

Under- versus overconfidence: an experiment on how others perceive a biased self-assessment

Carmen Thoma¹

Received: 19 November 2013 / Revised: 1 March 2015 / Accepted: 9 March 2015 /
Published online: 28 March 2015
© Economic Science Association 2015

Abstract This paper reports findings of a laboratory experiment, which explores how reported self-assessment regarding the own relative performance is perceived by others. In particular, I investigate whether overconfident or underconfident subjects are considered as more likeable, and who of the two is expected to win in a tournament, thereby controlling for performance. Underconfidence beats overconfidence in both respects. Underconfident subjects are rewarded significantly more often than overconfident subjects, and are significantly more often expected to win. 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-assessment · Overconfidence · Signaling · Experiment

Mathematics Subject Classification C91 · D03 · J16

Electronic supplementary material The online version of this article (doi:[10.1007/s10683-015-9435-2](https://doi.org/10.1007/s10683-015-9435-2)) contains supplementary material, which is available to authorized users.

✉ Carmen Thoma
carmen.thoma@gmail.com

¹ Ludwig-Maximilians-Universität (University of Munich), Munich, Germany

1 Introduction

Many earlier psychological and social–psychological studies claim that people are overconfident, which has sometimes been labelled the *better-than-average* effect.¹ However, recent economic studies analyzing choice behavior instead of verbal reports, show that individuals are able to estimate their skills correctly or even underestimate it.² A biased self-assessment can lead to systematic biases in individuals' decision making, thus the topic is of high interest to economists.³ Yet, barely any research has been done to analyze how overconfidence in comparison to underconfidence is perceived by others, and whether individuals adapt their reported 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 likeable than overconfident individuals. Secondly, I explore whether under- or overconfidence is perceived as a stronger signal for ambition and effort. As one's self-assessment might be a signal of one's actual performance, and therefore not the self-assessment but the estimated performance might be rewarded, I control for performance as follows: the individuals compared exhibit the same relative rank based on their performance in a task, conducted at the beginning of the experiment. My findings contribute to the expanding literature analyzing the advantages of overconfidence in comparison to underconfidence. Thereby I add to the questions why individuals might (rationally) exhibit a bias in their reported self-assessment and in which situations we should expect individuals to over- or underestimate themselves. How reported over- or underconfidence is perceived by others 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' reported self-assessment.

The experiment consists of two parts. In the first part all subjects perform an incentivized real effort task which serves as the basis for their self-assessment. In the second part 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

¹ Svenson (1981), e.g., shows that more than 50 % of participants (77 % in a Swedish and 88 % in a US study) report to drive more safely than the median, taking this result as evidence for overconfidence. Other studies demonstrating the *better-than-average* effect are Alicke (1985), Dunning et al. (1989) and Messick et al. (1985).

² In laboratory experiments Hoelzl and Rustichini (2005) show that people's choice behavior (competition vs. lottery) is underconfident when the task is 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.

³ 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.

each of them over- or underestimated himself.⁴ 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 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 agents' (incentivized) beliefs of 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 possible explanation is that 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.⁵ 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.⁶ Thus, it might be the case that women have somehow imprinted the social norm of

⁴ 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.

⁵ This is in line with women's shame of overestimation observed by the experimental study by Ludwig and Thoma (2012).

⁶ 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.

modesty and even downgrade their self-assessment in environments in which the (monetary) need for modesty is absent.⁷

There is an extensive and expanding literature on overconfidence. 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 literature identifying potential consequences of a biased self-assessment. Within this literature many papers focus on non-payoff maximizing decisions caused by a biased self-assessment, e.g., overinvestment, value-destroying mergers of CEOs (Camerer and Lovo 1999; Malmendier and Tate 2005, 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 Yestrumskas 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 (but not true performances) are observed by an audience (anonymity is lifted) in comparison to the situation in which reports are private.⁸ Yet, they find that individuals do not bias their reported self-assessments if the true performance is also publicly revealed. Thus, in their paper subjects do not try to signal high confidence or 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 to strategically bias their reported self-assessment. Moreover, their study does not analyze how one's reported 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.

