Thoma, Carmen:

Is Underconfidence Favored over Overconfidence? An Experiment on the Perception of a Biased Self-Assessment

Is Underconfidence Favored over Overconfidence? An Experiment on the Perception of a Biased Self-Assessment.*

Carmen Thoma[†]

November 19, 2013

Abstract

This paper reports findings of a laboratory experiment, which explores how self-assessment regarding the own relative performance is perceived by others. In particular, I investigate whether overconfident subjects or underconfident subjects are considered as more likable by others, and who of the two is expected to achieve a higher performance in a real effort task. I observe that underconfidence beats overconfidence in both respects. Underconfident subjects are rewarded significantly more often than overconfident subjects, and are significantly more often expected to win the competitive real-effort task. It seems as if subjects being less convinced of their performance are taken as more congenial and are expected to be more ambitious to improve, whereas overconfident subjects are rather expected to rest on their high beliefs. While subjects do not anticipate the stronger performance signal of underconfidence, they anticipate its higher sympathy value. The comparison to a non-strategic setting shows that men strategically deflate their self-assessment to be rewarded by others. Women, in contrast, either do not deflate their self-assessment or do so even in non-strategic situations, a behavior that might be driven by non-monetary image concerns of women.

Keywords: Self-confidence, Overconfidence, Experiment
JEL-Classification: C91, D03, J16

1   Introduction

Earlier, mainly psychological and social psychological studies claim that people are overconfident. However, recent economic studies show that individuals' self-assessment is rather precise or underconfident.[1] As a biased self-assessment can lead to systematic biases in individuals' decision making the topic is of high interest to economists.[2] Yet, barely any research has been done to analyze how overconfidence in comparison to underconfidence is perceived by others, and whether individuals adapt their self-assessment to others' perception.

I use a controlled laboratory study to address this topic, thereby focusing on two aspects: First, I analyze whether underconfident individuals are perceived as more or less likable than overconfident individuals. Secondly, I explore whether under- or overconfidence is perceived as a stronger signal for ambition and effort. These findings contribute to the expanding literature analyzing which advantages or disadvantages overconfidence in comparison to underconfidence involves. Thereby, adding to the questions why individuals might (rationally) exhibit a bias in their self-assessment and in which situations we should expect individuals to over- or underestimate themselves. The perception of one's self-assessment is difficult to analyze in the field as self-confidence interacts with other characteristics in many ways. The anonymous laboratory setting allows me to separate the causal effects of over- and underconfidence on others' appraisal, by only varying the accuracy of subjects' self-assessment.

The experiment consists of two parts. In part 1 all subjects perform an incentivized real effort task which serves as the basis for their self-assessment. In part 2 two thirds of the subjects (agents) are assigned a rank based on their relative performance, whereas each rank is assigned to two subjects. The two agents having the same rank are assigned to one of the remaining participants (principals). Both agents estimate their relative rank and the principal learns by how many ranks each of them over- or underestimated himself.[3] In treatment $SYMP$ the principal chooses to whom of the two agents he wants to give 5 Euros. In treatment $PERF$ the principal has a monetary incentive to choose the agent

---

[1] First psychological and social psychological evidence for overconfidence has sometimes been labelled better-than-average effect (Alicke, 1985; Dunning et al., 1989; Messick et al., 1985; Svenson,1981). In laboratory experiments Hoelzl and Rustichini (2005) show that choice behavior changes from overconfidence to underconfidence when the task changes from easy and familiar to non-familiar. Krueger (1999) and Moore and Cain (2007) also find that people tend to be underconfident rather than overconfident when the task is (perceived as) difficult. Clark and Friesen (2009) test for overconfidence in people's forecasts of their absolute and relative performance and observe a correct self-assessment or underconfidence more often than overconfidence.

[2] For example Niederle and Vesterlund (2007) show that overconfidence makes bad performing men selecting competitive payment schemes too often (regarding payoff maximization), and that underconfidence makes high performing women selecting competitive payment schemes too little.

[3] The type of overconfidence observed in this study is overplacement as termed by Larrick et al. (2007). See also Moore and Healy (2008) for a more precise distinction of the different types of overconfidence.

who performs better in a repetition of the real effort task. The only information the principal gets is the deviation of the agents' self-assessments and the information that both agents have the same actual rank. This element of the design is essential as subjects on higher ranks are more likely to be underconfident, while subjects on lower ranks are more likely to be overconfident, due to the limited scale for self-assessment. If subjects' actual ranks differed, principals might choose the underconfident agent not because they prefer underconfidence, but because underconfidence might signal a higher actual rank.

The results show that it can be advantageous to be underconfident with respect to the perception of others. In $SYMP$ principals reward the underconfident agent significantly more often than the overconfident agent. In $PERF$ principals bet on the underconfident agent significantly more often than on the overconfident agent. Questionnaire data reveals that underconfidence is preferred over overconfidence, and that the less self-confident agent is expected to exert more effort to improve himself, while the more self-confident agent is expected to rest on his high self-perception.

I also analyze whether the antipathy towards overconfidence is anticipated by eliciting the agents' (incentivized) beliefs of the principals' selection choices. Moreover, to test whether agents strategically bias their self-assessment in order to increase their selection chances, I conduct two control treatments without monetary incentives to be selected by the principal. Agents' beliefs in $PERF$ show that they do not expect underconfidence to signal a higher performance than overconfidence. Correspondingly, there is no difference in self-assessment between $PERF$ and its control treatment. In contrast, subjects anticipate that underconfidence is rewarded significantly more often than overconfidence, and men state marginally significantly lower ranks in $SYMP$ than in the non-strategic control treatment. Yet, there is no difference in self-assessment for women. One explanation could be that women do not downgrade their self-assessment strategically. Yet, I rather suggest that they even lower their self-assessment in the non-strategic setting, as its accuracy is still observable. Thus, they might still be afraid that their image might suffer when being overconfident.[4] Furthermore, women and men might downgrade their self-assessment in non-strategic settings due to an idea which goes back to Myerson (1991). He suggests that people internalize optimal behavior from certain situations and behave the same way in similar but different situations.[5] Thus, it might be the case that people have somehow imprinted the social norm of modesty and even downgrade their self-assessment in environments in which the (monetary) need for modesty is absent.[6]

---

[4]This is in line with women's shame of overestimation observed by the experimental study by Ludwig and Thoma (2012).

[5]Note that the accuracy of agents' self-assessment is still observable in the control treatments, but the agents do not have a monetary incentive to be chosen.

[6]Also compare Charness et al. (2012) who use this argumentation to explain overconfidence in non-strategic competitive settings.

There is an extensive and expanding literature on overconfidence. While one strand of this literature analyzes whether people are overconfident (e.g. Clark and Friesen, 2009; Hoelzl and Rustichini, 2005; Svenson, 1981), thereby focusing on the definition of overconfidence, the appropriate measurement, and influencing factors (see e.g. Benoît and Dubra, 2011; Moore and Healy, 2008), this paper is rather related to the strand of the overconfidence literature identifying potential consequences of a biased self-assessment. Thereof, many papers focus on non-payoff maximizing decisions caused by a biased self-assessment, e.g. overinvestment, value-destroying mergers of CEOs (Camerer and Lovallo, 1999; Malmendier and Tate, 2005 and 2008; Odean, 1999), and suboptimal selection of payment schemes (Dohmen and Falk, 2011; Niederle et al., 2013; Niederle and Vesterlund, 2007), or work environments (Niederle and Yestrumkas, 2008). Another strand of the literature, mainly theoretical work, identifies utility enhancing aspects of being overconfident, providing (behavioral) explanations for overconfidence at the same time. Overconfidence may directly enhance well-being (Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005; Caplin and Leahy, 2001; Koszegi, 2006), boost one's motivation and willpower (Bénabou and Tirole, 2002; Brocas and Carrillo, 2000), or increase performance (Compte and Postlewaite, 2004).

Only very few recent papers consider the impact of one's self-assessment on others, and whether individuals account for others' perception when stating their self-assessment. Ewers and Zimmermann (2012) theoretically and experimentally analyze whether individuals bias their self-assessment due to image concerns. They find that individuals state a higher self-assessment if reports are observed by an audience (anonymity is lifted) than in private.[7] Yet, they find that self-assessments do not differ if true performance is also publicly revealed, thus subjects do neither try to signal high ability nor modesty if the accuracy of one's self-assessment is observable. However, in contrast to my study, in Ewers and Zimmermann (2012) individuals have no monetary incentives for strategically biasing their self-assessment. Moreover, their study does not analyze how one's self-assessment is perceived by others, i.e. if one's social image or expected ability is actually increased by stating a higher self-assessment.

These issues are addressed in an experimental study by Charness et al. (2012). They investigate whether individuals bias their stated confidence about their performance strategically to deter or motivate others to enter a two-player tournament, and whether others react to it. They find that males inflate their stated confidence when deterrence is strategically optimal, and that men and women deflate their confidence if encouraging entry is strategically optimal. Moreover, they observe that individuals are less likely to enter

---

[7]Another study claiming that individuals inflate their self-assessment due to non-monetary image concerns is Burks et al. (2013). In a large survey with male truck drivers, they find a correlation between self-reports about how much one cares about one's image and overconfidence, consequently claiming that "overconfidence is social signaling bias".

the competition, the higher the confidence of the other person is. In line with these results Reuben et al. (2012) observe that men inflate their self-assessment to be voted as the group leader, which turns out to be a successful strategy.[8] However, in both studies individuals' actual performance is unknown and might strongly differ. Expecting all individuals to exhibit the same bias in self-assessment, the ranking of subjects' self-assessment most likely corresponds to the ranking of subjects' true performance. Thus, it is not the bias in self-assessment, which reveals information about the actual performance, but the self-assessment per se. This is different in my experiment, in which both agents have the same actual (relative) performance, enabling me to investigate whether the bias in self-assessment serves as a performance signal or influences one's image. In many real-life situations, in which individuals have to assess their performance, e.g. in promotion interviews or wage negotiations, the true performance is somehow appraised or at least partly known. Thus, the accuracy of the self-assessment might also be evaluated. On the one hand it might influence whether an individual is liked or not, on the other hand it might serve as a performance signal.

