Do You Know That I Am Biased? An Experiment

Sandra Ludwig[a*] and Julia Nafziger[b†‡]

[a]University of Munich

[b]ECARES and Solvay Business School, Université Libre de Bruxelles

July 26, 2007

Abstract

This experiment explores whether individuals know that other people are biased. We confirm that overestimation of abilities is a pervasive problem, but observe that most people are not aware of it, i.e. they think others are unbiased. We investigate several explanations for this result. As a first one, we discuss a possible unfamiliarity with the task and the subjects' inability to distinguish between random mistakes and a real bias. Second, we show how the relation between a subject's belief about others and his belief about himself might be driven by a false consensus effect or self-correction mechanism. Third, we identify a self-serving bias when comparing how a subject evaluates his own and other people's biases.

Keywords Bias, Overconfidence, Beliefs, Experimental Economics, Self-Serving Bias
JEL Classification: D83, C91, D01

*University of Munich, Ludwigstrasse 28 (Rgb), 80539 Munich, Germany, Email: sandra.ludwig@lrz.uni-muenchen.de, Tel.: +49 (0)8921803677.

†ECARES, Université Libre de Bruxelles, Avenue F D Roosevelt 50, CP 114, 1050 Bruxelles, Belgium, Email: julia.nafziger@ulb.ac.be, Tel.: +32 (0)26504602.

‡We would like to thank Johannes Abeler, Ethan Cohen-Cole, Simon Gächter, Paul Heidhues, Denis Hilton, Erik Hölzl, Hannah Hörisch, Steffen Huck, Philipp Kircher, Alexander Koch, David Laibson, Ulrike Malmendier, Felix Marklein, Luis Santos-Pinto, Burkhard Schipper, Karl Schlag, and Matthias Sutter for helpful comments and discussions. Financial support from the Bonn Graduate School of Economics is gratefully acknowledged.

1 Introduction

Biased individuals act differently from unbiased individuals. There are innumerable examples for this. For instance, having overly optimistic expectations about the future or the environment results in acting against long odds when choosing an investment strategy, making insurance choices, or entering a new market. One major reason for traffic accidents is that people overestimate their driving abilities. CEOs, who overestimate their success probability, may agree too often to a merger. Overestimation of one's (relative) abilities can lead to wrong job application or search behavior in the labor market.
When interacting with other people, it thus is important to know whether they are biased to react appropriately. Consider shareholders, who delegate to a manager the decision whether or not to conduct a merger. If the manager overestimates the success probability of the merger, his decision will be suboptimal for the shareholders and it might be better not to delegate the decision to him. Also in strategic interactions between biased agents, a player's behavior depends on the knowledge about the bias of his opponents.[1] Consider a manger's decision whether or not to enter a new market, where the success of entry depends on the (relative) quality of his good and the number of competitors in the market. Excess entry can be driven by two forces: managers overestimate the (relative) quality of their own good and/or they do not know that their opponents are overconfident. These two forces are, however, hard to disentangle and it might not be clear what drives excess entry. Thus, in this experiment, we want to abstract from all kinds of strategic interactions and ask in a simple decision problem whether people know that others are biased.

Our baseline experiment considers in the simplest possible setting the question whether people know about the bias of others. In a first treatment, subjects answered several difficult – but not tricky – multiple-choice questions and estimated how many questions they have correct. In a second treatment, we then informed subjects about the first experiment, showed them the questions, told them the others' average guess about the number of correct answers and "asked"[2] whether they think the true number is roughly equal, higher or lower.
While we observe in the first treatment that subjects overestimate their ability, subjects in the second treatment are not aware of this. As overconfidence is such a prominent phenomenon in the real world, this result is surprising. The aim of our extensions is to further examine this result and understand the driving forces behind it. One possible explanation for it is that people have no real life experience with the task and that this causes their unawareness of overconfidence. In a first extension, we cause this familiarity by letting subjects answer the questions and estimate their own number of correct answers before assessing the

[1] Theoretical models typically assume that an agent knows that (a proportion of) his opponents are biased.