The perception of self-assessment and the strategic incentives to bias one's self-assessment are addressed in an experimental study by Charness et al. (2012). They investigate whether individuals bias their stated confidence about their performance 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

⁷ Also compare Charness et al. (2012) who use this argumentation to explain overconfidence in non-strategic competitive settings.

⁸ Another study claiming that individuals inflate their reported 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 how much one cares about his image and overconfidence, consequently claiming that "overconfidence is a social signaling bias".

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 the competition, the higher the stated 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.⁹ However, in contrast to my experimental design, in Charness et al. (2012) and Reuben et al. (2012) 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. 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.

To the best of my knowledge this study is the first, to experimentally test 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 on gender differences in self-assessment.¹⁰ I suggest that the observed antipathy towards overconfidence adds to the explanation of the gender difference in reported self-assessment as women report to experience emotions, i.e., the negative attitude towards overconfidence, more intensely 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 observe that women are ashamed to overestimate themselves in public, while men have not.

The rest of this paper is structured as follows. In the next Sect. I describe the experimental design and the two different treatments. In Sect. 3 I present the selection behavior of principals. In Sect. 4 I analyze whether agents anticipate

⁹ 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.

¹⁰ See e.g., Balafoutas and Sutter (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) and Niederle and Vesterlund (2007).

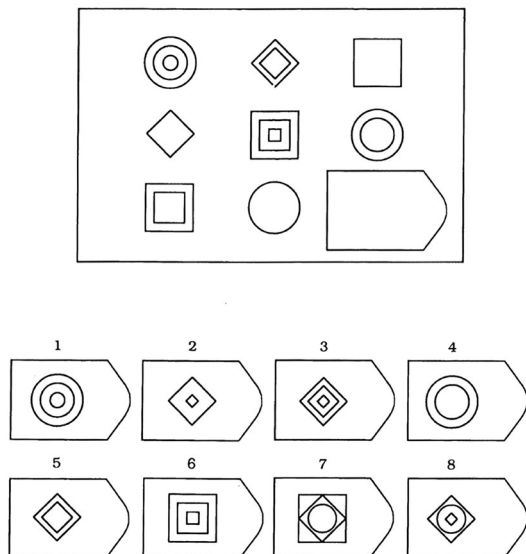
principals' preferences and whether they strategically bias their self-assessment. I conclude in Sect. 5.

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 to solve Raven's Advanced Progressive Matrices, a measure of cognitive ability (Raven 2003). For each matrix participants have to select one out of eight symbols fitting the visual pattern of the matrix. An example of a matrix is given in Fig. 1.

In this task ability and effort are needed to succeed in the task. The participants have 5 min to solve as many matrices as possible. After choosing a symbol, they receive feedback whether their chosen symbol is correct, and thereafter, the next matrix appears. Once they have chosen a symbol, they cannot go back and correct it. It is not possible to skip a matrix without making a choice either. 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 not about their relative performance or the performance of others. They receive five tokens for each matrix they solve correctly. For each wrong answer five tokens are deducted from their earnings. Yet, they receive at least zero 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 task starts, participants solve two matrices as a trial without payment. After part 1 is finished the instructions for part 2 are distributed. The instructions for both parts are read aloud.

Fig. 1 Example of a Raven Advanced Progressive Matrix



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). All 16 agents are ranked from 1 to 8, according to their performance in task 1, where 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 8.¹¹ The two agents having the same rank, are merged to a pair, i.e., there are eight pairs in each session. Each pair is randomly assigned to one of the eight principals.

I conduct two main treatments and two control 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 treatments are the principal's incentives for his choices.

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. Otherwise an inequity averse principal might reward the agent who did not estimate his rank correctly independent of his preferences for an accurate self-assessment.¹²

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. Therefore, 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. Then, the deviations of agents' self-assessments might only hint on a difference in the absolute performance, the agents' expected performance or their self-confidence.

2.1 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. I use the strategy vector method (SVM): before the principal learns the actual deviations of the agents, each principal takes the decision for the same 12 different situations. His choice becomes relevant for the situation, which actually applies to the agents' deviations. For the case that none of the situations applies, the principal also chooses whether the agent who estimates the higher (better) rank or the agent who estimates

¹¹ Two agents receive the same rank to have the same initial position for their self-assessment.