To the best of my knowledge this study is the first, which experimentally tests whether a principal prefers an over- or an underconfident agent. Yet, theoretical studies exist, providing different predictions. Gervais and Goldstein (2007) suggest that skill and effort are complements, thus an overconfident agent makes a higher effort choice due to underestimating the cost of effort or overestimating his marginal productivity. Sautmann (2013) suggests that overconfident agents overestimate their expected payoff, thus receiving higher incentives with the same wage. In contrast, Santos-Pinto (2008) suggests that a positive self-image and effort are substitutes, as an overconfident agent thinks that he has to exert less effort for the same outcome than an underconfident agent.

This paper might also contribute to the literature observing gender differences in self-assessment.[9] I suggest that the observed antipathy towards overconfidence adds to the explanation of the gender difference in self-assessment as women seem to experience emotions, i.e. the negative attitude towards overconfidence, stronger than men (see e.g. Brody, 1997; Grossman and Wood, 1993). In addition, they might even be punished more harshly than men when being self-confident (Eagly, 1987; Rudman, 1998). My results could moreover provide an explanation for the findings of Ludwig and Thoma (2012) who observes that women have shame to overestimate themselves in public, while men have

---

[8]In an experimental study Montinari et al. (2012) observe that the ex ante low ability type is chosen more often, as he is expected to exert more effort when receiving a fixed wage. However, the reason is a higher expected reciprocity when hiring the low ability type, which is absent in my study, as agents do not receive a fixed wage and do not learn whether they've been chosen or not until the end of the experiment.

[9]See e.g. Balafoutas et al., 2012; Beyer, 1990; Beyer and Bowden, 1997; Charness et al., 2012; Datta Gupta et al., 2013; Dohmen and Falk, 2011; Möbius et al., 2012; Niederle et al., 2013; Niederle and Vesterlund, 2007.

not.

The rest of this paper is structured as follows. In the next section, I describe the experimental design and the two different treatments. In Section 3 I present the main experimental results, i.e. the selection behavior of the principals. In Section 4 I analyze whether agents anticipate principals' preferences and whether they strategically bias their self-assessment. I conclude in Section 5.

2   Experimental Design

The experiment consists of two parts and a questionnaire, with separate instructions for each part. In part 1 all participants conduct a real effort task (task 1), which is solving Raven's Advanced Progressive Matrices (APM), a measure of cognitive ability (Raven, 2000). For each matrix participants have to select one out of 8 symbols fitting the visual pattern of the matrix. An example of a matrix is given in Figure 1.

Figure 1: Example of a Raven Advanced Progressive Matrix
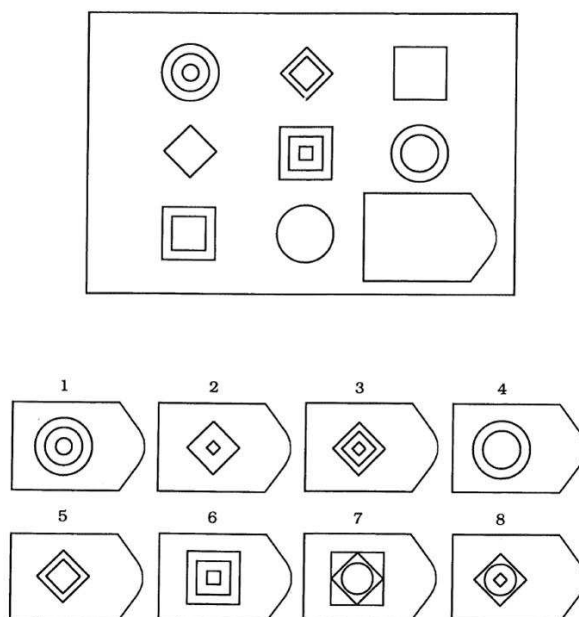
In this task ability and effort are needed to succeed in the task. The participants have five minutes to solve as many matrices as possible. After choosing a symbol, they receive feedback whether or not their chosen symbol is correct, and thereafter, the next matrix appears. Once having chosen a symbol, they cannot go back and correct it, neither is it possible to skip a matrix without making a choice. On subjects' screens the remaining time as well as the number of correctly and wrongly solved matrices is displayed. The maximum number of matrices is 22, none of the subjects managed to get to the last

matrix. Subjects are informed about their absolute performance, but neither about their relative performance nor the performance of others. They receive 5 tokens for each matrix they solve correctly. For each wrong answer 5 tokens are deducted from their earnings. Yet, they receive at least 0 tokens for part 1. During the whole experiment participants earn tokens, which are converted into Euros at the end of the experiment, at an exchange rate of 1 Euro for 10 tokens. Before the five minutes start, participants solve two matrices as a trial without payment. After the five minutes part 1 is finished and the instructions for part 2 are distributed. The instructions for both parts are read aloud.

At the beginning of part 2 subjects are randomly assigned a role. Out of the 24 participants in each session, the role A is assigned to 8 participants (principals) and the role B is assigned to 16 participants (agents). According to their performance in task 1, all 16 agents are ranked from 1-8, whereas each rank is assigned to two agents. The best and second best agent receive rank 1, the third and fourth best receive rank 2 and so on. The worst and second worst agent receive rank 8.[10] The two agents having the same rank, are merged to a pair, i.e. there are 8 pairs in each session. Each pair is randomly assigned to one of the 8 principals.

I conduct two treatments. In each treatment both agents estimate their rank between 1 and 8 and the principal selects one of the two agents. The only information the principal receives when making his choice, is the deviation of the agents' self-assessment, i.e. whether an agent under- or overestimated himself and to what extent. What differs between the two treatments is the incentive of the principal whom to pick.

Agents receive 20 tokens if their estimated rank corresponds with their actual rank. The payment for the accuracy of the guessed rank was not announced in the instructions, but only on the screen of the agents. Hereby, I exclude that inequity aversion affects the choice of the principals.[11]

As the scale for subjects' relative self-assessment is limited, it naturally occurs that subjects on a higher rank are more likely to underestimate their rank and subjects on a lower rank are more likely to overestimate their rank. If the actual ranks of the two agents differed, underconfidence might signal a higher actual rank than overconfidence. Thus, the principal's choice might be influenced by beliefs about the agents' actual ranks, mitigating the attitude towards over- or underconfidence. I exclude this effect by merging two subjects on the same rank. Thus, the sign and the magnitude of their deviation does not reveal a difference in their relative performance as there is none. The deviations of agents' self-assessments might only hint on a difference in the absolute performance, the agents' expected performance or self-confidence, which is what I am interested in.

---

[10]Two agents receive the same rank to have the same initial position for their self-assessment.

[11]Answers in the follow-up questionnaire show that principals nevertheless expected the agents to state their true belief about their relative ranks.

Treatment Sympathy ($SYMP$):

In this treatment each principal selects one of his two agents, who then receives 50 tokens. To exclude fairness concerns of the principal, it is neither possible to split the 50 tokens nor to avoid the decision. The only information the principal receives about the two agents is the deviation of their guessed ranks from their actual rank. Based on this information, he chooses whom to give the 50 tokens. Yet, I use the strategy vector method (SVM): Before the principal learns the actual deviations of the agents' guessed ranks, he takes the decision for 12 potential cases. Each case combines two different deviations of the agents' estimated ranks, including the sign of the deviation, i.e. if an agent over- or underestimates himself. For example case 1a is the following:

- one person overestimates his rank by 1 rank
- one person underestimates his rank by 1 rank

If one agent over- and the other agent underestimates himself each by 1 rank, the decision the principal takes for this case becomes relevant. The principal takes the decision for 12 different cases, which are listed in Table 1, where a negative deviation means that the agent underestimates himself and a positive deviation means that the agent overestimates himself. As accuracy might matter for the principal's choices, using the SVM decreases the number of sessions needed and allows for a cleaner data analysis, as each principal takes the decisions for the same cases. I picked these 12 cases in order to include the most realistic outcomes and to check the robustness of principals' preferences.[12]

Table 1: The Cases for which Principals Take a Decision

Case	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
Deviation agent 1	1	2	2	-2	3	-3	1	-1	2	-2	2	-2
Deviation agent 2	-1	-2	-1	1	-1	1	0	0	0	0	1	-1

The deviation is calculated as "actual rank - guessed rank", i.e. a positive deviation corresponds to overconfidence, a negative deviation to underconfidence. If the absolute deviations of the two agents differ (cases 2a-6b), the agent with the larger absolute deviation is listed on top.

In cases 1a and 1b both agents have the same absolute deviation. In cases 2a - 6b agents' accuracies differ and the less accurate agent is listed on top. In the a-cases the more self-confident agent has a larger absolute deviation, in the b-cases he has a smaller absolute deviation. For every case the principal selects one of the two agents, i.e. the one who he wants to receive the 50 tokens. His choice becomes relevant for the case, which actually applies to the two agents assigned to him. If none of the cases 1-12 applies to the two agents, principals take a 13th decision: Principals choose whether the agent who estimates the higher (better) rank or the agent who estimates the lower (worse) rank

---

[12]On subjects' screens cases were listed in a different order and I varied whether the more or less self-confident agent was listed first, what actually did not lead to different results.

shall receive the 50 tokens. The accuracy of the agents' estimated ranks is not considered in this decision.[13] If the two agents estimate the same rank, and therefore have the same deviation, a case not covered by the decisions of the principal, chance determines which agent receives the 50 tokens.[14]

After the 13 decisions, the principal learns the actual deviations of the two agents and thus which case becomes relevant. He neither learns the actual nor the estimated ranks of the agents. The agents learn their actual rank at the end of the experiment, but do not learn the guessed rank of the other agent. They are informed about whether the principal selected them or not, or whether the selection happened by chance.

Every principal receives 50 tokens for part 2 independent of his decision.