[2] We should mention here that we never asked subjects directly, nor did we use words like bias, or overestimation, but implemented simple decision problems to elicit beliefs.

others. While now more subjects recognize that others are overconfident, still the majority thinks that they are unbiased. In a second extension, we try to further increase the subjects' familiarity – by writing the instructions in a non-neutral way using words like over- and underestimation. Our result shows that this helps to make subjects recognize that others are *biased*, but not that they are overconfident. Lastly, we let subjects evaluate other subjects who answered tricky questions. Before the evaluation, we showed them the correct answers. This strong signal finally helped to make subjects recognize that others are overconfident, however, they still do not recognize the full extent of the overconfidence bias.

These observations indicate that people are not fully aware that others are on *average* overconfident. What do they think about single subjects? Do they think an individual makes mistakes when evaluating her number of correct answers? Do they think these mistakes are just random or that they mirror a real bias? For this we ask subjects about a single other subject. And indeed, they notice that a single subject's self-assessment can be wrong. We show that "wrong" for them means that the other subject makes a random mistake, but has no bias. Thus, they think not only the average of subjects is unbiased, but also single subjects are.

As in all these extension subjects answer the questions and evaluate themselves, they provide us furthermore with the opportunity to examine how the belief about others is shaped by the subjects' beliefs about themselves. First, we find evidence for a false consensus effect, i.e. subjects adjust their belief about others in the direction of their own belief. Moreover, subjects think that others, who are more similar to them, are more likely to be correct. Second, we observe a self-correction mechanism: For example, those subjects, who recognize that others are overconfident, are underconfident themselves. It is likely that – being aware of the overconfidence problem – they adjust their belief about themselves too heavily – which results in own underconfidence.

The aim of our last extension is to deepen the understanding what people know about biases. When overconfident agents interact with each other, it is not only important that they know about the bias of their opponent, but also about the relation between their own bias and the one of their opponent. To examine this, we let subjects (in all mentioned extensions) make an additional choice that reveals their beliefs about the relation of their own bias/mistake and the average bias of the others. Concerning these relative biases, the largest group of subjects thinks that they are themselves more likely to judge their ability correctly than is the average population. The result is consistent with a self-serving bias. Still it is surprising because "the others" represent an average: If one believes that people are unbiased, mistakes should cancel out *on average*, but not for a single subject. Thus, while subjects are aware that others make mistakes, they seem not to be aware that they can make mistakes themselves.

Much of the evidence for overconfidence comes from calibration studies by psychologists, in which subjects make probability judgements. People's confidence often exceeds their actual accuracy (for a review of this literature see Yates, 1990). Besides being poorly calibrated, people also state confidence intervals that are too narrow. There are also experiments by economists who confirm that individuals are overconfident (Camerer and Lovallo, 1999; Hoelzl and Rustichini, 2005). In Camerer and Lovallo's (1999) market entry game, people's behavior might not only be driven by their overconfidence, but also by their belief about the bias of others – something which is, however, hard to disentangle in their experiment.
There are only few papers that deal with the knowledge about biases. One strand considers time inconsistent preferences: People behave differently depending on whether they do or do not know that they behave inconsistently over time (for theoretical models see e.g. Laibson (1997), O'Donoghue and Rabin (1999), for a field study where time inconsistent preferences matter see Della Vigna and Malmendier (2006)). Another strand examines the behavior of individuals in so-called beauty contests (Bosch-Domenech et al., 2002; Ho et al., 1998). In these games, the strategy of a player depends on how rational he thinks that his opponents are, i.e. how many steps of iterated elimination of strictly dominated strategies they are able to perform. Finally, several theoretical papers – that ask about the behavior of overconfident agents in economic situations – assume that individuals know that others are biased. We discuss these papers in Section 6.

The paper is structured as follows. In Section 2, we describe our baseline experiment, which consists of two treatments (Treatment A and B). This baseline experiment answers our very basic question whether people are aware that other people are biased. We then introduce the three extensions of our baseline experiment: we make subjects more familiar with the task (Treatments Q and W, which we describe in Section 3), we distinguish between random mistakes and biases (Treatment S, described in Section 4), and finally, we ask what people think about the relation between their own and other peoples' bias (Section 5). In Section 6, we discuss how to deal with one possible drawback of our experiment (the selection of questions), the procedure to measure the subjects' biases, and applications of our results to organizations and markets. In Section 7, we conclude.