¹² Answers in the follow-up questionnaire show that principals nevertheless expected the agents to state their true belief about their relative ranks.

Table 1 The situations for which principals take a decision

Situations	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
Deviation of agent 1 (A1)	1	-1	2	-2	2	-2	2	-2	3	-3	1	2
Deviation of agent 2 (A2)	0	0	0	0	1	-1	-1	1	-1	1	-1	-2

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 (situations 1–5), the agent with the larger absolute deviation is listed on top

the lower (worse) rank shall receive the 50 tokens. The accuracy of the agents’ estimated ranks is not considered in this decision.¹³

The different situations are listed in Table 1. The deviation is negative if an agent underestimates himself and it is positive if an agent overestimates himself.¹⁴

In situations 1a, 1b, 2a, and 2b one agent estimates his rank correctly and the other agent either under- or overestimates himself. In situations 3a and 4b both agents either under- or overestimate their ranks, but the absolute deviation differs. These situations are to test whether accuracy is preferred over over- and underconfidence. In situations 4a, 4b, 5a, and 5b one agent is over- and the other agent is underestimating his rank and the absolute deviations of the agents differ. The principals’ choices in these situations will reveal whether the size of the deviation is more important than the direction of the deviation. In situations 6a and 6b one agent under- and the other overestimates his rank and both agents have the same absolute deviation. Here the decisions of the principals will reveal whether they prefer under- or overconfidence.¹⁵

For every situation the principal selects the agent he wants to receive the 50 tokens. If the two agents estimate the same rank, and therefore have the same deviation, a situation not covered by the decisions of the principal, chance determines which agent receives the 50 tokens.¹⁶

After his decisions, the principal learns the actual deviations of the two agents and thus which situation 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 get informed 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.

Control-treatment for agents (SYMP-CON) To check whether agents bias their self-assessment in order to be selected by the principal, I conduct a control treatment in which the agent receiving the 50 tokens is not selected by the principal, but by

¹³ This choice actually became relevant for 20 out of 56 principals.

¹⁴ On subjects’ screens the situations were listed in a different order and I varied whether the more or less self-confident agent was listed first, what actually did not influence results.

¹⁵ Besides my interest in whether accuracy is favored over over- and underconfidence, I picked the 12 different situations to cover the most realistic outcomes. Using the SVM instead of showing the actual deviations to the principal decreases the number of sessions needed and allows for a cleaner data analysis.

¹⁶ This situation became relevant in 8 out of 56 situations.

chance. However, to keep social image concerns constant, each pair of agents is assigned to a principal and the principal learns the actual deviations of the two agents. Moreover, I ask the principals to make the same (hypothetical) decisions as in *SYMP*, but without any monetary consequences for any participant. Agents do not know that principals make these hypothetical choices.

2.2 Treatment performance (*PERF*)

In *PERF* the principal takes the same decisions as in *SYMP*, i.e., using the SVM, the principal chooses one out of two agents, based on the deviations of their self-assessments. However, the motivation 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 situations 1–12. 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 who the principal bets on. After his decisions the principal learns about the actual deviations of both agents and which situation is relevant. As in *SYMP*, the principal does not learn the actual or 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 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 rank of the other agent, only when the two estimated the same rank, then they learn that a chance move decides on whom the principal bets.

Control-treatment for agents (*PERF-CON*) The only difference between *PERF-CON* and *PERF* is that in *PERF-CON* agents do not receive any money when selected by the principal. Thus, they do not have a monetary incentive to bias their reported self-assessment. As principals do not know about agents' payment when being selected, *PERF-CON* and *PERF* are identical for principals.