Treatment Performance ($PERF$):
In $PERF$ the principal takes the same 13 decisions as in $SYMP$, i.e. using the strategy vector method, the principal chooses one out of two agents, based on the deviations of their self-assessments. However, the motivation for the choice of the principal is different. While in $SYMP$ the principal's choice does not influence his monetary payoff, in $PERF$ it does. To maximize his expected earnings, the principal should choose the agent with the higher expected performance in a repetition of task 1, which is called task 2.

After the principals' decisions, all agents perform the same task as in part 1, with different matrices. Every agent solves the matrices for himself, but the payment scheme is competitive, with the two agents competing against each other. The agent, who achieves the higher difference of correctly minus wrongly solved matrices, receives 50 tokens. In case of a tie, the agent having solved more matrices correctly, receives the 50 tokens. If this number is also equal, chance decides.

The principals do not participate in task 2. They bet on the agent chosen in cases 1-13. If the agent, who the principal bets on, wins the competition, the principal receives 50 tokens. If both agents estimate the same rank, chance determines on whom the principal bets. After his decisions the principal learns about the actual deviations of both agents and which case is relevant. As in $SYMP$, the principal does neither learn the actual nor the estimated ranks of the agents. An agent receives 50 tokens if the principal bets on him. This monetary incentive is introduced to analyze whether agents bias their self-assessment in order to be selected. The analysis is shown in section 4. The payment is only announced on the agents' screens, thus principals do not know that an agent receives money when they bet on him, in order to avoid that feelings of sympathy and inequity aversion influence principals' choices. An agent learns whether the principal bet on him, but only after task 2. This is common knowledge. An agent does not learn the estimated

---

[13]This case actually became relevant in 20 out of 56 cases.
[14]This case became relevant in 8 out of 56 cases.

rank of the other agent, only when the two estimated the same rank, they learn that a chance move decides on whom the principal bets.

In both treatments all subjects are asked additional incentivized questions. In particular they estimate the mean deviation of agents and the choice behavior of principals. After the announcement of the payoffs, subjects complete a questionnaire asking for their age, gender, subject of study, choice motivation, and a self-assessment of risk preferences (Dohmen et al., 2011). Figure 2 illustrates the course of the treatment *PERF*. *SYMP* has the same course except that there is no task 2.

Figure 2: The Timeline of *PERF*
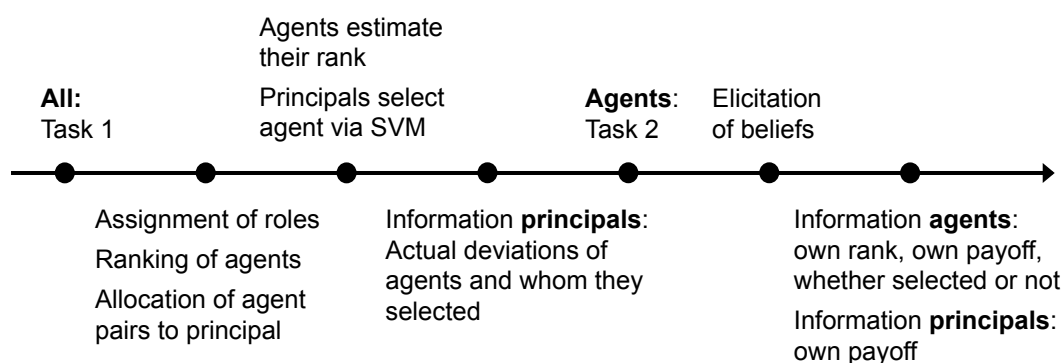
Experimental Procedure

I conducted the computerized experiment in the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) at the University of Munich during spring 2012. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007) and participants were recruited via ORSEE (Greiner, 2004). In total 216 subjects participated in the experiment (mainly students from the universities in Munich). I ran 5 sessions of *SYMP* and 4 sessions of *PERF*, whereof in each treatment 2 sessions were control sessions for agents, on which I will comment in Section 4.

Subjects were randomly assigned to sessions and could take part in one session only. Each session had 24 subjects and lasted a little less than one hour. Subjects earned 12.15 Euros on average (including a show-up fee of 4 Euros).

3   Experimental Results

3.1   Treatment *SYMP*

Table 2 reports the shares of principals choosing either the one or the other agent in each of the 13 different cases in *SYMP*. There are 40 principals in total. In 2 sessions (16 principals), which were control treatments for the agents, the principals' choices were only hypothetical and had no consequence on the decision, which agent received the 50 tokens. This decision was taken by chance to analyze whether agents' self-assessment is influenced by the principals' choices. The analysis is provided in the next section. The principals' decisions in these 2 sessions are not significantly different from the principals' choices in the other 3 sessions with actual choices, thus I pool the data.[15]

Table 2: Principals' Selection Behavior in *SYMP*.

| Case	1a	1b	2a	2b	3a	3b	
Deviation of agent 1	1	2	2	-2	3	-3	
Deviation of agent 2	-1	-2	-1	1	-1	1	
Share selecting agent 1	35%	12.5%	5%	32.5%	2.5%	27.5%	
Share selecting agent 2	65%	87.5%	95%	67.5%	97.5%	72.5%	
Binomial test (p-value)	0.040	0.000	0.000	0.019	0.000	0.003	
McNemar test (p-value)	0.004		0.007		0.002		
Case	4a	4b	5a	5b	6a	6b	7
Deviation of agent 1	1	-1	2	-2	2	-2	high
Deviation of agent 2	0	0	0	0	1	-1	low
Share selecting agent 1	10%	5%	2.5%	7.5%	10%	10%	32.5%
Share selecting agent 2	90%	95%	97.5%	92.5%	90%	90%	67.5%
Binomial test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.019
McNemar test (p-value)	0.688		0.500		1.000		

Every principal took every decision 1a-7. # of observations is 40 (including 16 hypothetical choices). A positive deviation represents "overconfidence", whereas a negative deviation represents "underconfidence". The Binomial test tests for a difference between selection rates and a 50:50 split for each case. The McNemar test tests for a difference between cases a and b.

The first important observation is that principals favor agents who are underconfident rather than overconfident. In case 1a 65% of principals (26 out of 40) reward the agent who underestimates his rank by 1 rank, whereas only 35% of principals (14 out of 40) reward the agent who overestimates his rank by 1 rank. This distribution is significantly different from a 50:50 split (one-sided binomial test, p=0.040). The result is even stronger in case 1b, in which 87.5% of principals (35 out of 40) reward the agent who underesti-

---

[15]Results go in the same direction when excluding the principals with hypothetical choices, but are less significant due to fewer observations. The separated data is listed in the appendix.

mates his rank by 2 ranks and only 12.5% of principals (5 out of 40) reward the agent who overestimates his rank by 2 ranks (one-sided binomial test, p=0.000). In case 7 in which the absolute deviation is unknown, 67.5% of principals reward the agent estimating the lower rank (*low_guess*) and only 32.5% reward the agent estimating the higher rank (*high_guess*) (one-sided binomialtest, p=0.019). These results are illustrated in Figure 3.

Figure 3: Principals' Choices for Cases 1a, 1b, and 7 in *SYMP*
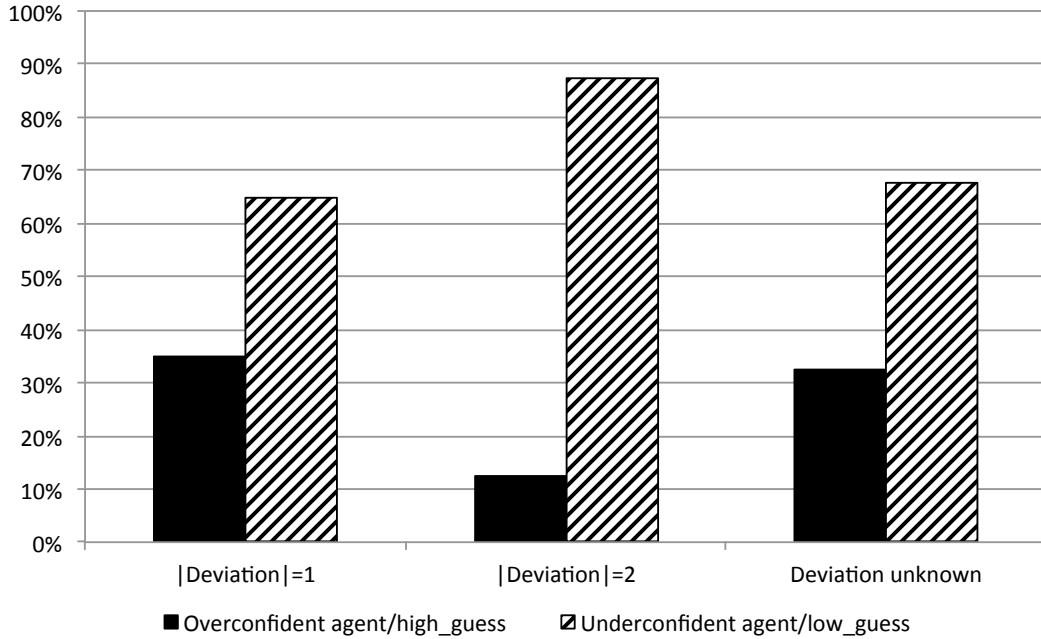
If agents' deviations differ (cases 2a-6b), the agent with the smaller absolute deviation is selected significantly more often - in every case - as shown by the results in Table 2. Figure 4 illustrates principals' choice behavior, whereas for each case the right bar represents the share of principals selecting the more accurate agent.

Despite the preference for accuracy, I observe that underconfidence is preferred over overconfidence. In 2a (3a) only 5% (2.5%) of principals choose the less accurate, *overconfident* agent, whereas in 2b (3b) 32.5% (27.5%) choose the less accurate, *underconfident* agent. The differences between cases 2a vs. 2b and 3a vs. 3b are significant (McNemar p=0.007 and p=0.002). Note that I do not observe a difference in selection behavior between cases 4a vs. 4b and 5a vs. 5b, meaning that the less accurate *underconfident* agent is not significantly more often selected than the less accurate *overconfident* agent if the more accurate agent is estimating his rank correctly (instead of being over- or underconfident). Thus, the higher selection rate of the less accurate agent in cases 2b and 3b, in comparison to 2a and 3b, seem to be driven by an antipathy towards overconfidence and not by a preference for underconfidence over accuracy.

