correlates with the time the offer was received. First, adding the variable to Model (1) does not change the significance of the treatment dummy. The variable itself has a negative coefficient but is not significant (p=0.106), weakly indicating that the later an early offer is received, the more likely it is rejected. In Model (2), the variable has no significant effect. This excludes the possibility that workers update their self-confidence when receiving late early offers because they believe that other workers must have rejected the early offer before.

[12] More precisely, among high- and middle-productivity workers, 44% of workers in OVER believe they are among the five most productive workers while only 15.5% hold this belief in UNDER (two-sided Fisher exact p<0.01). Note that while these workers have a dominant strategy to reject all offers, given their belief, we observe that they reject 90% of offers in OVER and 73% of offers in UNDER (p=0.10). Second, among high- and middle-productivity workers, 49% and 40% of them believe that they are among the middle workers in OVER and UNDER respectively (two-sided Fisher exact p=0.52). These workers reject 73% and 64% of offers in OVER and UNDER respectively (p=0.32). Finally, only 4% and 9% of workers stated the belief to be among the three least-productive workers, and there is no significant difference in their acceptance behavior (p=0.33).

Female is never significant in any model of Table 4. This indicates that women are not more likely to accept early offers than men.

The dashed line of Figure 2 presents the average acceptance rates of early offers in RANDOM depending on the rank of the firm that is making the offer. Note that we have no clear theoretical prediction regarding the comparison of acceptance behavior in RANDOM with OVER and UNDER, but in the data, the decisions taken in RANDOM are more similar to OVER than to UNDER. In fact there is no significant difference between acceptances of early offers in OVER and RANDOM, while early offers are less likely to be accepted in RANDOM than in UNDER (see Appendix B for a more detailed analysis of the equilibrium choices in RANDOM and Appendix C for the regressions investigating the differences between RANDOM and treatments OVER and UNDER).

4.4 Gender differences in beliefs but not in acceptances of early offers in OVER and UNDER

Our results on gender differences so far show that in OVER and UNDER women are less confident than men when controlling for productivity, but that they do not accept early offers more often than men. In this subsection we analyze gender differences in acceptances and rejections for a given guessed rank.[13]

The difficulty with treatments OVER and UNDER is that the best responses to guessed ranks depend on the beliefs about other workers. However, as a thought experiment, we define behavior to be in line with beliefs if a worker accepts an offer coming from a firm that has a better or equal rank than her guessed rank and if she rejects an offer coming from a firm that has a worse rank than her guessed rank. These represent optimal choices of naïve players who disregard the selection effects in stage two and simply assume that the assortative matching of *all* workers and firms is the outcome of stage two. As argued above, such choices are always optimal for workers with a guessed rank among the best five or the bottom three workers. However, other workers might take actions which are not in line with their belief according to this definition, but which can be rationalized by their beliefs about the firms making early offers and the acceptance behavior of other workers. Furthermore, since this subsection is devoted to gender comparisons, it is enough to make the assumption that men and women hold similar beliefs about the firms' and the other workers' behavior to be able to study gender differences.[14] Finally,

[13] Note that in this section we refrain from analyzing decisions in RANDOM, since holding any belief other than 8.5 is irrational. We will turn to the analysis of individual decisions in RANDOM in the next section.

[14] If women and men differ in their beliefs about the decisions of other workers, this could rationalize differences in the observed choice patterns. For example consider the case that women are less confident than men but hold beliefs about others' behavior that make them optimistic about their stage 2 prospect. Women could think that workers who are better than they are think that these workers are bad and therefore accept any early offers, thereby making good firms available for the women at stage 2. This would rationalize fewer early acceptances of women than of men with the same guessed rank. Notice however that it is rather unlikely that women would be both less confident than men and believe that those workers are underconfident whom they believe to be better.

when comparing empirically optimal responses to the naïve responses analyzed here, it turns out that they differ only in 7% of the observations (see Appendix D).

Turning to the results, in OVER there is no difference in the proportion of decisions in line with beliefs between men and women. In UNDER, however, men take decisions in line with beliefs significantly more often than women (p=0.02).[15].

In order to get a more precise picture of what is going on, we split the early offers into four categories: (1) the offers such that the decision dictated by beliefs was to accept which were accepted, (2) the offers such that the decision dictated by beliefs was to accept which were rejected, (3) the offers such that the decision dictated by beliefs was to reject which were rejected and(4) the offers such that the decision dictated by beliefs was to reject which were accepted. Hence, both categories (1) and (3) correspond to decisions in line with beliefs whereas (2) and (4) correspond to deviations from beliefs. Figures 4 and 5 show the proportions of decisions in line with beliefs and those deviating from beliefs separately for each treatment and gender. It appears that men and women do not make the same kind of mistakes, understood as deviations from decisions in line with beliefs.

We focus on the decisions that are not in line with beliefs, i.e., the dotted and striped parts of the bars. In OVER, men were more prone than women to accept offers that were bad compared to their stated beliefs, indicated by the striped part of the bar in Figure 4 (clustered regression p=0.07), while there is no significant difference between men and women's propensity to reject offers that were good with regard to their stated beliefs, indicated by the dotted area (clustered regression p=0.12). Altogether, there is no difference in the proportion of decisions inconsistent with beliefs made by men and women in OVER, but note that women and men deviate in opposite directions.

[15] This is the p-value of the coefficient from the probit regression of the probability of taking decisions in line with beliefs on the female dummy with standard errors clustered at the session level.

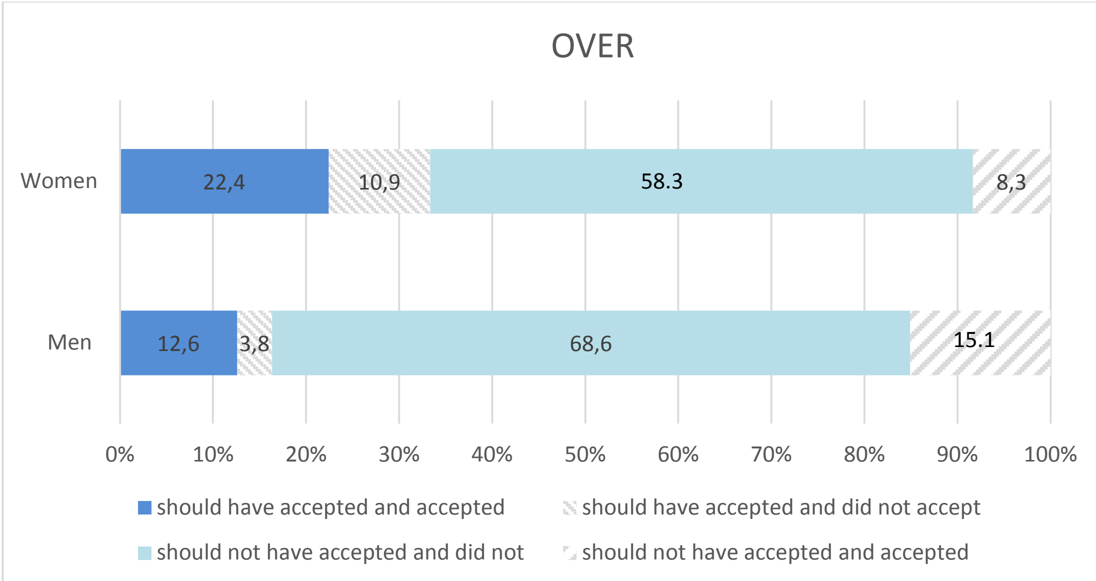

Figure 4. The proportions of decisions in line with beliefs for each gender in OVER