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

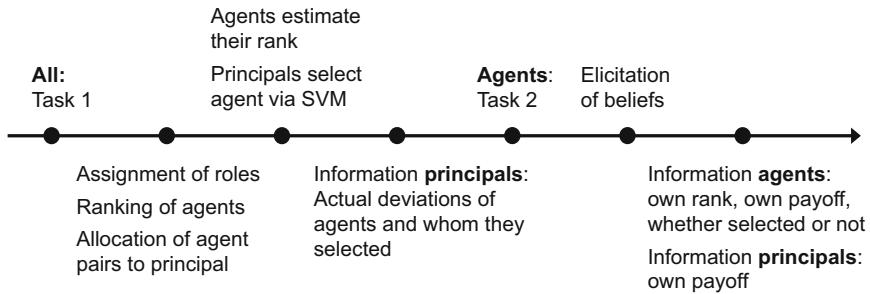


Fig. 2 The timeline of *PERF*

timeline of the treatment *PERF*. *SYMP* has the same course except that there is no task 2.

2.3 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 three sessions of *SYMP* and two sessions of each of the other treatments *SYMP-CON*, *PERF* and *PERF-CON*. 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 1 h. Subjects earned 12.15 Euros on average (including a show-up fee of 4 Euros).

3 Experimental results

In this section I analyze the choice behavior of principals. I will analyze agents' self-assessment in the next section.

3.1 Treatment *SYMP*

For the analysis of principals' choices I pool the real (*SYMP*) and hypothetical (*SYMP-CON*) choices as they do not differ significantly.¹⁷ Moreover, in this section, I refer to the results in *SYMP* which are actually the pooled results of *SYMP* and *SYMP-CON*. Table 2 reports the shares of principals choosing agent 1 or agent 2 for each of the 12 different situations.

¹⁷ Results go in the same direction when excluding hypothetical choices, but are less significant due to fewer observations. The separated data is listed in the [Appendix](#).

Table 2 Principals' selection behavior in *SYMP*

Situations	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
Deviation of A1	1	-1	2	-2	2	-2	2	-2	3	-3	1	2
Deviation of A2	0	0	0	0	1	-1	-1	1	-1	1	-1	-2
Share selecting A1 (%)	10	5	2.5	7.5	10	10	5	32.5	2.5	27.5	35	12.5
Share selecting A2 (%)	90	95	97.5	92.5	90	90	95	67.5	97.5	72.5	65	87.5
Binomial test (p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.01	0.08	0.00
McNemar test (p-value)	0.69		0.50		1.00		0.01		0.00		0.00	

Every principal took every decision 1a–6b. Number of observations is 40 (24 real and 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 situation. p-values are for two-sided tests. The McNemar test tests for a difference between situations a and b

The results clearly show that principals have a preference for accuracy. If agents' absolute deviations differ (situations 1a–5b), the agent who guesses his rank correctly (situations 1a–2b) or who has the smaller absolute deviation (3a–5b) is selected significantly more often.

Result 1 Principals reward agents with a correct guess or with smaller deviations significantly more often than agents with larger deviations, regardless of the sign of the deviation.

If the absolute deviations are the same (situations 6a and 6b) principals favor agents who are underconfident rather than overconfident. In situation 6a 65 % of principals (26 out of 40) reward the agent who underestimates his rank by one rank, whereas only 35 % of principals (14 out of 40) reward the agent who overestimates his rank by one rank. This distribution is weakly significantly different from a 50:50 split (two-sided binomial test, $p = 0.080$). The result is even stronger in situation 6b, in which 87.5 % of principals (35 out of 40) reward the agent who underestimates his rank by two ranks and only 12.5 % of principals (5 out of 40) reward the agent who overestimates his rank by two ranks (two-sided binomial test, $p = 0.000$). If the absolute deviation is unknown, and principals are asked to choose either the agent estimating the lower rank or the agent estimating the higher rank, 67.5 % of principals reward the agent estimating the lower rank (two-sided binomial test, $p = 0.038$).

Result 2 Agents who are underconfident are rewarded significantly more often than agents who are overconfident if the absolute deviation is the same.

The preference for underconfidence over overconfidence can also be seen when taking a closer look at situations 4 and 5. In 4a (5a) only 5 % (2.5 %) of principals

choose the less accurate, *overconfident* agent, whereas in 4b (5b) 32.5 % (27.5 %) choose the less accurate, *underconfident* agent. The differences between situations 4a versus 4b and 5a versus 5b are significant (McNemar $p = 0.007$ and 0.002). Note that there is no difference in selection behavior between situations 1a versus 1b and 2a versus 2b. Thus, the higher selection rates of the less accurate agent in situations 4b and 5b, in comparison to 4a and 5a, seem to be driven by an aversion towards overconfidence.