Figure 4: Principals' Choices for Cases 2 - 6 in *SYMP*
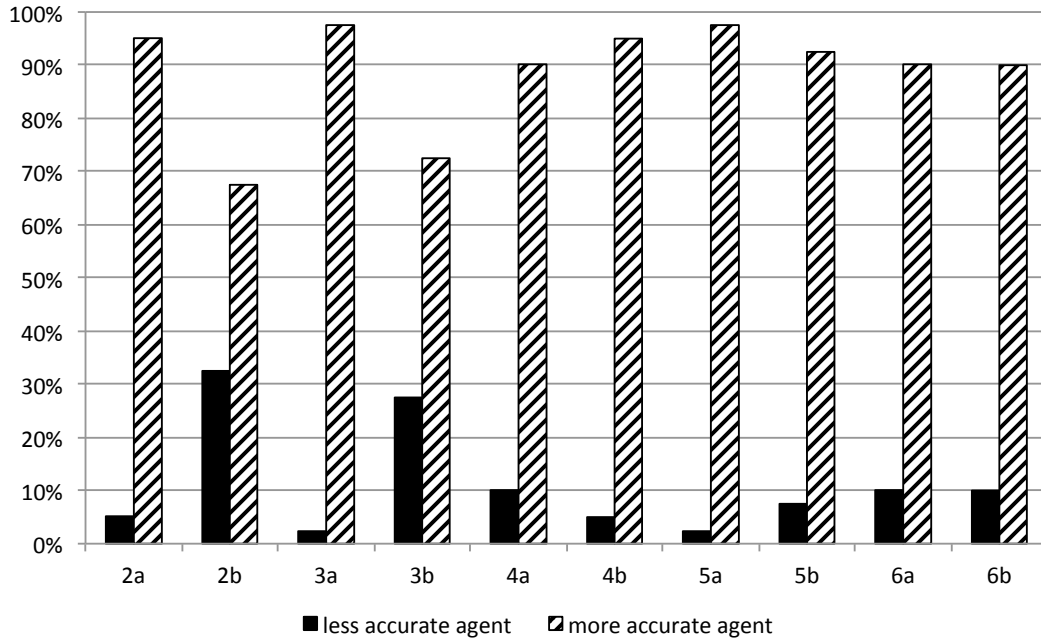
These results are confirmed by probit regressions, which are reported in Table 3.[16] The dependent variable is 1 if the agent with the larger absolute deviation is selected. *inacc_low* is a dummy for the cases, in which the agent with the larger absolute deviation (the more inaccurate agent) is underconfident, thus stating the lower rank (cases 2b, 3b, 4b, 5b, and 6b). Column (1) includes the principals' choices for the cases 2 and 3, i.e. all cases in which agents have a different absolute deviation and one agent is under- and the other agent is overconfident. Column (2) includes principals' choices for the cases 4-6, i.e. the cases in which either one agent has a deviation of 0 (cases 4-5), or both agents are either under- or overconfident (cases 6a and 6b). In both regressions I include control variables for risk and gender and cluster on principals.[17]

The coefficient of *inacc_low* in column (1) is highly significant, while in column (2) it is not. This shows that the less accurate agent is more likely to be selected if he is under- and the other agent is overconfident.

Considering the main focus of this study, i.e. the different perception of over- and underconfidence, the main result of *SYMP* is the following:

---

[16]The results are robust to OLS regressions.

[17]To elicit risk preferences, individuals indicated on a scale ranging from 0 to 10 whether they are willing to take risks (or try to avoid risks). 0 represented a very weak willingness to take risks, while 10 represented a strong willingness to take risks. Dohmen et al. (2011) show that this general risk question is a good predictor of actual risk-taking behavior.

Table 3: Probit of Selection of the Less Accurate Agent in *SYMP*

	(1) Cases 2-3	(2) Cases 4-6
	Coefficient (p-value)	Coefficient (p-value)
inacc_low	1.233	0.003
	(0.000)	(0.990)
Female	0.015	-0.235
	(0.960)	(0.508)
Risk	-0.110	-0.044
	(0.121)	(0.509)

I cluster standard errors on principals. Number of observations is 160 (1) and 240 in (2). In cases 2-3 one agent is overconfident, the other agent is underconfident. In cases 4-5 one agent has a deviation of 0, the other agent is under or overconfident; in cases 6 both agents' deviations have the same sign.

Result 1 *If controlling for accuracy, in treatment SYMP agents who are underconfident are rewarded significantly more often than agents who are overconfident.*

A modest agent who underestimates his performance is more likely to be rewarded by the principal than an overconfident agent, who believes that his performance has been better than it actually was. In the questionnaire answered by subjects at the end of the experiment, a reasoning for this behavior is revealed: most subjects prefer modest persons to self-confident persons. Some subjects even stated explicitly that they dislike people who are overconfident.

3.2 Treatment *PERF*

The principals' choice behavior for all 13 decisions in *PERF* is reported in Table 4.

I observe that the underconfident agent is selected more often than the overconfident agent when the agents' absolute deviations are equal or unknown. In case 1a 62.5% of principals (20 out of 32) bet on the agent who underestimates himself by 1 rank, while only 37.5% (12 out of 32) bet on the agent who overestimates himself by 1 rank. This result is not significant, but close to marginally significant (one-sided binomial-test, $p = 0.108$). Yet, in case 1b 75% of principals (24 out of 32) bet on the agent underestimating himself by 2 ranks, while only 25% (8 out of 32) bet on the agent overestimating himself by 2 ranks. This is significantly different from a 50:50 split (one-sided binomial-test, $p = 0.004$). In case 7 62.5% select the agent estimating the lower rank and 37.5% select the agent estimating the higher rank ($p = 0.108$). The selection rates for cases 1a, 1b, and 7 are illustrated in Figure 5.

The principals' choice rates in *PERF* for the cases 2-6 are illustrated in Figure 6. Almost

Table 4: Principals' Selection Behavior in *PERF*

Case	1a	1b	2a	2b	3a	3b	
Deviation agent 1	1	2	2	-2	3	-3	
Deviation agent 2	-1	-2	-1	1	-1	1	
Share selecting agent 1	37.5%	25%	12.5%	53.1%	15.6%	46.9%	
Share selecting agent 2	62.5%	75%	87.5%	46.9%	84.4%	53.1%	
Binomial test (p-value)	0.108	0.004	0.000	0.430	0.000	0.430	
McNemar test (p-value)	0.424		0.004		0.002		

Case	4a	4b	5a	5b	6a	6b	7
Deviation agent 1	1	-1	2	-2	2	-2	high
Deviation agent 2	0	0	0	0	1	-1	low
Share selecting agent 1	28.1%	53.1%	15.6%	40.6%	21.9%	46.9%	37.5%
Share selecting agent 2	71.9%	46.9%	84.4%	59.4%	78.1%	53.1%	62.5%
Binomial test (p-value)	0.010	0.430	0.000	0.189	0.001	0.430	0.108
McNemar test (p-value)	0.057		0.057		0.021		

Every principal took all 13 decisions. # of observations is 32. A positive deviation represents "overconfidence", whereas a negative deviation represents "underconfidence". The Binomial test tests for a difference between selection rates and a 50:50 split for each case. The McNemar test tests for a difference between cases a and b.
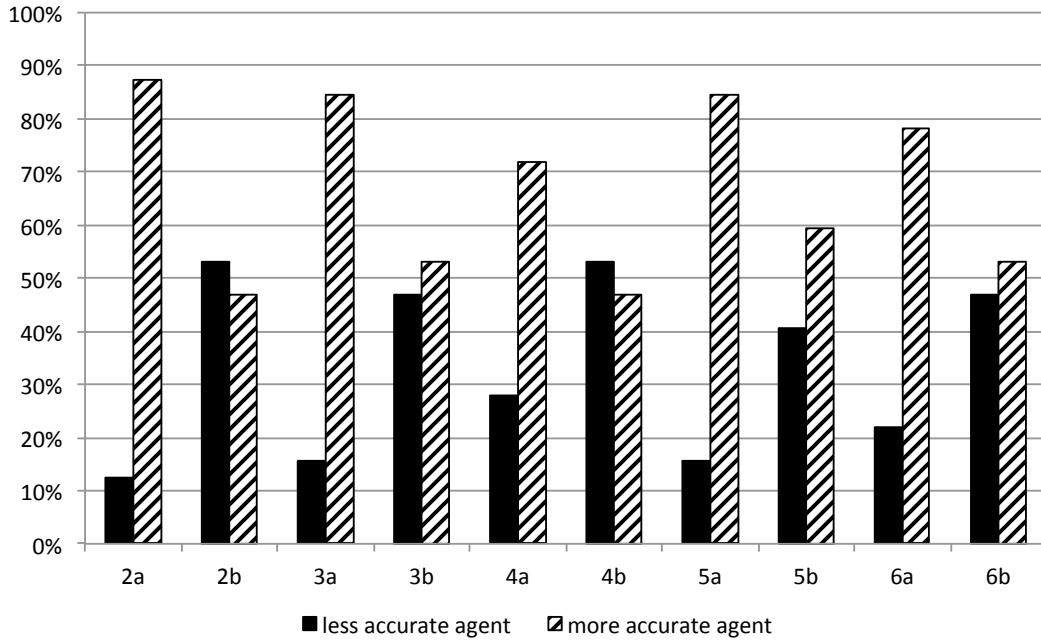
Figure 5: Principals' Choices for Cases 1a, 1b, and 7 in *PERF*
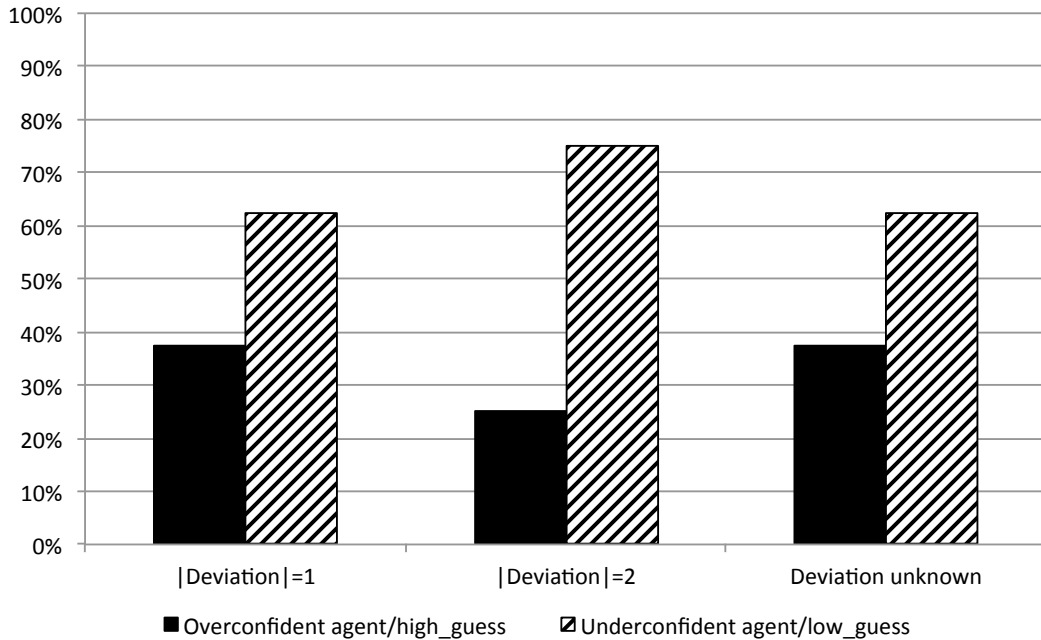
all principals select the agent stating the *lower* rank if he has the smaller absolute deviation (cases 2a, 3a, 4a, 5a, 6a). Yet, there is not a single case, in which the agent stating the *higher* rank is selected significantly more often than the agent estimating