2 Baseline Treatments: Do You Know That Others Have a Bias?

We conducted the computerized experiment at the University of Bonn. We programmed the experiment with the software z-Tree Fischbacher (1999) and recruited subjects via the internet by using the software ORSEE (Online Recruitment System for Economic Experiments) developed by Greiner (2004). A total of 116 subjects participated in six treatments (with

17 to 20 participants each). Each subject participated in only one of the treatments. Before the experiment started, we read out loudly the instructions[3] and the subjects answered clarifying questions to make sure that they understand the experimental procedure. We kept the wording of all but one instructions (see later) neutral to avoid framing effects: We did not use terms like self-assessment, ability, or overconfidence – although we use them in the following to describe the design. Subjects earned Tokens during the experiment, where 210 Tokens = 1 Euro. Average hourly earnings were 8 Euros.

2.1 Experimental Design

To address our main question whether people know about the overconfidence bias of others, we conducted two treatments, A and B. In Treatment A, 20 subjects ("*A subject*" or "she") answered multiple-choice questions and estimated their number of correct answers. In treatment B, we asked 20 other subjects ("*B subject*" or "he") whether they think that the *A subjects* estimated their number of correct answers *on average* correctly or not. This means that we define the bias of the *A subjects* as the average difference between their estimated and true number of correct answers. In contrast, from observing that a single *A subject's* true and estimated number diverge, one cannot say whether she is really biased or makes a random mistake.

2.1.1 Treatment A

Treatment A consists of two stages. In the first stage, 20 *A subjects* answer seven difficult multiple-choice questions from different fields of general knowledge. Questions are such that a subject should know the answer more or less for sure or not at all. We paid subjects 190 Tokens for each correct answer. In the second stage, an *A subject* estimates her number of correctly answered questions without knowing her true number of them $t \in \{0, 1 \ldots 7\}$. We denote this estimate by $q \in \{0, 1 \ldots 7\}$.

To achieve that a subject states her true belief about t, we paid her 525 tokens if $q = t$ and 30 else. Subjects infer these payoffs from Table 1. Like this we can elicit beliefs via a neutral decision task without asking subjects directly "how many question do you think you answered correctly" as such a question could influence their choice e.g. via self-esteem protection channels.[4]

[3]Instructions are available upon request.

[4]The procedure ensures that subjects select the alternative on which they place the highest probability (see the appendix.). In the following we just say "the *A subject* thinks she has x answers correct", which should read as "the *A subject* puts the highest probability on the event that she has x answers correct". As we later on take the average over all subjects this makes no difference: choosing x when one is roughly indifferent between x and say $x - 1$ and $x - 1$ is true is an random mistake if subjects were unbiased, i.e. it should cancel out on average.

	Correct Answers (t)			
	0	1	...	7
Action 0	525	...	30	30
Action 1	30	525	...	30
...	...	...	...	...
Action 7	30	30	...	525

Table 1: Payoff table to elicit belief about number of correct answers in Treatment A.

While this payoff structure has the disadvantage that larger deviations are not punished more, it ensures truth telling under very mild behavioural assumptions ("more is better") as we show in the appendix. In contrast, to punish larger deviations and to ensure truth telling, one had to assume risk neutrality which is likely to be violated even for small stakes (see e.g. Holt and Laury, 2002; Rabin, 2001). Furthermore, we wanted to make sure that subjects do not use a hedging strategy, like answering no question in the first stage to be able to make the correct choice in the second stage. To circumvent this problem, we decided to implement the following procedure: When answering the questions, an *A subject* knows that she has to make a decision later on. She also knows that her payoff for this decision depends on her number of correctly answered questions but does not know yet the exact task and the relevant payoff table. This procedure is necessary as with the payoff structure itself, we could not eliminate hedging concerns completely unless we rewarded subjects much more for a correct answer than for a right decision. This would, however, not have mirrored the importance of the tasks.

2.1.2 Treatment B

In Treatment B, 20 subjects ("*B subject*" or "he") receive in their instructions detailed information about Treatment A: the full procedure of the experiment, the multiple-choice questions (without indication of correct answers), and the average belief about the number of correct questions (rounded to one decimal place), which we denote by $\bar{q}$ in the following. Given this information, *B subjects* have to state whether they think that *A subjects* are overconfident (meaning that $\bar{q}$ exceed the average true number of correct questions ($\bar{t}$) by more than 0.5), underconfident ($\bar{t} - \bar{q} > 0.5$) or unbiased ($|\bar{q} - \bar{t}| \leq 0.5$). By adding/subtracting 0.5 we capture rounding effects and small mistakes which remain on, even though A subjects are unbiased.