These results are confirmed by probit regressions, which are reported in Table 3.¹⁸ The dependent variable is 1 if the agent with the larger absolute deviation is selected. I include a dummy for the situations, in which the agent with the larger absolute deviation states the lower rank (situations 1b, 2b, 3b, 4b, and 5b). Column (1) includes principals' choices for the situations 1–3, i.e., the situations in which either one agent has a deviation of 0 (situations 1 and 2), or both agents are either under- or overconfident (situations 3). Column (2) includes the principals' choices for the situations 4 and 5, i.e., all situations in which agents have a different absolute deviation and one agent is under- and the other agent is overconfident. In both regressions I include control variables for risk and gender (of principals) and cluster on principals.¹⁹

The coefficient of the dummy variable in situations 1–3 is not significant, while in situations 4–5 it is highly significant. This shows that the less accurate agent is more likely to be selected if he is under- and the other agent is 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*

As the treatments *PERF* and *PERF-CON* are identical for the principals and only differ in the agents' payment about which principals are not informed, I pool the data of the principals' decisions and talk about the results in *PERF*, while actually meaning the results of *PERF* and *PERF-CON*. The principals' choice behavior is reported in Table 4.

In contrast to *SYMP* the preference for accuracy is less pronounced. In the situations 1a, 2a, 3a, 4a, 5a the more accurate agent is the one being less confident about his performance. Here, significantly more principals select the more precise and less confident agent. However, in situations 2b, 3b, 4b, 5b, 6b, the more accurate agent is the one being more confident about his performance. Here, the more accurate agent is not selected significantly more often, i.e., selection rates are

¹⁸ The results are robust to OLS regressions.

¹⁹ 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) Situations 1–3 Coefficient (p-value)	(2) Situations 4–5 Coefficient (p-value)
Dummy situation b	0.003 (0.990)	1.233 (0.000)
Female	−0.235 (0.508)	0.015 (0.960)
Risk	−0.044 (0.509)	−0.110 (0.121)

I cluster standard errors on principals. Number of observations is 240 in (1) and 160 in (2). In situations 1–2 one agent guesses his rank correctly, in situations 3 both agents' deviations have the same sign. In situations 4–5 one agent is overconfident, the other agent is underconfident. In situations b the agent with the higher absolute deviation states the lower rank

Table 4 Principals' selection behavior in *PERF*

Situations	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
Deviation of A1	1	−1	2	−2	2	−2	2	−2	3	−3	1	2
Deviation of A2	0	0	0	0	1	−1	−1	1	−1	1	−1	−2
Share selecting A1 (%)	28.1	53.1	15.6	40.6	21.9	46.9	12.5	53.1	15.6	46.9	37.5	25
Share selecting A2 (%)	71.9	46.9	84.4	59.4	78.1	53.1	87.5	46.9	84.4	53.1	62.5	75
Binomial test (p-value)	0.02	0.86	0.00	0.38	0.00	0.86	0.00	0.86	0.00	0.86	0.21	0.01
McNemar test (p-value)	0.06		0.02		0.06		0.00		0.00		0.42	

Every principal took all 12 decisions. Number 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 situation. Reported p-values are for two-sided tests. The McNemar test tests for a difference between situations a and b

not significantly different from a 50:50 split as confirmed by binomial tests reported in Table 4. Note that even for situations 1b and 2b, in which one agent correctly assesses himself and the other agent underestimates himself, roughly 50 % of principals select the underconfident agent.

Result 3 More accurate agents are expected to perform better than overconfident agents.

Result 4 Underconfident agents are expected to perform as good as agents who estimate their rank more precisely or even correctly.