Figure 6: Principals' Choices for Cases 2 - 6 in *PERF*

the lower rank, even if he has the smaller absolute deviation (cases 3b, 4b, 5b, 6b). In these cases selection rates are not significantly different from a 50:50 split as confirmed by binomial tests reported in Table 4. Note that even for cases 4b and 5b, in which one agent correctly assesses himself and the other agent underestimates himself, roughly 50% of principals select the underconfident agent. This shows that principals, betting on the underconfident agent in cases 2b and 3b, do not only choose the underconfident agent to avoid betting on the overconfident agent (as it happens in *SYMP*). Instead, they seem to be convinced that the underconfident agent will perform better in task 2. This also becomes apparent when conducting the same two probit regressions as in *SYMP*. The results are reported in Table 5.[18]

As above, the dependent variable is 1 if a principal selects the less accurate agent. The results in column (1) show that the less accurate agent is more likely to be selected if he is underconfident and the more accurate agent is overconfident than if he is overconfident and the more accurate agent is underconfident. However, in contrast to *SYMP* I observe that the coefficient of *inacc_low* in column (2) is also highly significant. This means that even if the more accurate agent is assessing himself correctly (or the deviation of his self-assessment goes in the same direction), the less accurate agent is more likely to be selected if he is underconfident than if he is overconfident.

In addition, the within-data analysis of principals shows that roughly one third of the

---

[18]The results are robust to OLS regressions.

Table 5: Probit of Selection of the Less Accurate Agent in *PERF*

	(1) Cases 2-3	(2) Cases 4-6
	Coefficient (p-value)	Coefficient (p-value)
inacc_low	1.080	0.700
	(0.000)	(0.001)
Female	-0.093	0.150
	(0.785)	(0.633)
Risk	0.015	0.060
	(0.827)	(0.337)

I cluster standard errors on principals. Number of observations is 128 (1) and 192 in (2). In cases 2-3 one agent is overconfident, the other agent is underconfident. In cases 4-5 one agent has a deviation of 0, the other agent is under or overconfident; in cases 6 both agents' deviations have the same sign. *inacc_low* is 1 for the b cases in which the agent stating the lower rank has a larger absolute deviation.

principals always selects the agent stating the lower rank, independent of the agents' absolute deviations. Another third of principals always selects the more accurate agent. Very few principals always select the agent stating the higher rank, and some principals seem to select randomly. Thus, in the cases in which selection shares are equal, the selection does not occur randomly, but most principals have a preference either for accuracy or a low self-assessment. This is also confirmed by answers given in the questionnaire. Thus, the main result of *PERF* is the following.

Result 2 *In PERF underconfidence seems to be perceived as a stronger signal for future performance than overconfidence.*

I observe that both agents are equally likely to win the competition independent of their self-assessment. Yet, only very few principals seem to chose randomly, but rather seem to have a clear preference. In the questionnaire, I asked the subjects to state reasons for their choice. The answer given most often was that they expected the underconfident agent trying harder to improve and thus exerting more effort in task 2 than the overconfident agent. The few principals choosing the overconfident agent stated that they expected self-confidence to enhance performance. The reason for selecting the accurate agent was that an agent being able to estimate his performance correctly was expected to have a high overall level of performance.

3.3 Treatment Comparison

Both treatments have the same main result: underconfidence is preferred over overconfidence. A difference between the two treatments is that in *PERF*, in which principals

select the agent who they expect to have a higher future performance, accuracy seems to be less important. In *PERF* the agent having the larger absolute deviation is significantly more often selected than in *SYMP*, especially if he stated the lower rank (one-sided test of proportions, case 2b: p=0.039, 3b: p=0.045, 4b: 0.000, 5b: p=0.000, 6b: p=0.082). Interestingly, this also holds true for the agent stating the higher rank (one-sided test of proportions, case 2a: p=0.126, 3a: p=0.023, 4a: 0.010, 5a: p=0.000, 6a: p=0.082). I conclude that in *PERF* more principals value a low self-assessment over an accurate self-assessment, and some also value a high and confident self-assessment, which is not the case in *SYMP*. Note that principals' expectations about agents' self-assessments do not differ across treatments. Principals expect an average deviation of 0.4 in both treatments (MWU, two-sided, p=0.891).[19]

These treatment differences are also confirmed by two probit regressions, which are reported in Table 6.[20] As in the earlier probit regressions, the dependent variable is 1 if the principal selects the less accurate agent, and it is 0 if he selects the more accurate agent. I conduct two separate regressions for the cases 2-3 and the cases 4-6. *inacc_low* is a dummy, being 1 for the cases, in which the less accurate agent is the agent stating the lower rank, *PERF* is a treatment dummy, and *PERF*inacc_low* the interaction of the two, in order to check whether the treatment difference is even larger for the agent stating the lower rank. In both regressions I cluster standard errors on principals.

Table 6: Probit of Selection of the Less Accurate Agent

	Cases 2-3 Coefficient (p-value)	Cases 4-6 Coefficient (p-value)
inacc_low	1.245	0.004
	(0.000)	(0.985)
PERF	0.718	0.657
	(0.033)	(0.025)
*PERF*inacc_low*	-0.160	0.694
	(0.709)	(0.026)
Female	-0.085	0.011
	(0.700)	(0.867)
Risk	-0.047	0.011
	(0.330)	(0.807)

Includes decisions of all principals, number of observations is 288 in column 1 and 432 in column 2, SE are clustered on principals.

---

[19]This also indicates that principals do not expect agents to downgrade their self-assessment. Moreover, principals were asked in the questionnaire if they think that agents stated the rank they actually think they have. The answer given was "yes" with almost no exception although principals did not know that agents' guess was incentivized.

[20]The results are robust to OLS regressions.

I observe that the coefficient of the treatment dummy *PERF* is significantly different from zero and positive in both regressions. Thus, the probability to be selected when being the less accurate agent is significantly higher in *PERF* than in *SYMP*. This result does not only hold true for the agent stating the lower rank (*inacc_low*), but also for the agent stating the higher rank. However, for the cases 4-6 the probability when being the less accurate agent and underconfident is even higher, as the coefficient of the interaction term *PERF*inacc_low* is positive and significant. Thus, the main treatment difference is the following:

Result 3 *In SYMP principals are more concerned about the accuracy of the agents' self-evaluation than in PERF. In PERF the less accurate agent is selected significantly more often than in SYMP, especially if he is underconfident.*

4 Further Analysis

The results in the former section show that it might have negative consequences when being overconfident. This might explain why individuals are modest in their self-assessment. In this section I first analyze whether agents anticipate the preference for underconfidence, and second, whether they (consciously) downgrade their self-assessment.

4.1 Anticipation of Principals' Preferences

After the principals' selection, agents estimated principals' selection behavior for the cases listed in Table 7. For each case they estimated how many of the 8 principals in their session selected the agent with the deviation listed first. One question was randomly chosen for payment and participants received 1 Euro if their answer did not differ more than +/-0.5 from the correct answer. The agents' average estimations are reported in Table 7.

In *SYMP* agents anticipate a stronger sympathy towards underconfidence than towards overconfidence. In *PERF* agents do not seem to expect the strong preference for underconfident agents. Instead agents seem to believe that overconfidence signals high future performance to principals, even though it does not. In other words, agents think that principals select the overconfident agent significantly more often in *PERF* than in *SYMP*, which is actually not true (case 1a and 2a). However, they do anticipate correctly that the less accurate agent is selected significantly more often in *PERF* than in *SYMP*. There are no gender differences in answers.