Furthermore, *B subjects* state their belief about the exact size of $\bar{t}$ by choosing a number $z \in \{0, 0.1, 0.2, \ldots, 6.9, 7\}$. To avoid hedging possibilities, this choice has to be consistent with the statement about the bias of the *A subjects*: e.g. z has to be larger than $\bar{q} + 0.5$ if a *B subject* stated that the *A subjects* are underconfident.

	Action		
	left	middle	right
$\bar{t} < \bar{q} - 0.5$	315	315	1680
$\bar{q} - 0.5 \geq \bar{t} \leq \bar{q} + 0.5$	315	1680	315
$\bar{t} > \bar{q} + 0.5$	1680	315	315

Table 2: Payoff table to elicit the *B subjects'* beliefs whether the *A subjects* are under- or overconfident or unbiased.

The procedure to elicit these beliefs is analogue to the one described in Section 2.1.1 and uses Table 2: *B subjects* receive a high payoff (1680 Tokens for the assessment about the *A subjects'* bias and 105 for the guess about $\bar{t}$) if they make the correct choice and a low payoff else.

2.2 Results Baseline Treatments

Treatment A confirms that subjects are overconfident. 60% of the *A subjects* overestimate their number of correct answers and, moreover, they are overconfident on average (we measure the bias by $\bar{t} - \bar{q}$): their bias is -1.1. According to a Wilcoxon signed-rank test, we can reject the null-hypothesis that the difference between the individual values of q and t has a median value of zero (p-value: 0.006) and thus say that they differ significantly. The distribution of the values of t and q in Figure 1 illustrates this: the q-distribution is right skewed, the t-distribution is left skewed.[5]

Do the *B subjects* know about this bias? The answer is no. Roughly 75% say that the *A subjects* are correct and only 15% think that they are overconfident. Also the *B subjects'* average belief about the true number of correct answers of the *A subjects* points only to some awareness of overconfidence: They guess that $\bar{t}$ is on average 0.2 units smaller than $\bar{q}$ (Mann-Whitney U test, $p = 0.002$, two-sided). This does not mirror the true degree of overestimation. Both observations show that *B subjects* overestimate the accuracy of the *A subjects'* belief on average.

3 Extension I: Familiarity with the Task

Given that overconfidence is a phenomenon that is present in everyday's life, the above result is surprising. One possible explanation is that estimating the number of correct questions is an unfamiliar task, in which *B subjects* have no further experience, and that

[5]Treatment A is not an outlier: In all other treatments, where subjects estimated their number of correct answers, they overestimated it, where the percentage of subjects who overestimated it varies from 53 to 76 percent (for the difficult questions).

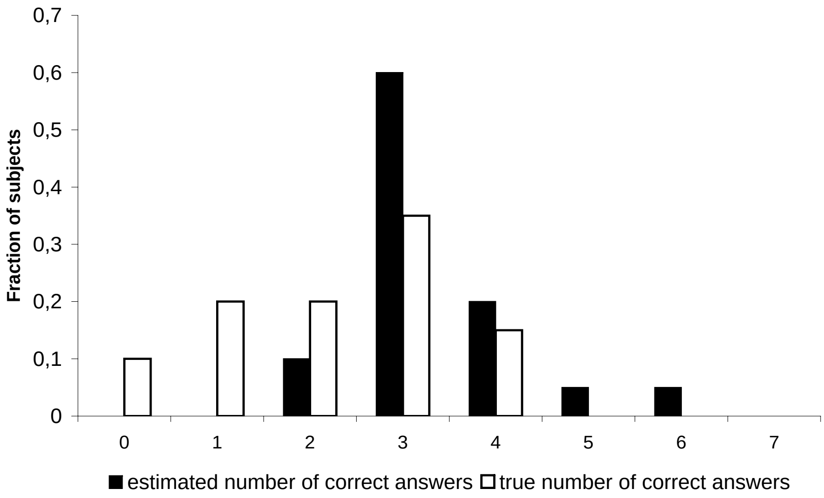

Figure 1: Distribution of true and estimated number of correct answers in Treatment A.

this unfamiliarity causes the *B subjects'* unawareness. To further investigate this issue, we conduct several extensions of our baseline experiment. In these, we give *B subjects* small pieces of information about the task: We let them answer the questions and estimate their number of correct answers themselves, we use non-neutrally written instructions, or tricky questions to which we show them the correct answers. Next to making *B subjects* more familiar with the task, letting them answer the questions themselves, allows us to shed some light on the reasons for their belief about others: Is there a relation between the own bias and the belief about the bias of others?