If agents' absolute deviations are equal (situations 6a and 6b), the underconfident agent is selected more often than the overconfident agent. In situation 6a 62.5 % of principals (20 out of 32) bet on the agent who underestimates his rank by one rank,

Table 5 Probit of selection of the less accurate agent in *PERF*

	(1) Situations 1–3 Coefficient (p-value)	(2) Situations 4–5 Coefficient (p-value)
Dummy situation b	0.700 (0.001)	1.080 (0.000)
Female	0.150 (0.633)	−0.093 (0.785)
Risk	0.060 (0.337)	0.015 (0.827)

I cluster standard errors on principals. Number of observations is 192 (1) and 128 in (2). In situations 1–2 one agent guesses his rank correctly, in situations 3 both agents' deviations have the same sign. In situations 4–5 one agent is overconfident, the other agent is underconfident. In situations b the agent with the higher absolute deviation states the lower rank

while only 37.5 % (12 out of 32) bet on the agent who overestimates his rank by one rank. However, the result is not significant when using a two-sided binomial-test ($p = 0.216$). Yet, in situation 6b 75 % of principals (24 out of 32) bet on the agent underestimating his rank by two ranks, while only 25 % (8 out of 32) bet on the agent overestimating his rank by two ranks. This is significantly different from a 50:50 split (two-sided binomial-test, $p = 0.008$). Moreover, 62.5 % select the agent estimating the lower rank and 37.5 % select the agent estimating the higher rank ($p = 0.216$).

Result 5 Underconfidence seems to be perceived as a stronger signal for future performance than overconfidence.

These results also become apparent when conducting the same two probit regressions as in *SYMP*. The results are reported in Table 5.²⁰

As above, the dependent variable is 1 if the principal selects the agent with the higher absolute deviation. The significance of the coefficients of the dummy variables show that the less accurate agent is more likely to be selected if he is underconfident. Note that in contrast to *SYMP* the coefficient in column (2) is also highly significant. This shows 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 under- and not overconfident.

In addition, the within-data analysis of principals shows that roughly one third of the 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 situations in which selection is distributed equal, the selection does not occur randomly, but most principals have

²⁰ The results are robust to OLS regressions.

a preference for accuracy or a low self-assessment. This is also confirmed by answers given in the questionnaire.

The data shows that both agents are equally likely to win the competition, independent of their self-assessment. Considering all agents stating the worse rank, a share of 52 % (14 out of 27) wins task 2.²¹ Yet, only very few principals seem to choose randomly, but rather seem to have a clear preference. In the questionnaire I asked subjects to state reasons for their choice. The most frequently given answer was that they expect the underconfident agent to try harder to improve and thus to exert more effort in task 2 than the overconfident agent. The few principals who choose the overconfident agent stated that they expect self-confidence to enhance performance. The reason for selecting the accurate agent was that an agent being able to estimate his performance correctly is expected to have a high overall level of performance.

4 Further analysis

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

4.1 Anticipation of principals' preferences

After the principals' choices, agents estimate principals' selection behavior for the situations listed in Table 6. For each situation they estimate how many of the eight principals in their session selected the agent with the deviation listed first. One question is randomly chosen for payment and participants receive 1 Euro if their answer does not differ more than ± 0.5 from the correct answer. The agents' average estimations are reported in Table 6.

In *SYMP* agents anticipate the preference for a correct self-assessment (1a, 1b) but underestimate the preference for smaller mistakes (4b). However, they anticipate principals' preference for underconfidence and their aversion towards overconfidence (6a). In *PERF* agents correctly anticipate that underconfident agents are expected to perform as good as more precise agents (1b, 4b), but do not anticipate the strong preference for underconfident agents (6a).

4.2 Strategic adaptation of reported self-assessment

Agents only anticipate principals' preference for underconfidence in *SYMP*, but not in *PERF*. If agents adapt their reported self-assessment to be selected by the principal, agents' self-assessment should thus be lower in *SYMP* than in *PERF*. This actually

²¹ Seven out of 15 agents stating the worse rank win task 2 in *PERF* and 7 out of 12 in *PERF-CON*. 10 pairs of agents state the same rank.