Table 7: Agents' Average Estimation of Principals' Choice Behavior

	# obs.	Case 1a (+1, -1)	Case 2a (+2, -1)	Case 2b (-2, +1)	Case 4a (1, 0)	Case 4b (-1, 0)
SYMP	48	2.33	1.40	4.44	1.19	2.06
PERF	32	4.0	3.19	3.79	3.34	3.62
p-value (MWU)		0.000	0.000	0.038	0.000	0.002

Including all agents, excluding the control sessions, in which principals made hypothetical choices. The numbers are the average estimation of the agents to the question, how many of the 8 principals chose the agent with the deviation listed first. E.g. the average agent thinks that on average 2.33 principals in *SYMP* picked the overconfident agent (deviation +1) in case 1a. The last row reports p-values of MWU-tests of treatment differences.

4.2   Strategic Adaptation of Self-Assessment

As agents only seem to anticipate principals' preference for underconfidence in *SYMP*, but not in *PERF*, agents' self-assessment should be lower in *SYMP* than in *PERF*. This actually holds true as reported further below. Yet, other factors might also trigger a difference in stated ranks, because the two settings are different. Therefore, I conduct two control treatments.

In *SYMP-CON*, which is the control treatment for *SYMP*, the agent receiving the 50 tokens is not selected by the principal, but by chance. To keep as much equal to *SYMP* as possible, each pair of agents is assigned to a principal and the principal learns the actual deviations of the two agents. Thus, a treatment difference in self-assessment cannot be caused by social image concerns or shame of overestimation (Ludwig and Thoma, 2012), but only by the (absent) incentive to be selected by the principal. I ask the principals to make the same (hypothetical) 13 decisions as in *SYMP*, but without any monetary consequences for any participant. Agents do not know that the principals make these hypothetical choices.

In *PERF-CON*, which is the control treatment for *PERF*, agents do not receive any tokens if the principal bets on them, i.e. they do not have a monetary incentive to be selected by the principal.[21]

To control for potential differences in performance across treatments, I calculate the accuracy of agents' stated ranks instead of only comparing agents' self-assessment. As agents' ranks are determined endogenously within a session, agents' actual ranks might be influenced by performance differences across sessions. Therefore, I use an agent's optimal guessed rank instead of the actual rank to calculate his deviation. The optimal guessed

---

[21]Note that there is no difference between *PERF* and *PERF-CON* for principals as they are not informed about agents monetary incentive in *PERF*. I pool *PERF* and *PERF-CON* in the section analyzing principals' behavior.

rank is the rank that is most likely assigned to an agent, given his performance and the performance distribution of all agents in all treatments.[22] I calculate the deviation of an agent as his optimal guessed rank minus his actual rank. Thus, a negative deviation represents underconfidence, a positive deviation represents overconfidence. Table 8 gives an overview of agents' average guessed and optimal guessed ranks as well as their average deviation in all treatments.

Table 8: Guessed and Optimal Guessed Ranks of Agents

	# obs.	Guessed Rank	Optimal guessed Rank	Deviation
SYMP	48	4.08	3.96	-0.13
SYMP-CON	32	4.22	4.47	0.25
PERF	32	4.47	5.28	0.81
PERF-CON	32	4.00	4.19	0.19

An agent's deviation is calculated as optimal guessed rank - guessed rank.

While agents are slightly underconfident in *SYMP* (average deviation -0.13), they are slightly overconfident in the other three treatments.[23] The deviation across *SYMP* and *PERF* is significantly different (MWU, two-sided, p=0.021). I cannot differentiate whether agents state a higher rank in *PERF* or a lower rank in *SYMP*.[24] Moreover, the differences between treatments and control treatments are not significant (MWU, two-sided, SYMP: p=0.562; *PERF*: p=0.236). As participants do not seem to anticipate the preference for underconfidence in *PERF* the absence of a difference in deviations between *PERF* and *PERF-CON* is not surprising. However, I expected a lower self-assessment in *SYMP* than in *SYMP-CON*. A possible explanation for why this is not the case is that in *SYMP-CON* as in *SYMP*, principals get to know whether an agent over- or underestimated himself. Ludwig and Thoma (2012) show that women have shame to overestimate themselves, while men have not. Women's shame might bias their self-assessment in the same direction as the ambition of being selected by the principal, leading to a low self-assessment in both treatments. As men seem to be less prone to this kind of shame, I conduct ordered probit regressions controlling for gender, in order to explore whether gender differences in stated ranks exist.

---

[22]To calculate the optimal guessed ranks, I ran Monte-Carlo simulations, in which I randomly drew 500,000 groups consisting of 15 participants out of the performance distribution of all agents (with replacement). I then calculated for any given performance level the rank within each simulated group. The optimal guessed rank equals the mode of all 500,000 simulated ranks.

[23]I cannot confirm the highly debated finding that subjects are overconfident. In this experiment, the subjects have a rather precise self-assessment. 22 % of agents estimate their rank correctly, 35 % have a deviation of +/-1 rank, 25 % have a deviation of +/-2 ranks and only 9% of subjects overestimate their rank by 3 or 4 ranks. In comparison also 9 % underestimate their rank by 3-5 ranks.

[24]Note that optimal guessed ranks in *PERF* are lower than in *SYMP*. As participants on lower ranks are more likely to overestimate themselves, the difference across treatments might not be due to a treatment difference, but only due to a performance difference and the limited scale. To take care of this issue I conduct Ordered Probit regressions below.

Table 9 reports the results of ordered probit regressions for each gender separately.[25] The dependent variable is the guessed rank. The independent variables are the optimal guessed rank (as performance measure), a dummy for *CON*, being one for *PERF-CON* and *SYMP-CON*, a dummy for *PERF*, being one for PERF and PERF-CON, and an interaction dummy for *PERF-CON*.[26] I also control for risk aversion.

Table 9: Ordered Probit of Guessed Rank

	(1) WOMEN	(2) MEN
Opt. guessed rank	0.288***	0.418***
PERF	-0.264	0.125
CON	0.055	-0.566*
PERF*CON	0.160	0.211
Risk	0.001	0.214
# of observations	78	66

The table reports the coefficients. Base case is *SYMP*.
The sample consists of all agents and the regression clusters on sessions.
*** significant on 1% level, ** 5% level, * 10% level

Column 1 reports the results for men, column 2 reports the results for women. I observe that for men the coefficient of *CON* is marginal significant (p=0.072) and negative, i.e. men rank themselves lower in *SYMP* than in *CON*. It seems as if men strategically lower their self-assessment in *SYMP* to increase their selection chances.

There is no significant difference in self-assessment for women across *SYMP* and *CON*. Note that there is no gender difference in self-assessment except in *SYMP-CON*, in which men rank themselves higher than women (MWU, two-sided, p=0.032).

Besides a conscious downgrade of self-assessment, it can be the case that the social antipathy towards overconfidence influences subjects' self-assessment unconsciously. Subjects' self-assessment might not only be influenced in situations, in which others learn their self-assessment, but per se. Charness et al. (2012) show that individuals might act out of unconscious strategic concerns, even in situations, in which strategic concerns are absent. They pick up the idea made by Myerson (1991) and further developed by Samuelson (2001) that people make the same decisions in situations that appear to be similar for the sake of convenience. This can lead to suboptimal behavior in certain situations. While Charness et al. (2012) argue that subjects might be overconfident due to an internal-

---

[25]The results are robust when conducting OLS regressions.

[26]I do not use three dummies for three treatments, but one for *PERF* and one for *CON*, and an interaction of the two for *PERF-CON*. Results are the same when including three dummies for *PERF*, *PERF-CON* and *SYMP-CON*, except that the coefficient of *PERF-CON* is marginally significant for men.

ization of the positive impact of overconfidence, I suggest that the opposite can be the case.

5 Conclusion

In this paper I analyze how overconfidence in comparison to underconfidence is perceived by others. The results reveal that underconfident agents are perceived as more likable than overconfident agents and are expected to exhibit a higher performance in a real effort task. Questionnaire answers suggest that modest agents are expected to be more ambitious to improve, while overconfident agents rather have the reputation to rest on their high self-confidence.

Elicited beliefs of agents show that they do not expect the principals to select the underconfident agent more often when performance is the critical selection criterion. However, they anticipate that underconfidence is deemed more likeable than overconfidence. The comparison of stated self-assessments to a treatment, in which the principal cannot make a selection choice (non-strategic setting) shows that men slightly deflate their self-assessment strategically to be rewarded by the principal. Yet, women do not. An explanation might be given by women's shame of overestimation (Ludwig and Thoma, 2012) suggesting that women even deflate their self-assessment in the non-strategic setting, as principals still learn the deviation of agents' self-assessment. Thus, apart from monetary consequences women might expect their social image to suffer. Moreover, besides the conscious adaptation of self-assessment, individuals might have internalized the negative attitude towards overconfidence and might be modest in situations, in which no strategic concerns are at place.

While further research is needed to precisely identify in which situations individuals might bias their self-assessment consciously or unconsciously, the results reported in this paper provide an important fact one should consider when eliciting and interpreting individuals' self-assessment: subjects might not state the self-assessment that they actually have, but that has the largest signaling value.

References

Akerlof, G. A., Dickens, W. T., 1982. Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics*, 97(4), 543–569.

Alicke, M. D., 1985. Global Self-Evaluation as Determined by the Desirability and Controllability of Trait Adjectives. *Journal of Personality and Social Psychology*, 49, 1621–1630.

Balafoutas, L., Sutter, M., 2012. Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Lab. *Science*, 335(6068), 579–582.

Bénabou, R., Tirole, J., 2002. Self Confidence and Personal Motivation. *Quarterly Journal of Economics*, 117(3), 871–915.

Benoît, J.-P., Dubra, J., 2011. Apparent Overconfidence. *Econometrica*, 79(5) , 1591–1625.

Beyer, S., 1990. Gender Differences in the Accuracy of Self-Evaluations of Performance. *Journal of Personality and Social Psychology*, 59, 960–970.

Beyer, S., Bowden, E. M., 1997. Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. *Personality and Social Psychology Bulletin*, 23, 157–172.