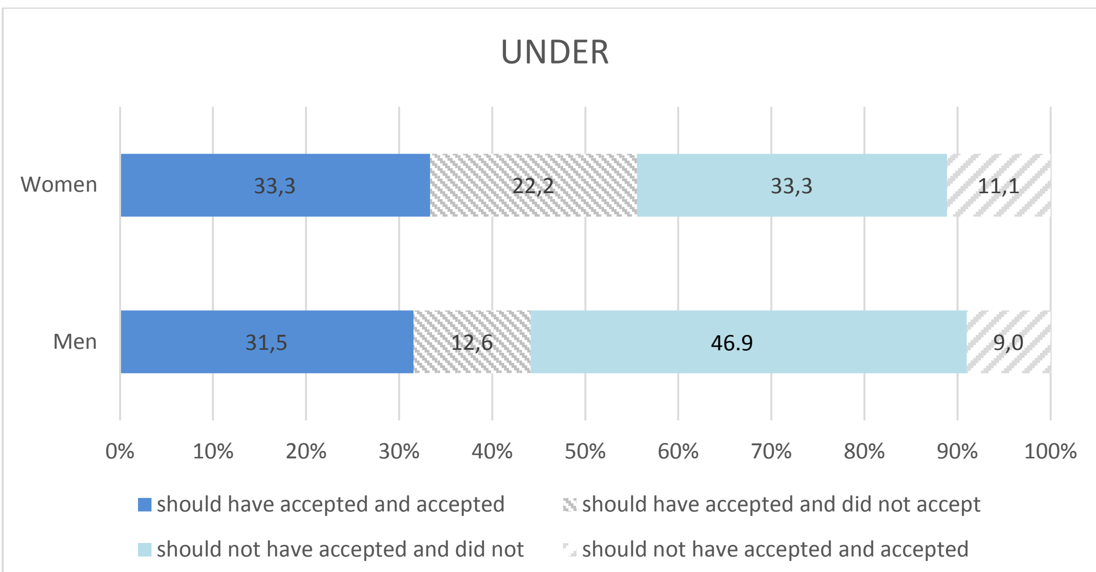

Figure 5. The proportions of decisions in line with beliefs for each gender in UNDER

In UNDER, women are more prone than men to reject offers that were good with regard to their stated beliefs (clustered regression $p<0.01$) while there is no gender difference regarding the second type of mistake (clustered regression $p=0.53$). This adds up to a greater overall rate of deviations from beliefs for women in UNDER.

Thus, across both treatments women reject too many offers given their guessed rank compared to men. This can explain the observation that although women are less confident than men (also when controlling for productivity), they do not accept early offers more often than men. Put differently, women are more self-confident as expressed by their choices than what the elicitation of their beliefs reveals. This is a puzzling finding, since higher risk aversion of women would predict more early acceptances.

There are several possible explanations. One potential reason is that women are too humble when assessing their relative productivity, but when they are confronted with a mediocre offer, they have the correct impulse to reject it (Ibanez et al 2009). Daubmal et al. (1992) observe that women express less optimistic beliefs about their performance in public than in private. This resonates well with our findings if women treat the belief task as a public representation of their confidence while actions as more private since they only indirectly reveal their confidence level.[16] Finally in line with evidence from psychology (Estes and Sydney 2012), women might prefer to avoid binding decisions: accepting an offer is an active choice that may turn out to be wrong while rejecting an offer delegates the determination of the payoff to a fair system, the clearinghouse, that produces the assortative matching in the second stage.[17]

4.5 Payoffs of workers

Since our manipulation of beliefs affected the proportion of optimal decisions, we study how the treatment matters for the workers' payoffs. The realized payoffs of workers depend on random factors such as whether they received an early offer from a particular firm and whether it was a real offer. To check for a possible treatment effect on the workers' payoffs, we therefore run simulations.

We do not have enough observations to evaluate each worker *i*/firm *j* pair. Therefore, we create groups of firms in the following way. Two firms end up in the same group if they are accepted and rejected by the same proportion of workers. To do this, we use Kolmogorov-Smirnoff tests of the equality of distributions for pairs of firms. We conclude that in UNDER, offers from firms 6, 7, 8 and 9 are treated in the same way (i.e. workers are about as likely to accept offers from firms 6, 7, 8, 9), and offers from firms 10, 11, 12 and 13 are also treated in the same way but differently than the first group of firms.[18] Note that this observed grouping coincides with our categorization of firms into those of low and intermediate quality.

16 This interpretation is also in line with the results of Ludwig and Thoma (2014) and Thoma (2016), showing that women are ashamed to overstate their performance, unlike men.

[17] Alternatively, women might be less confident in their beliefs, leading them to behave in line with them less often. However this cannot explain why they mostly deviate from their beliefs in the direction of rejecting offers that are better than their guessed rank.

[18] Testing for the equality of distributions of acceptances, we get the following p-values: p=1.00 for firms 6 and 7, p=0.72 for firms 6 and 8, p=0.47 for firms 6 and 9, p<0.01 for firms 9 and 10, p<0.01 for firms 6 and 10, p<0.01 for firms 9 and 13, p=1.00 for firms 10 and 13, and p=1.00 for firms 10 and 12.

In OVER, the results are slightly different: the tests indicate that offers from firms 6, 7 and 8 are treated in the same way by workers, and offers from firms 9, 10, 11, 12 and 13 are also treated in the same way but differently than the first group of firms.[19] Indeed, the distribution of acceptances for firm 9 is significantly different or close to being significantly different from that for firms 6, 7 and 8 (p=0.11, 0.04 and 0.14 respectively) but it is not different from that for firms 10, 11, 12 and 13. Thus, in OVER, the worst of the middle-quality firms, firm 9, is treated by the subjects just as the low-quality firms, which is consistent with workers being overconfident.

Dep var : Payoff	Model (1)
OVER_dummy	4.25*** (0.06)
Productivity	-3.33*** (0.00)
productivity*OVER	-0.36*** (0.01)
Constant	53.89*** (0.04)

Table 7. OLS regression of workers' payoffs from simulated data. Notes: Values in parentheses represent standard errors. *p<0.1,**p<0.05,***p<0.01

We calculate the average acceptance rate of early offers from each group of firms by each worker, that is, for workers at each rank separately.[20] First, we fix the probability of each firm to make early offers. Then we run 10,000 simulations and record the payoffs of each worker. Table 7 presents the results of regressing the payoffs of the workers in the simulated markets on the treatment dummy and Productivity, that is, the performance rank.

Workers on average earn higher payoffs in OVER compared to UNDER, documented by the significance of the treatment dummy. The payoff difference is due to the best five workers making it to stage two more often in OVER than in UNDER and thereby realizing a payoff of 50. Indeed, the only possible situation where the sum of the workers' payoffs is lower than in the case of the assortative matching is the one where some of the top five workers are not hired by one of the top five firms.