Table 6 Agents' average estimation of principals' choice behavior

	Situations				
	1a	1b	4a	4b	6a
Deviation A1	1	-1	2	-2	1
Deviation A2	0	0	-1	1	-1
Treatment <i>SYMP</i>					
Expected share selecting A1 (%)	14.9	25.8	17.5	55.5	29.1
Expected share selecting A2 (%)	85.1	74.2	82.5	44.5	70.9
Treatment <i>PERF</i>					
Expected share selecting A1 (%)	41.8	45.3	39.9	47.4	50
Expected share selecting A2 (%)	58.2	54.7	60.1	52.6	50
Mann-Whitney <i>U</i> -test (p-value)	0.000	0.002	0.000	0.038	0.000

Number of observations is 48 in *SYMP* and 32 in *PERF*. *SYMP-CON* and *PERF-CON* are not included. The last row reports p-values of two-sided MWU-tests for treatment differences

holds true as reported further below. Yet, as the settings in the two treatments are different, I cannot exclude that other factors are also responsible for the difference in reported self-assessments. To cleanly check whether agents adapt their reported self-assessment to principals' preferences, I compare the stated ranks in *SYMP* and *PERF* with the stated ranks in the control treatments *SYMP-CON* and *PERF-CON*.

To control for potential differences in performance across treatments, I calculate the deviations of agents' stated ranks instead of only comparing agents' reported self-assessments. However, as agents' ranks are determined endogenously within a session, agents' actual ranks might be influenced by performance differences across sessions. Therefore, I use *simulated* ranks instead of the actual ranks to calculate agents' deviations. The *simulated* rank of an agent is the rank that is most likely assigned to him, given his performance and the performance distribution of all agents in all treatments.²² I calculate the deviation of an agent as his *simulated* rank minus his actual rank. Thus, a negative deviation represents underconfidence, a positive deviation represents overconfidence. Table 7 gives an overview of agents' average guessed and simulated ranks as well as their average deviation in all treatments.

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 bias their reported self-assessment upwards in *PERF* or

²² To calculate the *simulated* 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 *simulated* rank equals the mode of all 500,000 calculated ranks.

²³ I cannot confirm the highly debated finding that subjects are overconfident. In this experiment, 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.

Table 7 Guessed and *simulated* ranks of agents

	Number of observations	Guessed rank	<i>Simulated</i> ranks	Deviation
<i>SYMP</i>	48	4.08	3.96	-0.13
<i>SYMP-CON</i>	32	4.22	4.47	0.25
<i>PERF</i>	32	4.47	5.28	0.81
<i>PERF-CON</i>	32	4.00	4.19	0.19

An agent's deviation is calculated as *simulated* rank – guessed rank

downwards 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 reported self-assessment in *SYMP* than in *SYMP-CON*. A possible explanation for the missing difference 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 state a lower self-assessment if overestimation becomes public, whereas men do not. This effect might bias women's reported self-assessment in the same direction as the ambition of being selected by the principal, leading to a low reported self-assessment in both treatments. As men do not seem to be influenced by observable overestimation I conduct ordered probit regressions for each gender separately. The results are reported in Table 8.²⁵ The dependent variable is the guessed rank. The independent variables are the *simulated* rank (as performance measure) and dummies for the treatments *PERF*, *SYMP-CON*, and *PERF-CON*. I also control for risk aversion.

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

There is no significant difference in reported self-assessment for women across *SYMP* and *SYMP-CON*. Note that there is no gender difference in reported self-assessment (each treatment considered separately), with one exception: *SYMP-CON*. Here, men rank themselves higher than women (MWU, two-sided, $p = 0.032$) which fits to the hypothesis that men downgrade their reported self-assessment in *SYMP*.

Besides a conscious downgrade of reported self-assessment, it can be the case that the social aversion towards overconfidence influences subjects' self-assessment unconsciously. Subjects' reported self-assessment might not only be influenced in situations, in which others learn their self-assessment, but per se. Charness et al.

²⁴ Note that *simulated* 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.

²⁵ The results are robust when conducting OLS regressions.