Brocas, I., Carrillo, J. D., 2000. The Value of Information when Preferences are Dynamically Inconsistent. *European Economic Review*, 44, 1104-1115.

Brody, L. R., 1997. Gender and Emotion: Beyond Stereotypes. *Journal of Social Issues*, 53, 369–394.

Brunnermeier, M. K., Parker, J. A., 2005. Optimal Expectations. *American Economic Review*, 95(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L., Rustichini, A., 2013. Overconfidence and Social Signalling. *Review of Economic Studies*, forthcoming.

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

Caplin, A., Leahy, J., 2001. Psychological Expected Utility Theory and Anticipatory Feelings. *The Quarterly Journal of Economics*, 116(1), 55–79.

Charness, G., Rustichini, A., Van de Ven, J., 2012. Self-Confidence and Strategic Behavior. Mimeo.

Clark, J., Friesen, L., 2009. Overconfidence in Forecasts of Own Performance: An Experimental Study. *Economic Journal*, 119(534), 229–251.

Compte, O., Postlewaite, A., 2004. Confidence-Enhanced Performance. *American Economic Review*, 94(5), 1536–1557.

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

Dohmen, T., Falk, A., 2011. Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences and Gender. *American Economic Review*, 101(2), 556–590.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G., 2011. Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association*, 9(3), 522–550.

Dunning, D., Meyerowitz, J. A., Holzberg, A. D., 1989. Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability. *Journal of Personality and Social Psychology*, 57(6), 1082–1090.

Eagly, A. H., 1987. *Sex Differences in Social Behavior: A Social-Role Interpretation.* Hillsdale, NJ: Erlbaum.

Ewers, M., Zimmermann, F., 2012. Image and Misreporting. IZA Discussion Paper No. 6425.

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

Gervais S., Goldstein, I., 2007. The Positive Effects of Biased Self-Perceptions in Firms. *Review of Finance*, 11(3), 453–496.

Greiner, B., 2004. An Online Recruitment System for Economic Experiments. *Forschung und wissenschaftliches Rechnen 2003,* GWDG Bericht 63, Eds. K. Kremer and V. Macho, Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, 79–93.

Grossman, M., Wood, W., 1993. Gender Differences in Intensity of Emotional Experience: A Social Role Interpretation. *Journal of Personality and Social Psychology*, 65, 1010–1022.

Hoelzl, E., Rustichini, A., 2005. Overconfident: Do You Put Your Money on It? *Economic Journal*, 115, 305–318.

Koszegi, B., 2006. Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association*, 4(4), 673–707.

Krueger, J. I., 1999. Lake Wobegon be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgments. *Journal of Personality and Social Psychology*, 77(2), 221–232.

Larrick, R. P., Burson, K. A., Soll, J. B., 2007. Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and when it Does not). *Organizational Behavior and Human Decision Processes*, 102, 76–94.

Ludwig, S., Thoma, C., 2012. Do Women Have More Shame than Men? An Experiment on Self-Assessment and the Shame of Overestimating Oneself. Munich Discussion Paper 2012-15.

Malmendier, U., Tate, G., 2005. CEO Overconfidence and Corporate Investment. *The Journal of Finance*, 60(6), 2661–2700.

Malmendier, U., Tate, G., 2008. Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction. *Journal of Financial Economics*, 89, 20–43.

Messick, D. M., Bloom, S., Boldizar, J. P., Samuelson, C. D., 1985. Why We Are Fairer than Others? *Journal of Experimental Social Psychology*, 21, 480–500.

Montinari N., Nicolo, A., Oexl, R., 2012. Mediocrity and Induced Reciprocity. Jena Economic Research Papers 2012-053.

Moore, D. A., Cain, D. M., 2007. Overconfidence and Underconfidence: When and Why People Underestimate (and Overestimate) the Competition. *Organizational Behavior and Human Decision Processes*, 103, 197–213.

Moore, D., Healy, P. J., 2008. The Trouble with Overconfidence. *Psychological Review*, 115(2), 502–517.

Möbius, M. M., Niederle, M., Niehaus, P., Rosenblat, T. S., 2012. Managing Self-Confidence: Theory and Experimental Evidence. Mimeo.

Myerson, R. (1991), *Game Theory: Analysis of Conflict*, Harvard University Press: Cambridge,MA.

Niederle, M., Segal, C., Vesterlund, L., 2013. How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1), 1–16.

Niederle, M., Vesterlund, L., 2007. Do Women Shy Away from Competition? Do Men Compete too Much? *Quarterly Journal of Economics*, 122 (3), 1067–1101.

Niederle, M., Yestrumskas, A. H., 2008. Gender Differences in Seeking Challenges: The Role of Institutions. National Bureau of Economic Research Working Paper 3922.

Odean, T., 1999. Do Investors Trade too Much? *American Economic Review*, 89, 1279–1298.

Raven, J. C., 2000. *Raven's Advanced Progressive Matrices* (APM). Pearson.

Reuben, E., Rey-Biel, P., Sapienza, P., Zingales, L., 2012. The Emergence of Male Leadership in Competitive Environments. *Journal of Economic Behavior and Organization*, 83(1), 111–117.

Rudman, L. A., 1998. Self-Promotion as a Risk-Factor for Women: The Costs and Benefits of Counterstereotypical Impression Management. *Journal of Personality and Social Psychology*, 77, 629–645.

Samuelson, L., 2001. Analogies, Adaptation, and Anomalies. *Journal of Economic Theory*, 97(2), 320–366.

Santos-Pinto, L., 2008. Positive Self-image and Incentives in Organisations. *Economic Journal*, 118(531), 1315–1332.

Sautmann, A., 2013. Contracts for Agents with Biased Beliefs: Some Theory and an Experiment. *American Economic Journal: Microeconomics*, forthcoming.

Svenson, O., 1981. Are We All Less Risky and More Skillful than our Fellow Drivers? *Acta Psychologica*, 47, 143–148.

6 Appendix A2

A2.1 Tables

Table A2.1: Principals' Choices in $SYMP$ Separated

			$SYMP$	$SYMP\text{-}CON$
Case 1a	Deviation of agent 1	1	62.5%	68.8%
	Deviation of agent 2	-1	37.5%	31.2%
Case 1b	Deviation of agent 1	2	91.7%	81.3%
	Deviation of agent 2	-2	8.3%	18.7%
Case 2a	Deviation of agent 1	2	95.8%	93.8%
	Deviation of agent 2	-1	4.2%	6.2%
Case 2b	Deviation of agent 1	-2	70.8%	62.5%
	Deviation of agent 2	1	29.2%	37.5%
Case 3a	Deviation of agent 1	3	95.8%	100%
	Deviation of agent 2	-1	4.2%	0%
Case 3b	Deviation of agent 1	-3	66.7%	81.2%
	Deviation of agent 2	1	33.3%	18.8%
Case 4a	Deviation of agent 1	1	91.7%	87.5%
	Deviation of agent 2	0	8.3%	12.5%
Case 4b	Deviation of agent 1	-1	95.8%	97.8%
	Deviation of agent 2	0	4.2%	2.2%
Case 5a	Deviation of agent 1	2	95.8%	100%
	Deviation of agent 2	0	4.2%	0%
Case 5b	Deviation of agent 1	-2	91.7%	93.8%
	Deviation of agent 2	0	8.3%	6.2%
Case 6a	Deviation of agent 1	2	91.7%	87.5%
	Deviation of agent 2	1	8.3%	12.5%
Case 6b	Deviation of agent 1	-2	91.7%	87.5%
	Deviation of agent 2	-1	8.3%	12.5%
Case 7	higher guess		33.3%	31.2%
	lower guess		66.7%	68.8%

Each principal took every decision 1a-7. # of observations is 24 in $SYMP$ and 16 in $SYMP\text{-}CON$.
A positive deviation represents "overconfidence", whereas a negative deviation represents
"underconfidence". $SYMP\text{-}CON$ is the control treatment of $SYMP$, in which principals make
hypothetical choices without any payoff consequences.

A2.2 Instructions (translated from German)

We welcome you to this experiment. Please read these instructions carefully. After the start of the experiment, please follow the instructions on your screen.

At the end of the experiment you will get paid according to your decisions and the decisions of the other participants as described below. In addition, you will get a fixed payment of 4 Euro for your attendance.

During the whole experiment you are not allowed to talk to other participants, to use mobile phones, or to start other programs on your computer. If you disobey these rules, we have to exclude you from the experiment and all payments. If you have any questions, please raise your hand. An experimenter will come to your seat to answer your questions.

During the experiment, we are not talking about Euros but tokens. Your payment will be calculated in tokens. At the end of the experiment your overall score will be converted to Euro, whereas

1 Token = 10 Eurocent.

The experiment consists of two parts and a questionnaire. Part 1 will be explained below. Once all participants have finished part 1, you will receive the instructions for part 2. After part 2 a questionnaire will follow.

Instructions Part 1 (equal for *SYMP* and *PERF*)

In Part 1 of the experiment all participants are requested to solve an assignment, which will be explained more precisely in the following:

On your screen a matrix, i.e. a rectangular arrangement of different symbols will appear in a framed box. The matrix has 3 columns and 3 rows. The symbol in the lower right corner is missing. There will be 8 symbols beneath the matrix, one of which fits schematically in the lower right corner of the matrix. An example is given below.

In this example the correct solution is symbol number 5.

Your task is to choose the right symbol. After you made your choice, a new matrix including 8 new symbols will appear. Again exactly one of the 8 symbols fits in the lower right corner of the matrix.

You have to choose a symbol for each matrix. You cannot move on on to the next matrix without making a choice. Moreover you will not be able to go back to the previous matrix and change your choice after you have confirmed it.