Moreover, we find a heterogeneous treatment effect displayed in Table 7: Treatment OVER helps higher-productivity workers whereas UNDER leads to relatively lower payoffs of high-performance subjects. In other words, by manipulating the beliefs exogenously, we create a heterogeneous effect for different groups of workers:

[19] Testing for the equality of distributions of acceptances, we get the following p-values: p =1.00 for firms 6 and 7, p=1.00 for firms 6 and 8, p=0.11 for firms 6 and 9, p=1.00 for firms 9 and 10, p<0.01 for firms 6 and 10, p=0.41 for firms 9 and 13, p=0.81 for firms 10 and 13, and p=0.59 for firms 10 and 12.

[20] Here we group workers by their true rank. Thus, across sessions we have five workers of true rank 1, for example. We calculate the probability of the worker of rank 1 to accept an early offer from the group of middle firms. This probability is different in each of five sessions (in each treatment), but we average it across the five workers that had rank 1 in the same treatment.

overconfidence is relatively more costly for lower-ranked workers (because it can lead them to reject early offers they would have been better off accepting) while underconfidence is especially bad for the most productive types.

We can infer from these findings, together with the observed stronger overconfidence of men than of women, that especially low-performing men and high-performing women are negatively affected by their biased beliefs.

4.6 Stability

In this section, we turn to the analysis of the stability of the market allocations realized in the experiment. Note that unraveling occurred in every session of every treatment; thus the assortative match was never reached. As a continuous measure of the distance between the realized matching and the assortative matching, we calculate the number of blocking pairs in the final allocation. Once again, as in the previous section, we use simulations to generate 10,000 allocations for OVER and UNDER each. The cumulative density functions of the number of blocking pairs by treatment are presented in Figure 6.

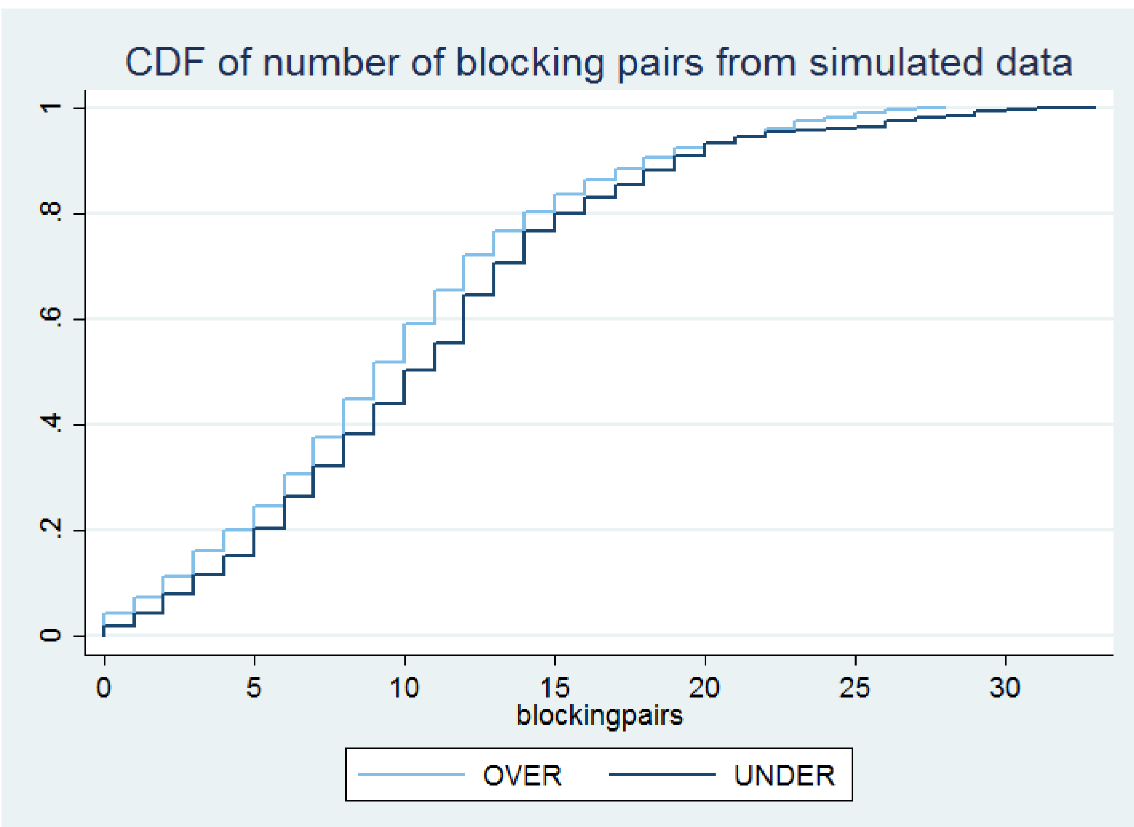

Figure 6. Cumulative density function of number of blocking pairs in simulated data by treatments.

The distributions of the number of blocking pairs are significantly different between UNDER and OVER (Wilcoxon rank-sum two-sided $p<0.01$). Moreover, the distribution of blocking pairs in UNDER first-order stochastically dominates the distribution of blocking pairs in OVER. Thus the exogenous shift of beliefs in UNDER, making subjects less confident, leads to the expected increase in the number of blocking pairs, and thus leads to a matching that is further away from the fair and efficient assortative match than the matching reached in OVER.

4.6 Payoffs, beliefs, and early offers of firms

In this section, we ask which firms make early offers and how this relates to their second-order beliefs about the workers. First, consider the number of early offers made by firms in the three treatments. Note that for the firms, the treatments OVER and UNDER are indistinguishable as we only told the subjects in the role of firms that the workers have to perform a real-effort task and are ranked accordingly. Not surprisingly, we observe no significant difference between the number of early offers in both treatments. In particular, 27 out of 40 firms made an early offer in OVER, and 19 out of 40 in UNDER (Fisher exact test comparing OVER and UNDER: $p= 0.11$). The number of early offers in RANDOM was 15 out of 32 in RANDOM.

We start by investigating the impact of making an early offer on the firm's payoff. Note that the actual payoff of a firm depends on a random outcome, namely the first worker who is randomly chosen to receive the offer and who accepts it. In order to calculate the expected profits from early offers more precisely, we run simulated markets based on the proportion of offers that are accepted or rejected by the workers for each treatment.

We calculate the average acceptance rate of early offers for each group of firms (as defined in the previous subsection) by each worker. Given this average acceptance rate, we then calculate the payoff of each firm, keeping the behavior of other firms fixed, by simulating the entire market. We run 2,000 simulations for each firm, 1,000 assuming that the firm made an early offer and 1,000 assuming the firm did not make an early offer. We test whether the resulting distributions are different. Figure 3 shows the average simulated payoffs for each firm with or without having made an early offer.