3.1 Answer Questions Themselves: Treatment Q

The first extension is Treatment Q, which is a combination of Treatments A and B: 17 subjects ("*Q subject*" or "he") first answer the same seven multiple-choice questions which the *A subjects* answered and then estimate their number of correct answers. Afterwards, they receive the same information about Treatment A as the *B subjects* did and state whether they think that the *A subjects* are over-, underconfident, or unbiased. By answering the questions himself, a *Q subject* might not only get a better feeling for the difficulty of the questions and whether the *A subjects'* average guess is realistic, but he might also start to reason better about the problem at hand (compare e.g. Croson, 2000). Our results show that this is indeed the case: Subjects now recognize more often that others are overconfident. 24% of the subjects state that the *A subjects* overestimate their number of correct answers. While this percentage is higher than the 15% in Treatment B, still the majority of 59% thinks that the *A subjects* are on average correct. Moreover, we cannot reject the hypothesis that there is no relation between the number of subjects in Treatments B and Q, who think that the *A subjects* are unbiased or biased, according to a Fisher's exact test ($p = 0.234$, one-sided). Thus, the greater familiarity makes subjects more often recognize that overconfidence is a problem in the population, but still they are not fully aware of it.

By letting the *Q subjects* complete the task of the *A subjects*, we do not only increase their familiarity with the task, but we can also analyze whether there is a relation between a subject's own bias and the belief about the bias of others. To investigate this issue, we plot in Figure 2 three cumulative distribution functions of *Q subjects'* biases – one function for *Q subjects* who think that *A subjects* are unbiased/overconfident/underconfident. We see that the cumulative distribution function of those subjects, who think that the *A subjects* are on average correct, is always below the other two functions – indicating that they have less extreme overconfidence biases.[6] *Q subjects* who think that others are underconfident are in fact most overconfident: They overestimate their number of correct answers by 3 on average. A possible explanation behind this result is the presence of a self-correction mechanism: Those *Q subjects*, who think others underestimate their number of correct answers may also think that they themselves underestimate it – thus, adjusting their estimate about the own number of correct answers, q, upwards – leading in fact to a greater overestimation. In contrast, subjects who think that others overestimate it, adjust their belief downwards, leading more often to correct/better choices. We do not observe in this treatment that subjects, who think that others are overconfident, are more biased (underconfident) than those, who think that others are unbiased – something we observe, however, in the treatment described in the following section, where the familiarity with the task is further increased.

3.2 Answer Questions Themselves, Non-Neutral Wording, Tricky Questions: Treatment W

We have seen in Section 3.1 that some familiarity with the task helps to recognize that others are biased, but does not make subjects fully aware of it. In Treatment W, we increase the subjects' ("*W subject*" or "he") familiarity with the task further by using non-neutral wording in the instructions. We do so for two different types of questions and randomly select which part is paid to avoid hedging.

[6] If we test the biases of those *Q subjects*, who think others are biased (either overconfident or underconfident), against the biases of those, who think others are unbiased, a Mann-Whitney U test indicates a significant difference (p=0.033).

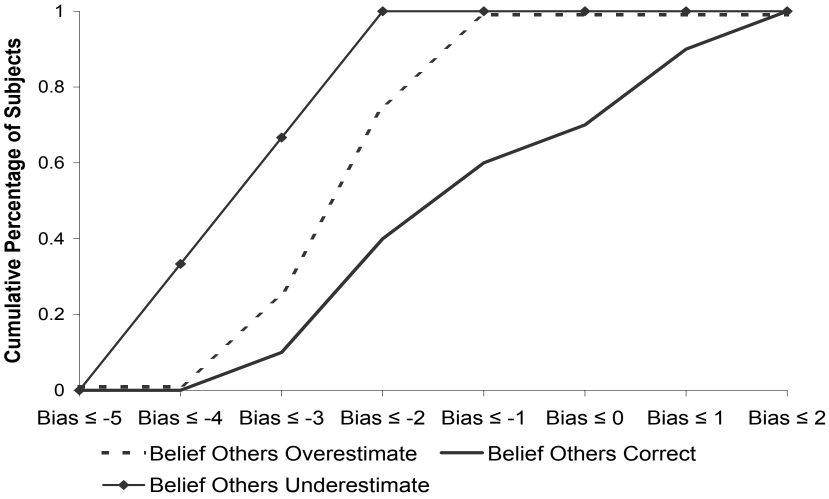

Figure 2: *Q subject's* bias given belief about the bias in *A subjects'* self-assessment. A negative bias refers to overconfidence and a positive bias to underconfidence.