Table 8 Ordered probit of guessed rank

	(1) Women	(2) Men
<i>Simulated rank</i>	0.288***	0.418***
<i>PERF</i>	-0.264	0.125
<i>SYMP-CON</i>	0.058	-0.566*
<i>PERF-CON</i>	-0.049	-0.230
Risk	0.001	-0.100
Number of observations	78	66

Table reports the coefficients. Base case is *SYMP*. The sample consists of all agents and the regression clusters on sessions

*** Significant on 1 % level,
* 10 % level

(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 internalization 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 likeable 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 reported self-assessments to a treatment, in which the principal cannot make a selection choice (non-strategic setting) shows that men slightly deflate their reported 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 reported self-assessment in the non-strategic setting, as principals still learn the deviation of agents' reported self-assessment. Thus, apart from monetary consequences women might expect their social image to suffer. Moreover, besides the conscious adaptation of reported 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 reported self-assessment consciously or unconsciously, the results reported in this paper provide an important fact one should consider when eliciting and interpreting individuals' stated self-assessment: subjects might not

state the self-assessment that they actually have, but that has the largest signaling value.

Acknowledgments I would like to thank Martin Kocher, Sandra Ludwig, Ernesto Reuben, Klaus Schmidt, Marta Serra-Garcia, Sebastian Strasser, Ferdinand Vieider, participants of the MELESSA Brown Bag Seminar at the University of Munich for helpful comments and discussions. Financial support from Sonderforschungsbereich (SFB)/Transregio (TR) 15 is gratefully acknowledged. For providing laboratory resources I kindly thank MELESSA of the University of Munich.

Appendix

See Table 9.

Table 9 Principals' choices in *SYMP* separated

		<i>SYMP</i> (%)	<i>SYMP-CON</i> (%)
Situation 1a			
Deviation of A1	1	91.7	87.5
Deviation of A2	0	8.3	12.5
Situation 1b			
Deviation of A1	-1	95.8	97.8
Deviation of A2	0	4.2	2.2
Situation 3a			
Deviation of A1	2	95.8	100
Deviation of A2	0	4.2	0
Situation 3b			
Deviation of A1	-2	91.7	93.8
Deviation of A2	0	8.3	6.2
Situation 2a			
Deviation of A1	2	91.7	87.5
Deviation of A2	1	8.3	12.5
Situation 2b			
Deviation of A1	-2	91.7	87.5
Deviation of A2	-1	8.3	12.5
Situation 4a			
Deviation of A1	2	95.8	93.8
Deviation of A2	-1	4.2	6.2
Situation 4b			
Deviation of A1	-2	70.8	62.5
Deviation of A2	1	29.2	37.5
Situation 5a			
Deviation of A1	3	95.8	100
Deviation of A2	-1	4.2	0
Situation 5b			
Deviation of A1	-3	66.7	81.2

Table 9 continued

		<i>SYMP</i> (%)	<i>SYMP-CON</i> (%)
Deviation of A2	1	33.3	18.8
Situation 6a			
Deviation of A1	1	62.5	68.8
Deviation of A2	-1	37.5	31.2
Situation 6b			
Deviation of A1	2	91.7	81.3
Deviation of A2	-2	8.3	18.7
Situation 7			
Higher guess		33.3	31.2
Lower guess		66.7	68.8

Each principal took every decision 1a–7. Number of observations is 24 in *SYMP* and 16 in *SYMP-CON*. A positive deviation represents “overconfidence”, whereas a negative deviation represents “underconfidence”. *SYMP-CON* is the control treatment of *SYMP*, in which principals make hypothetical choices without any payoff consequences

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). Affirmative action policies promote women and do not harm efficiency in the lab. *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*, *80*(3), 949–983.
- Camerer, C., & Lovo, 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 ready-made 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. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63 (pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.
- 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.
- Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2012). *Managing self-confidence: Theory and experimental evidence*. Mimeo.
- 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.
- Myerson, R. (1991). *Game theory: Analysis of conflict*. Cambridge, MA: Harvard University Press.
- 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., Raven, J. C., & Court, J. H. (2003). *Raven's Progressive Matrices und Vocabulary Scales. Grundlagenmanual*. Frankfurt: Pearson Assessment.
- 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*, 5(3), 124–156.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47, 143–148.