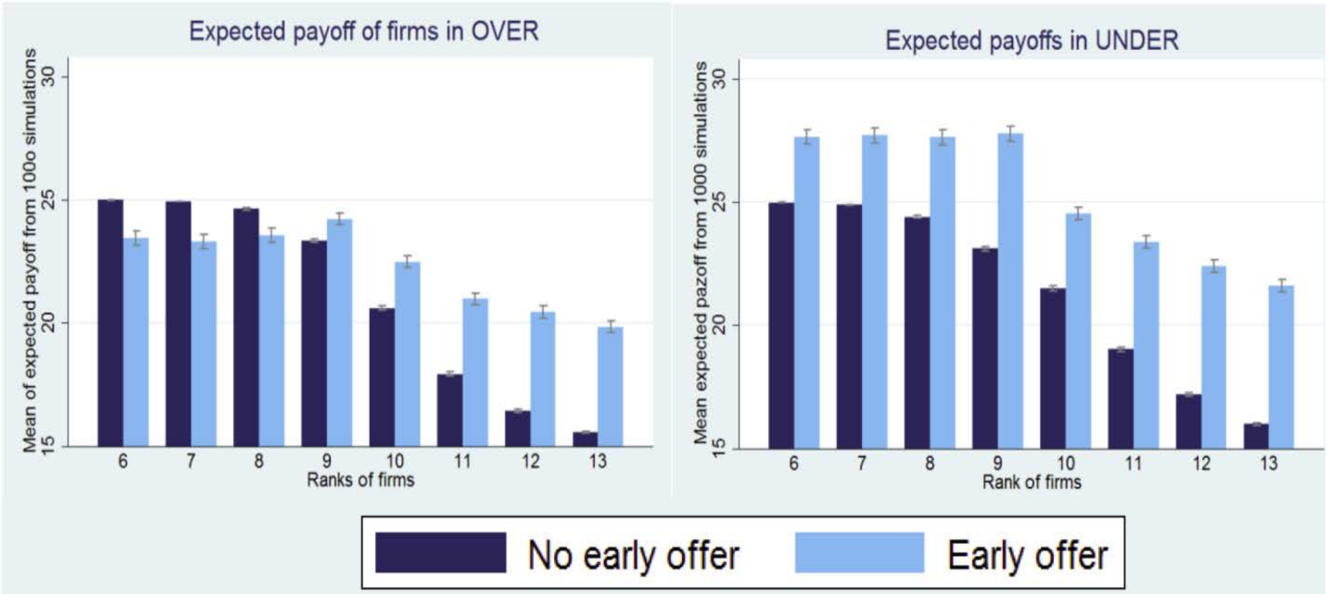

Figure 6: Expected payoffs of firms with and without an early offer, depending on the firms' ranks in treatments OVER and UNDER.

In OVER, presented in the left panel, firms 6, 7, and 8 on average lose money by making early offers while all lower-quality firms profit from early offers.[21] In contrast, it is evident from the right panel of Figure 6 that all firms in UNDER on average benefit from making an early offer. This is because underconfidence will push good workers to accept early offers and top five firms cannot make early offers in our setting as they would never be better off by making them. Thus, low-quality firms always profit from early offers while it matters for medium-quality firms that workers are not too overconfident.

The question is whether firms understand the relationship between the self-confidence of workers and the profitability of early offers. To address this question, we asked participants in the role of firms for their second-order beliefs, that is, the beliefs that the firms hold about the perceived rank of each worker. Remember that firms were asked to guess the guessed rank of each of the 16 workers. Firms do not know the task that workers had to perform (IQ questions or additions) in UNDER and OVER, and we also see no difference between second-order beliefs between the two treatments, see Figure 7. Note that on average firms believe that workers on ranks 1 to 6 are underconfident while workers on ranks 7 to 16 are overconfident, a pattern that implies the belief that the workers' beliefs are biased towards the average rank.

[21] The reason that for the middle-quality firm 9 the expected payoff in the assortative match is lower than for the other middle firms is that the probability that at least one middle-quality worker has left the market in the first stage is quite high. Thus, firm 9 is likely to be matched with a low-ability worker in the assortative match .

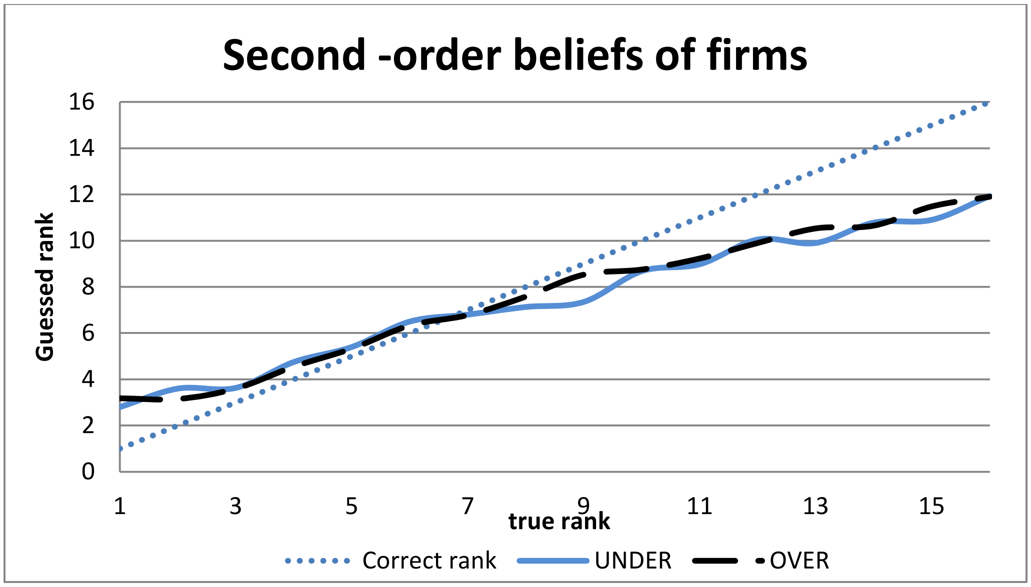

Figure 7: Average second-order beliefs by firms on the workers' guessed ranks.

We can finally turn to the decisions made by the firms and the consequences of their decisions in terms of profits. Remember that firms 1 to5 were computerized and therefore only firms 6 to13 have the possibility of making an early offer. We regress the early-offer dummy (which equals 1 if the firm decided to make an early offer) on the firm's average second-order belief, a dummy that is 1 for female participants and the firm's rank. The variable Second-order belief is the average of all 16 guesses of a firm. A higher average second-order belief indicates that the firm believes the workers to be less confident. The regression shows that firms are more likely to make an early offer, the more they believe workers are underconfident, as can be seen from the positive and significant coefficient of Second-order belief in Table 8. The probability of making an early offer decreases with the quality of the firm while the gender of the subject in the role of the firm has no effect.

Dep var : Early offer	Model (1)
Second-order belief	0.092** (0.044)
Female	0.009 (0.116)
Firm rank	0.06** (0.026)

Table 8: Marginal effects of probit regression of early offers on second-order beliefs of firms, i.e., averages over beliefs for all workers for a given firm. Based on data from OVER and UNDER. Values in parentheses represent standard errors. *p<0.1,**p<0.05,***p<0.01

5. Conclusions

We have designed an experimental labor market to investigate the effects of over- and underconfidence on the market outcome. With the help of two different real-effort tasks, each of them employed in a separate treatment, we shift the subjects' self-confidence and observe an effect of this shift in beliefs on the market outcome. Namely, we see that in the treatment with the easy task, fewer offers are being accepted early than in the case of the hard task. As a consequence, underconfidence of workers causes the matching to be less stable than when workers are overconfident. We find that apart from shifting the beliefs of the workers, the treatment has no additional effect on the acceptance of early offers.