You have 5 minutes (300 seconds) to solve as many matrixes as possible. Your remaining time will be displayed in the upper right corner of your screen. On the left bottom of the
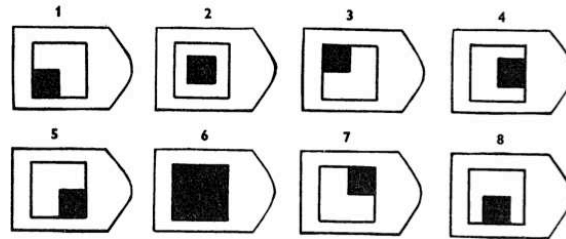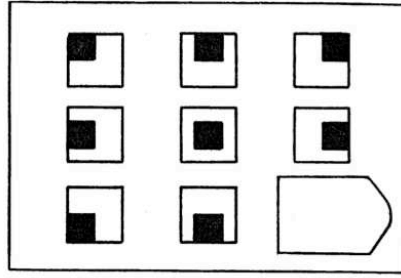
screen, you will see how many matrices you have solved correctly and incorrectly so far, and whether your last choice was correct or wrong. The difficulty of the matrices will increase as time elapses.

Your payment for Part 1
You will receive 5 tokens for each matrix you solve correctly. For each matrix you solve wrongly, 5 tokens will be deducted from your earnings. Thus, your overall score is:

5 x (number of correctly solved matrices - number of incorrectly solved matrices)

In the following, we call the difference *(number of correctly solved matrices - number of incorrectly solved matrices)* your performance. If your performance is negative, i.e. if you solve more matrices incorrectly than correctly, you will receive 0 tokens for part 1. Therefore, your payment for part 1 is:

5 x your performance if your performance > 0 or 0 if your performance < / = 0

The Course of Part 1
After every participant has read these instructions, we will tell you a password. After having entered the password, two trial matrices will successively appear on your screen, in order for you to get used to the type of matrices and to the selection of symbols. You won't receive a payment for solving the trial matrices. Afterwards, the 5 minutes will start as described above. After the 5 minutes part 1 is completed and the instructions for part 2 will be handed out.

At the beginning of part 2, role A and role B are assigned randomly to every participant. There are 24 participants in the room. Role A is randomly assigned to 8 participants ("A-persons") and role B is randomly assigned to 16 participants ("B-persons"). You will be informed about your role on your screen at the beginning of part 2.

Ranking

The **16 B-persons** will be ranked according to **their performance in part 1**. There will be **8 ranks in total**. Each rank will be allocated to **two B-persons**. The two B-persons having the highest and the second highest performance will be assigned rank 1, the two B-persons with the third and the fourth highest performance will be assigned rank 2, ..., the two B-persons with the lowest and the second lowest performance will be assigned rank 8.

Remember:

Performance= # of correctly solved matrices - # of incorrectly solved matrices

If several B-Persons have the same performance and cannot be assigned to the same rank, the B-Person, who solved more matrices correctly, gets the higher rank. If this number is also equal, ranks will be assigned by chance.

Thus, a B-Person will be assigned rank...
...1 if 0 or 1 of the other 15 B-persons performed better than him.
...2 if 2 or 3 of the other 15 B-persons performed better than him.
...3 if 4 or 5 of the other 15 B-persons performed better than him.
...4 if 6 or 7 of the other 15 B-persons performed better than him.
...5 if 8 or 9 of the other 15 B-persons performed better than him.
...6 if 10 or 11 of the other 15 B-persons performed better than him.
...7 if 12 or 13 of the other 15 B-persons performed better than him.
...8 if 14 or 15 of the other 15 B-persons performed better than him.

Self-assessment of the B-persons

Each of the 16 B-persons will estimate his rank between 1 and 8.
The B-persons will be informed about their actual rank at the end of the experiment.

Matching of B-pairs

The two B-persons who have achieved the same rank will be matched together to a so-called B-pair. Since there are 8 ranks, there will be 8 B-pairs in total.

[[**Task 2 of the B-persons**

After the estimation of their rank, all B-persons will conduct the same task as in part

---

[27]Passages only occurring in *SYMP* are indicated by [...], passages only occurring in *PERF* are indicated by [[...]].

1 with different matrices, called task 2. The difficulty of the matrices will be similar as in part 1. As in part 1, each B-person will solve the matrices by himself. However, the payment of a B-person for task 2 will now depend on his performance in comparison to the performance of the B-person, who he is matched with.

The payment scheme for task 2 is similar to a competition. The B-person, who will achieve the higher performance in Task 2 within a pair, will win 50 points.

If both B-persons achieve the same performance in Task 2, the B-person, who solved more matrices correctly, will win. If this number is also equal, chance will decide who of the two B-persons will win the competition, thus getting 50 points. The other B-person will receive 0 points for Task 2.]]

Assignment
Each A-person will be assigned one of the 8 B-pairs. This allocation will be random and anonymous. No participant will get to know the identity of the participants assigned to him during or after the experiment.

[Payment Part 2
Every A-person will receive 50 points. One of the B-persons within each B-pair will also receive 50 points. The A-person who is assigned to this pair will decide who of the two B-persons will receive the 50 points.]

[[A-persons' bet
The A-persons are not taking part in Task 2. Instead, they will bet on one of the two B-persons, who were assigned to them. If the B-person, who the A-person has bet on, wins, the A-person will receive 50 points.]]

A-persons' selection decisions
The A-person will decide, [who of the B-persons will get the 50 points] [[on whom of the two B-persons he will bet on]], in dependence of the B-persons' deviations of their estimated ranks from their actual rank.

The A-person will decide for 12 possible combinations of deviations of the estimated ranks of the B-persons from their actual rank, [who of the B-person will receive the 50 points] [[on whom of the B-persons he will bet on]] (case 1-12). In each case, the two B-persons are under- or overestimating themselves differently. For example, the cases 1, 2 and 12 are as follows:

Case 1:
- One person is overestimating himself by 1 rank
- One person is underestimating himself by 1 rank

Case 2:

- One person is overestimating himself by 2 ranks

- One person is underestimating himself by 2 ranks

...

Case 12:

- One person is under-/overestimating himself by 0 ranks

- One person is underestimating himself by 2 ranks

For each of the listed cases, the A-person will choose the person [who he wants to receive the 50 points] [[he wants to bet on]], in case this case will occur. Only one case can apply to the B-pair that is assigned to the A-person. If a case applies to the B-pair, the selection decision for this case will become relevant. [The B-person chosen by the A-person receives 50 points.] [[If the selected B-person wins the competition in Task 2, the A-person will receive 50 points.]]

Note: It is *not* the task of the A-person to decide which self-assessments of the B-persons are most likely to occur. If the selection decision, e.g. for case 1, becomes relevant, one of the two B-persons overestimated himself by 1 rank and the other person underestimated himself by 1 rank. The A-person [decides who of the two persons will receive 50 points] [[chooses the person who he will bet on in this case]].

Example: One of the two B-persons overestimated himself by 2 ranks and the other one underestimated himself by 2 ranks. Thus, the decision taken by the A-person in case 2 is relevant. [The B-person chosen by the A-person in case 2 gets the 50 points.] [[If the A-person selects the person, who overestimated himself by 2 ranks, and this person wins the competition in Task 2, the A-person will receive 50 points. If the A-person selects the person, who underestimated itself by 2 ranks, and the selected person wins, the A person will also receive 50 points. Otherwise the A-person does not get any points.]]

Since the A-persons will not learn which case is relevant until having taken the selection decisions, they should take the decision for every case, as if this case was relevant for the assigned pair. In the event that none of the cases 1-12 will apply to the B-pair, the A-person will also take a more general selection decision, called case 13.

CASE 13: The selection decision in this case will become relevant if none of the cases 1-12 applies to the B-pair. In this case the A-person decides, whether [the person who stated the higher (better) rank, or the person who stated the lower (worse) rank out of the two B-persons will get the 50 points] [[he wants to bet on the person, who stated the higher (better) rank, or on the person, who stated the lower (worse) rank.]]. Note: A higher (better) rank means that the B-person has stated a smaller number. The accuracy of the stated ranks is not considered in case 13.

In Case 13 the A-person has the following choice:

[[To choose]] [[To bet on]]

- the person who estimated the higher (better) rank
- the person who estimated the lower (worse) rank

Example: If one of the two B-persons overestimated himself by 3 ranks and the other person underestimated himself by 3 ranks, a case that doesn't correspond to any of the cases 1-12, then the decision of the A-person in case 13 will become relevant. Since both B-persons are assigned the same rank, the person who overestimated his rank by 3 ranks is the person who stated the higher rank. The person who underestimated his rank by 3 ranks is the person who estimated the lower rank.

Equal self-assessment of the B-persons:

If both B-persons estimate the same rank, none of the cases 1-13 is relevant, and chance will decide [who of the two B-persons will receive the 50 points] [[on who of the two B-persons the A-person will bet on]].

A-persons' information

After the selection decisions, but before the B-persons will begin task 2, the A-person will learn about how many ranks the two B-persons over- or underestimated themselves. Thus, they learn which case is relevant, or which B-person has been selected randomly respectively. The A-person will neither learn the actual nor the estimated ranks of the B-persons.

Summary and course of action of part 2

Based on the performance in part 1, the 16 B-persons will be ranked, whereas always two B-persons will receive the same rank.

The two B-persons being assigned the same rank will be matched to a B-pair. Each B-pair will be randomly assigned to an A-person.

Each B-person will estimate his rank between 1 and 8. The B-persons will be informed about their actual rank at the end of the experiment.

At the same time each A-person will decide for 13 different combinations of deviations of the B-persons' estimated ranks from their actual rank, [which agent will receive 50 points] [[on whom he wants to bet]].

After the decisions the A-persons will learn the B-persons' actual deviations and which of the cases 1-13 has occurred. If both B-persons estimated the same rank, chance will decide [who of the two agents will get the 50 points] [[on whom the A-Person will bet]].

[The B-person chosen in this case receives 50 points as does the A-person.]

34

[[Subsequently, all B-persons conduct task 2. The B-person achieving the higher performance will receive 50 points. If the A-person chose this B-person in the relevant case, the A-person will also receive 50 points.

After Task 2 the B-persons will learn their actual rank.]]

During the experiment, all participants will get the chance to earn additional points by answering additional questions.