Female participants in our experiment are less self-confident than male participants but they are not more likely to accept early offers than the men. This is puzzling since we also observe that self-confidence is negatively correlated with the acceptance of early offers. In line with this discrepancy between the women's beliefs and actions, we find that women best-respond to their stated beliefs less often than men. In particular, in the treatment with the hard task (UNDER), women reject "good" offers (offers from firms that are better than their assortative match given their stated beliefs) more often than men. One possible explanation for this finding is that women are more humble when stating the belief about their relative performance than when making real decisions. Alternatively, the procedure in stage two that guarantees the assortative matching may seem relatively more attractive to women than to men.

Whether early offers of firms increase their profits, depends on the quality of the firm and the self-confidence of workers: In treatment UNDER, all medium and low-ranked firms profit from making early offers. In contrast, in OVER only the worst firms benefit from making early offers. Consistent with this, we observe that the more underconfident the firms believe the workers to be, the more early offers the firms make. This is evidence that firms understand the relationship between the self-confidence of workers and the profitability of early offers.

Regarding the implications of our results, note that unraveling effects similar to the ones described in our experiment not only occur when there is a centralized clearinghouse using a stable mechanism (such as in some US markets for doctors or the Canadian market for young lawyers, see Roth and Xing 1994). Unraveling that is affected by the applicants' self-confidence can also be at play in completely decentralized markets where some offers are made earlier than others. For example, the market for apprenticeships in Switzerland has unraveled such that many apprenticeships are filled more than a year before applicants have finished school. This makes it likely that applicants who do not know their final grades make suboptimal acceptance decisions that depend on their self-confidence.[22] Moreover, our results indicate that the amount of unraveling that a market experiences is

[22] See the article in the *Neue Züricher Zeitung* on August 19, 2016 (http://www.nzz.ch/wirtschaft/wirtschaftspolitik/berufsbildung-die-heikle-jagd-nach-den-besten-lehrlingen-ld.111764).

affected by the general level of overconfidence of the applicants. Overoptimism and overconfidence limit unraveling while lack of confidence can be a source of unraveling and instability. We believe that the identification of such instances, e.g., where beliefs are shifted due to a change in the modalities of a final exam or in the rules of admission, is a worthwhile future endeavor.

Finally, market designers need to take into account the possibility of cultural differences between groups of applicants, between women and men, and individual differences that are reflected in the relative self-assessments, since they can affect the labor market success. The evidence on the penalty for underconfident women for university admissions in certain regions of China illustrates such effects nicely (Ma et al. 2016, Pan 2016). Market design can protect biased agents in a number of ways. Apart from restrictions on early contracting, early and frequent feedback on individual performance can reduce the biases of agents.

References

Avery, Christopher, Christine Jolls, Richard A. Posner, and Alvin E. Roth "The market for federal judicial law clerks." *The University of Chicago Law Review* (2001). 74: 793-902.

Bardsley Nick. “Control Without Deception: Individual Behaviour in Free-Riding Experiments Revisited.” *Experimental Economics* (2000). 3: 215-40.

Costa-Gomes, Miguel A., Steffen Huck, and Georg Weizsäcker. “Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect.“ *Games and Economic Behavior* (2014). 88: 298-309.

Damiano, Ettore, Hao Li, and Wing Suen. “Unravelling of Dynamic Sorting”. *The Review of Economic Studies* (2005). 72.4: 1057-1076.

Daubman, Kimberly A., Laurie Heatherington, and Alicia Ahn. "Gender and the self-presentation of academic achievement." *Sex Roles* (1992). 27.3-4: 187-204.

Du, Songzi and Yair Livne. “Rigidity of Transfers and Unraveling of Matching Markets.” (2016). Mimeo.

Echenique, Federico, and Juan Sebastián Pereyra. "Strategic uncertainty and unraveling in matching markets." *Theoretical Economics* (2016), 11(1): 1-39.

Estes, Zachary, and Sydney Felker. "Confidence mediates the sex difference in mental rotation performance." *Archives of Sexual Behavior* (2012) 41.3: 557-570.

Fréchette, Guillaume R., Alvin E. Roth, and M. Utku Ünver. "Unraveling yields inefficient matchings: evidence from post-season college football bowls." *The RAND Journal of Economics* (2007). 38.4: 967-982.

Hałaburda, Hanna. "Unravelling in two-sided matching markets and similarity of preferences." *Games and Economic Behavior* (2010). 69.2: 365-393.

Jensen, Robert. “The (perceived) returns to education and the demand for schooling," *Quarterly Journal of Economics* (2010). 125(2): 515-548.

Kagel, John H., and Alvin E. Roth. "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment." *Quarterly Journal of Economics* (2000). 115.1: 201-235.

Kruger, Justin. "Lake Wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments." *Journal of Personality and Social Psychology* 77.2 (1999): 221.

Li, Hao, and Sherwin Rosen. "Unraveling in matching markets." *The American Economic Review* (1998): 371-387.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence Phillips. “Calibration of probabilities: The state of the art to 1980.” in D. Kahneman, P. Slovic and A. Tversky *Judgement under Uncertainty: Heuristics and Biases*. 306-334.

Ludwig, Sandra, & Thoma, Carmen “Do women have more shame than men? An experiment on self-assessment and the shame of overestimating oneself.” Mimeo, 2014.

Ma, Mingming, Binzhen Wu, and Xiahan Zhong. “Matching mechanisms, matching quality, and strategic behavior: evidence from China.” Mimeo, 2016.

McKinney, C. Nicholas, Muriel Niederle, and Alvin E. Roth. “The collapse of a medical labor clearinghouse (and why such failures are rare).” *American Economic Review* (2005). 95(3): 878-889.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. „Managing self-confidence: Theory and experimental evidence.“ Mimeo, 2014.

Moore, Don A., and Paul J. Healy. “The trouble with overconfidence.” *Psychological Review* (2008). 115 (2), 502-517.

Niederle, Muriel, and Alvin E. Roth. "Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match." *Journal of Political Economy* (2003). 111.6: 1342-1352.

Niederle, Muriel, and Alvin E. Roth. "The gastroenterology fellowship match: how it failed and why it could succeed once again." *Gastroenterology* (2004). 127.2: 658-666.

Niederle, Muriel, Alvin E. Roth, and M. Utku Ünver. "Unraveling results from comparable demand and supply: An experimental investigation." (2013). *Games* 4.2: 243-282.

Pan, Siqi. “The instability of matching with overconfident agents: laboratory and field investigations.” Mimeo, 2016.

Roth, Alvin E., and Xiaolin Xing. "Jumping the gun: Imperfections and institutions related to the timing of market transactions." *The American Economic Review* (1994). 84: 992-1044.

Roth, Alvin E., and Xiaolin Xing. "Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists." *Journal of Political Economy* (1997). 105.2: 284-329.

Roth, Alvin E. "A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom." *The American Economic Review* (1991). 81.3: 415-440.

Schwardmann, Peter and Joël van der Weele. „Deception and Self-Deception." Mimeo 2016.

Thoma, Carmen. "Under-versus overconfidence: an experiment on how others perceive a biased self-assessment." *Experimental Economics* (2016). 19: 1-22.

APPENDIX

Appendix A

Proof of the equilibrium prediction in treatment RANDOM (Section 3.1)

The following calculations prove that the unique equilibrium in treatment RANDOM consists of the following strategies: the top five firms do not make early offers, the middle firms (firms 6-9) make early offers which are accepted, and the workers always reject offers from bad firms (firms 10-13) which make these firms indifferent between making early offers or waiting for stage 2.

We start by showing that firm 1 will always choose not to make an early offer. The expected payoff from making an early offer is 27.5 (as offers from top firms will always be accepted). Note that the situation leading to the worst possible expected payoff from not making an early offer for a given firm is when all better firms wait for stage 2 while all worse firms make early offers and leave the market. For firm 1, this corresponds to the situation where all other 12 firms make early offers which are accepted. Let us now compute the expected payoff of firm 1 from not making an early offer in this situation. Firm 1 can get three possible payoffs from its behavior. First, it can get a payoff of 50 is at least one of the top five worker did not receive an early offer. This happens with probability 1-(C(5,5)*C(11,7))/C(16,12)=1490/1820. Second, it can get a payoff of 25 if all of the top five workers received an early offer and at least one of the 4 good workers did not. This happens with probability [C(5,5)*C(4,3)*C(7,4)+C(5,5)*C(4,2)*C(7,5)+C(5,5)*C(4,1)*C(7,6)+C(5,5)*C(7,7)]/C(16,12)=295/1820.

Finally, it can get a payoff of 15 if all of the nine top and good workers received an early offer. This happens with probability C(9,9)*C(7,3)/C(16,12)=35/1820. This yields an expected payoff from not making an early offer of 45.27 that is higher than the payoff from making an early offer.

Now, knowing that firm 1 will always choose not to make an early offer, what will firm 2 do? To see this, let us again consider the situation where the expected payoff from not making an early offer is minimal (the situation where firms 3-13 make early offers) and show that even then, the expected payoff from waiting until stage 2 will exceed that of making an early offer. Firm 2 can then again get three possible payoffs. It will get a payoff of 15 if at least eight out of the nine top and good workers receive and accept an early offer. This happens with probability [(C(9,8)*C(7,3))+ (C(9,9*C(7,2))]/C(16,11)=336/4368. Then, if at least four out of the five top workers but a maximum of 7 out of the 9 top and good workers receive and accept an early offer (probability of [C(5,4)*(C(4,0)*C(7,7)+C(4,1)*C(7,6)+C(4,2)*C(7,5)+C(4,3)*C(7,4))+C(5,5)*(C(4,0)*C(7,6)+C(4,1)*C(7,5)+C (4,2)*C(7,4))]/C(16,11)=1776/4368), Firm 2 will get a payoff of 25. Firm 2 will finally get a payoff of 50 if a maximum of 3 out of the 5 top workers receive and accept an eo (probability=[C(5,0)*C(11,11)+C(5,1)*C(11,10)+C(5,2)*C(11,9)+C(5,3)*C(11,8)]/C(16,11)==2256/4368). The expected payoff of waiting for stage 2 is therefore at least equal to 37.14, higher than the expected payoff from making an early offer.

Using the same reasoning, we can show that firm 3 will also choose not to make an early offer. Indeed, knowing that firms 1 and 2 will not make early offers, the lowest possible expected payoff of firm 3 from waiting for stage 2 (obtained when firms 4-13 make early offers which are accepted is 29.23, again higher than the expected payoff from making an early offer. Indeed, the probabilities that firm 3 will get payoffs of 15 (case a), 25 (case b) and 50 (case c) are, respectively: P(a)=[(C(9,7)*C(7,3)+C(9,8)*C(7,2)+C(9,9)*C(7,1)]/C(16,10), P(b)=[C(5,3)*(C(4,0)*C(7,7)+C(4,1)*C(7,6)+((4,2)*C(7,5)+C(4,3)*C(7,4))+C(5,4)*(C(4,0)*C(7,6)+C(4,1)*C(7, 5)+C(4,2)*C(7,4))+C(5,5)*(C(4,0)*C(7,5)+C(4,1)*C(7,4))]/C(16,10) and P(C)=[C(5,0)*C(11,10)+C(5,1)*C(11,9)+C(5,2)*C(11,8)]/C(16,10).

For firm 4, we used a different reasoning by noting that firms 10-13 will still be present at stage 2 (even if they make early offers, workers will reject them). Under the assumption that firms 1-3 and 10-13 will hire workers in stage 2, the worst case scenario for firm 4 is when firms 5-9 make early offers which are accepted. In this situation, firm 4's payoff will be 50 if a maximum of one top five worker received an early offer (probability=[C(5,0)*C(11,5)+C(5,1)*C(11,4)]/C(16,5)=2112/4368) and 25 otherwise, corresponding to a worst case scenario expected payoff of 37.09. Firm 4 will therefore choose not to make an early offer.

Likewise, firm 5 will get a higher expected payoff by not making an early offer in the worst case scenario where, firms 1-4 and 10-13 are still on the market for stage 2 and firms 6-9 have left the market. Indeed, its expected payoff in this situation will be 29.53. This comes from the fact, that it will get a payoff of 50 with probability [C(5,0)*C(11,4)/C(16,4)] and a payoff of 25 otherwise.

Considering the middle firms, given that the top five firms do not make early offers and that possible early offers by firms 13-16 are always rejected, firms 6, 7, 8, and 9 make early offers as they get 27.5 in expectation compared to at most 25 in the assortative matching in stage 2. Those early offers are accepted by the workers, since being matched to a middle firm yields at least a payoff of 29 which is better than 27.125 by waiting for stage 2 even in the best possible scenario where middle firms are still in the market in stage 2 (given that we know that the top five firms and firms 10-13 are still on the market at stage 2). Firms 10, 11, 12, and 13 are indifferent as their offers are always rejected.

Finally, we can show that firms 6-9 will make early offers that will be accepted. Knowing that firms 1-5 and 10-13 will still be present at stage 2, firm 6 will get an expected payoff of 25 in the best case scenario where firms 7,8 and 9 do not make early offers either. It will therefore choose to make an early offer. The same result holds for firms 7, 8 and 9.

Appendix B

Equilibrium decisions in RANDOM

In the treatment RANDOM, we look at whether participants in the role of workers accept and reject offers as predicted by the equilibrium for risk-neutral workers in this treatment. Table 5 shows the observed behavior compared to the equilibrium predictions. Remember that in equilibrium, a risk–neutral worker should accept offers from the middle firms (firm 9 and better) and reject offers from the bad firms (firm 10 and worse). We observe that only 6% and 10% of the decisions made by men and women deviate from the equilibrium when they receive offers from bad firms (p=0.52). These acceptances could be explained by risk aversion or, alternatively, by overly pessimistic beliefs about their random draw. More deviations are observed in the form of rejections of offers from middle firms, with rejection rates of 29% and 40% for men and women respectively (p=0.37). This could be interpreted as over-optimism with respect to the random draw. Overall, 16% of decisions made by men and 22% by women are not in line with the equilibrium (p=0.38).

Men	Reject	Accept	% out-of equil.
Offers from bad firms	48	3	6%
Offers from middle firms	12	30	29%
Total men			16%
Women			
Offers from bad firms	65	7	10%
Offers from middle firms	19	29	40%
Total women			22%
Total men and women	68%	32%	

Table 5: Acceptance and rejection decisions by workers in RANDOM.

Appendix C

Acceptance of early offers in RANDOM compared to OVER and UNDER

Dep var : accept offer	RANDOM and OVER	RANDOM and UNDER
RANDOM_dummy	0.03 (0.07)	-0.13*** (0.04)
Female	-0.01 (0.08)	0.03 (0.09)
Productivity	0.02*** (0.01)	0.01 (0.01)
Offer	-0.11*** (0.01)	-0.15*** ()0.02)

Table 10: Linear regression of the decision to accept an offer on explanatory variables for treatments RANDOM and OVER or UNDER, with standard errors clustered at the sessions level

Appendix D

Payoff-maximizing decisions by workers in OVER and UNDER

We study how often workers take optimal decisions given the actual choices of the other market participants. This will also allow us to gauge how often the decisions consistent with beliefs that we examine in Section 4.4 are in fact payoff-maximizing.

For each worker we calculate the outcome of the second stage if she would reject all offers, given the actual early matches of all other workers in the session. Then, if a worker accepts an offer which leads to at least as high a payoff as the hypothetical match in the second stage, we count this acceptance as optimal. Similarly for rejections: if the rejected offer leads to a lower payoff than the hypothetical match of the second round, we count it as an optimal rejection.

Panel A of Table 6 presents the proportion of optimal decisions when the empirical best response is to reject the offer (i.e. after receiving a bad offer), by treatment and gender. It is equivalent to the number of optimal rejections divided by the number of total decisions when the optimal decision was to reject. The proportion of payoff-maximizing decisions when “reject” is the best response in OVER is 84%, and in UNDER it is 66%, with the difference being significant ($p<0.01$). There are no significant gender differences in each of the treatments separately, nor in the pooled dataset.

Panel B of Table 6 presents the proportion of optimal decisions in the case when the best response is to accept the offer by treatments and gender (good offers). This is the number of optimal acceptances divided by the number of total decisions when the optimal decision was to accept. Note that the difference in the proportions of optimal decisions between the treatments is reversed relative to the case when the optimal decision is to accept, i.e., treatment OVER helps people to reject bad offers while UNDER helps them to accept good offers, but the difference is not significant (p=0.19). There are no gender differences either. When pooling optimal acceptances and optimal rejections in Panel C, the difference between optimal choices in OVER and UNDER is significant (p=0.02), but the gender difference is still not significant (p=0.47).

Panel A. Optimal decisions with bad offers			
Treatment	Men	Women	Both genders
UNDER	69%	64%	66%
OVER	85%	83%	84%
Both treatments	78%	75%	
Panel B. Optimal decisions with good offers			
Treatment	Men	Women	Both genders
UNDER	53%	57%	55%
OVER	46%	45%	45%
Both treatments	49%	44%	
Panel C. Optimal decisions overall			
Treatment	Men	Women	Both genders
UNDER	62%	61%	62%
OVER	69%	64%	67%
Both treatments	66%	63%	

Table 6. Proportion of optimal decisions, grouped by treatment, gender, and the optimal action (acceptance or rejection). Good offers are offers which should be accepted given the actual behavior of others, and bad offers should be rejected.

Note that participants are better at rejecting offers that should be rejected than at accepting offers that should be accepted both in OVER and UNDER (p=0.05 for UNDER and p<0.01 for OVER). In OVER, the difference is significant for both men and women (p=0.03 for men and p<0.01 for women). In UNDER, the difference is only marginally significant for men (p=0.09), and is not significant for women (p=0.56), consistent with the fact that women hold too low beliefs about themselves in UNDER, making it less likely that they reject bad offers, but that they are also slightly better than men at identifying good offers in UNDER.

Appendix E

Instructions

Instructions

General description

This experiment is about **workers** who try to find the best possible job and **firms** looking for the best possible workers. Each firm wants to employ exactly one worker. At the beginning of the experiment, it will be determined randomly whether your will be in the role of a worker or of a firm. You will keep this role for the entire experiment.

There are 13 firms in the market of which five are of high quality and the remaining eight are each of a different quality. There are also 16 workers who prefer to be matched to the firms of high quality relative to the firms of intermediate or lower quality.

The workers have different productivities for the firm. That is, the workers can be ranked in terms of their productivity, with one worker being the most productive and one worker being the least productive. All firms have identical preferences over workers, that is, they agree on which worker is the most productive etc.

Each round of the experiments consists of two stages.

First stage

At the beginning of the first stage, the quality of the firms is revealed to all participants. The productivity of the workers is not known, neither by the firms nor by the workers themselves.

During the first stage, all middle- and low-quality firms are allowed to make early offers to the workers. They do not differentiate between the workers, thus they make an offer, if any, to a random worker. Each worker is free to accept or reject the offer. The workers have 30 seconds to submit the decision of acceptance or rejection. If the offer is accepted, both the firm and the worker leave the market. The first stage consists of maximum of nine rounds. Thus if an offer of a firm was rejected by a worker it will be automatically sent to another worker. If it is rejected again it will be send to the third one. Thus any offer can be rejected a maximum of nine times. (Note that the procedure of distributing offers guarantees an offer is always sent to a new worker.)

Every worker who receives an offer receives two more fictitious offers. A worker must decide whether or not to accept each of the three offers. This means that he can accept all three offers, only two of them, only one or none. He does not know which offer is real and which offers are fictitious. If he rejects an offer, he will not get any more offers from this firm, independent of whether it was a real or a fictitious offer.

Now consider the following example: Firm 12 makes an offer to a randomly selected worker. The worker sees three offers on his screen: two randomly selected offers and the offer of firm 12. Let us suppose that the offers of firms 9 and 11 were made randomly. Let us also suppose that the worker accepts the offer of firm 9 but rejects the offers of firms 11 and 12. The worker is then told that only the offer of firm 12 was real, so he is still unmatched.

The rejection of the offer of firm 11 is final. This means that the worker will not receive an offer from firm 11 anymore. This means that a worker should consider each of the three offers as if it is the only one that he has received.

Second stage

All workers and firms who remain unmatched at the end of the first stage (that is, firms who decided to wait and did not make early offers, and firms whose offer was rejected) move on to the second stage.

At the beginning of the second stage, the quality of all workers is revealed. Moreover, it will be announced which firms and workers have already left the market in the first stage. Then the following matching is implemented: the five best unmatched workers are assigned to the five best firms, the sixth[s] best unmatched worker is assigned to the sixth best firm and so on. The three workers of the lowest productivity among all workers at the second stage remain unmatched and receive a payoff of 0.

Workers and firms only have to make decisions in the first stage. The second stage is executed by the computer, according to the above description.

Information and Payoffs

The payoffs of firms and workers in all rounds of the experiment have the same structure and are presented in the tables below.

Payoffs of the firms

	Most productive workers 1-5	Workers 6-9	Workers 10-13	Least productive workers 14-16
Payoff of firm (points)	50	25	15	10

Thus all firms receive 50 points if they are matched with any of the best five workers, 25 if they employ a worker who is ranked 6th to 9th, 15 for a worker ranked 10th to 13th, and 10 if they employ any of the least productive three workers.

Payoffs of the workers

The payoff of a worker depends on which firm it concludes a contract with. For a contract with one of the top five firms, the five most productive workers receive a payoff of 50, while all other workers receive a payoff of 32 for these firms. For all other firms the payoff is equal for all workers. For example, every worker who signs a contract with the tenth-best firm receives a payoff of 17, with the eleventh best firm only 16, etc.

All workers	Five most productive workers	All other workers
Top Firm	50	32
2nd best firm	50	32
3rd best firm	50	32
4th best firm	50	32
5th best firm	50	32
6th best firm	32	32
7th best firm	31	31
8th best firm	30	30
9th best firm	29	29
10th best firm	17	17
11th best firm	16	16
12th best firm	15	15
13th best firm	14	14
Unassigned	0	0
Unassigned	0	0
Unassigned	0	0

Rounds

There are a total of four rounds. The first three are practice rounds and do not affect your payoff. Only in the fourth round your payoffs will be determined. In each practice round, the rank of the participants in the role of the workers is drawn anew.

Productivity of workers

In each practice round the productivity of the workers is determined by a random draw. That is, the computer randomly assigns a productivity rank to each worker. A new ranking is randomly drawn in every round. The ranks thus determined are communicated to all workers and to all firms at the beginning of the second part.

In the third round of the experiment, which is relevant for the payoffs, the ranks of all workers are determined as follows:

- All workers are asked to work on a task for which they earn points.
- Whoever has reached the most points within a certain time is the most productive worker. The one who has reached the second most points is the second most productive worker and so on.
- If two or more workers are equal, the relative ranking of these workers is determined by chance.

Note that in the first stage, only firms of intermediate and low quality can make offers to workers. Thus, the only way for workers to be employed by one of the best five firms is to stay in the market until the second stage. Moreover every time a worker receives an offer in the first stage, he knows which firm made the offer, and thus the corresponding payoff. At the same time, every participant in the role of the firm knows its own quality. Also, keep in mind that a firm that makes an offer in the first part does not know what productivity the workers have.

If you have any questions about the experiment, please raise your hand.

Additional instructions for belief elicitation stage

(distributed in the final payoff-relevant round after subjects have worked on the real-effort task)

Belief elicitation

For workers:

All 16 workers including you have completed a task. According to the performance of all workers, each of them is attributed a rank. Rank 1 corresponds to the worker whose performance was best, rank 2 to the worker whose performance was the second best, and so on. We ask you to state your expectation about your rank. More precisely, we want you to tell us which rank you think you most likely have reached. The answer must be an integer between 1 and 16.

How is the payoff for your stated beliefs calculated? Your payoff is calculated based on your guess about which rank most likely corresponds to your relative performance.

Your payoff depends on the difference between your expected rank and your actual rank based on your performance. Your payoff is higher if you have stated a rank close to your true rank (which corresponds to the rank of your performance among all other workers' performances), and it is lower if you have stated a rank far away from your true rank.

The exact calculation of the payoff is as follows: We calculate a number reflecting how close your guess of your rank and your true rank is. We take this number to calculate your payoff. We consider the difference between your estimate of your rank and your true rank. We then multiply this difference by itself and the resulting number is multiplied by the factor 0.25. Thus, if you guessed a rank close to your true rank, this number will be smaller than if you guessed a rank far away from your true rank.

We then take the number thus calculated and deduct it from 5. This determines the number of points you receive for your statement of confidence (if the number of points obtained is smaller than 0, you get 0 points for this task).

As an illustration of how your payoff is calculated, let us consider three examples. Let us assume that your true rank is 10 and that you correctly guessed rank 10. This means that you have stated a rank that is exactly right. Consequently, you earn the following number of points:

$$5-0{,}25*(10-10)^2=5$$

Let us assume again that your true rank is 10. But now your guessed rank is 6. Consequently, you earn the following number of points:

$$5-0{,}25*(6-10)^2=1$$

Let us still assume that your true rank is 10, but your estimated rank is 2.

Consequently, you earn the following number of points:

$$5-0{,}25*(2-10)^2=-11$$

As this is a negative number, you will earn 0 points.

Please note that the numbers used in the examples have been chosen arbitrarily. They provide no indication for how you should decide.

The examples demonstrate that you will always receive a payoff of at least 0 points, and at most 5 points for your stated beliefs. And the closer your beliefs to the truth, the more money you earn. (You may be asking yourself why we have chosen such a payoff rule as described above. The reason is that with this payoff rule, you can expect the highest payment when you state the number that is closest to your true belief.)

For firms:

We have just asked the workers to tell us what they think their rank is. The question was: "All 16 workers including you have completed a task. According to the performance of all workers, each of them is attributed a rank. Rank 1 corresponds to the worker whose performance was best, rank 2 to the worker whose performance was the second best, and so on. We ask you to state your expectation about your rank. More precisely, we want you to tell us which rank you think you most likely have reached. The answer must be an integer between 1 and 16."

Please tell us what you think each worker thought about his true rank. In particular, please tell us which rank the worker thought is his most likely rank. Of course, the best performing workers are likely to have guessed differently than the worst performing ones. For this reason, you will be asked to guess the estimated rank of each of the 16 workers.

How is the payoff for your stated beliefs calculated? Your payoff is calculated after you have guessed the rank stated by all 16 workers. An integer between 1 and 16 will be chosen randomly and your payoffs will depend on the accuracy of your beliefs about the worker of the rank chosen randomly. For instance, if the number 3 is chosen randomly, your payoff will depend on the difference between your belief of the rank guessed by the worker whose true rank is 3 and the rank actually guessed by this worker.

Your payoff is higher when you have stated a rank close to the rank guessed by the selected worker, and it is lower when you have stated a rank far away from the rank guessed by worker.

The exact calculation of the payoff is as follows: For the randomly chosen worker, we calculate a number reflecting how close your estimate of the rank guessed by this worker is to the actual rank guessed by this worker. We take this number to calculate your payoff. We consider the difference between your estimate of the rank guessed by the randomly chosen worker and the actual rank guessed by this worker. We then multiply this difference by itself and the resulting number is multiplied by the factor 0.25. Thus, if you stated a rank close to the rank guessed by the randomly chosen worker, this number will be smaller than if you stated a rank far away from the rank guessed by this worker.

We then take the number thus calculated and deduct it from the number 5. This determines the number of points you receive for your belief statement. If the number of points obtained is smaller than 0, you get 0 points for this task.

As an illustration of how your payoff is calculated, let us consider the following example.

True rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Rank guessed by worker	1	1	6	3	8	10	2	7	8	3	10	15	5	8	6	8
Your estimate of each worker's guess	2	1	4	5	6	7	7	7	8	8	5	15	12	16	8	14

Let us assume that the number 5 was randomly chosen. We have to look at the difference between your estimate of the rank guessed by the worker whose true rank is 5 and the rank actually guessed by this worker. Consequently, you earn the following points for your beliefs:

$$5-0{,}25*(6-8)^2=4\ .$$

Let us now assume that the randomly chosen number was 11. We consider the difference between your estimate of the rank guessed by the worker whose true rank is 11 and the rank actually guessed by this worker. Consequently, you earn the following points for your beliefs:

$$5-0{,}25*(5-10)^2=-1{,}25$$

This is a negative number. Therefore you earn 0 points for your beliefs.

Please note: The numbers used in the examples have been chosen arbitrarily. They provide no indication for how you should decide.

The examples demonstrate that you will always receive a payoff of at least 0 points, and at most 5 points for your stated beliefs. And the closer your beliefs to the truth, the more money you earn. (You may be asking yourself why we have chosen such a payoff rule as described above. The reason is that with this payoff rule, you can expect the highest payment when you state the number that is closest to your true belief.)