3.2.1 Difficult Questions

In the first stage of Treatment W, we repeat Treatment Q, but the instructions use a non-neutral language. We explicitly asked the 19 *W subjects* "How many questions do you think you have correct?", "Do you think that others over-, underestimate their number of correct questions or estimate it correctly?" and "How many questions do you think the others answered on average correctly?" (i.e. the choice of the number, which we added compared to Treatment Q).

Our first observation is that the greater familiarity increases the awareness about the bias of others: In Treatment W, only 42% of subjects believe that the *A subjects* are unbiased, while 32% think that they are overconfident. Thus, the information that the non-neutral wording of the instructions provides leads to the insight that the *A subjects* are biased. Compared to Treatment B, a Fisher's exact test (p=0.038, one-sided) indicates that the percentage of subjects who think that others are biased (unbiased) increases (decreases). The wording does not, however, lead to the insight that *A subjects* are necessarily overconfident as roughly the same number of *W subjects* thinks that the *A subjects* are over- or underconfident (32% versus 26%). The words "overestimate" and "underestimate" pop up in the instructions in the same proportion – making people think about problems of biased estimates. But reading these words did not animate subjects to ponder whether over-, or understimation is more likely a problem.

As in Section 3.1, we can analyze the relation between a subject's own bias and the subject's belief about the bias of others. Compared to the results in Treatment Q, a self-correction mechanism now explains the results perfectly: Those subjects saying that the *A subjects* underestimate their number of correct answers, overestimate their own number by 1.2 on average. They adjust their belief too much upwards and end up being most heavily overconfident. Those subjects saying that the *A subjects* are roughly correct, overestimate their own number by 0.75 on average. They do not recognize the bias of others, do not adjust the own belief, and therefore, end up being overconfident. And finally those, who say that the *A subjects* are overconfident, are in fact underconfident with a bias of 0.5, they adjust their own belief too much downwards.[7]

In addition to the relation between the own bias and the belief about the bias of others, we can analyze the relation between the belief about one's own number of correct answers and the exact belief about the *A subjects'* true number of correct answers (the guess z). Figure 3 shows that *W subjects* with lower beliefs q (up to 3) guess on average that the *A subjects'* true number of correct answers is lower than $\bar{q} = 3.4$, whereas subjects with higher beliefs (from 4 on) estimate that the *A subjects'* true number of correct answers is larger than 3.4. Thus, *W subjects* adjust their belief about the *A subjects* in the direction of their own belief. The result is supported by the observation that the estimates q and z are positively correlated (Spearman rank-order correlation coefficient 0.737, $p = 0.0002$). A false consensus effect (Mullen et al., 1985) can explain the observed behavior. People who are prone to a false consensus effect tend to overestimate the degree to which, for example, their own behavior or beliefs are shared by other people. This effect implies that subjects overestimate the frequency with which their own estimate q is present in the population.

Up to now, we have seen that subjects tend to overestimate the own number of correct answers as well as the accuracy of other subjects' beliefs. These observations raise the question whether subjects have the same bias when evaluating themselves and when evaluating the *A subjects*. In a study by Baker and Emery (1993), individuals have extremely optimistic expectations assessing the likelihood that they get divorced themselves while they know quite accurately the likelihood of divorces in the population (about 50 percent of U.S. couples who marry). This finding suggests that people may be better at detecting "biases" of other people than biases of themselves since no self-image concerns are involved when evaluating others. This is consistent with our data: On average, the own bias is about 0.26 larger in absolute terms (it is more negative) than the bias in assessing the others' average number of correct answers $\bar{t}$ (Wilcoxon signed-rank test, p=0.015). Thus, a driving force behind overconfidence can be a kind of self-impression motive.

[7]The estimate q of those subjects, who say that others are overconfident (underconfident), is significantly smaller than the estimate q of those, who say that others are unbiased (according to Mann-Whitney U tests $p = 0.004$ and $p = 0.011$, respectively).

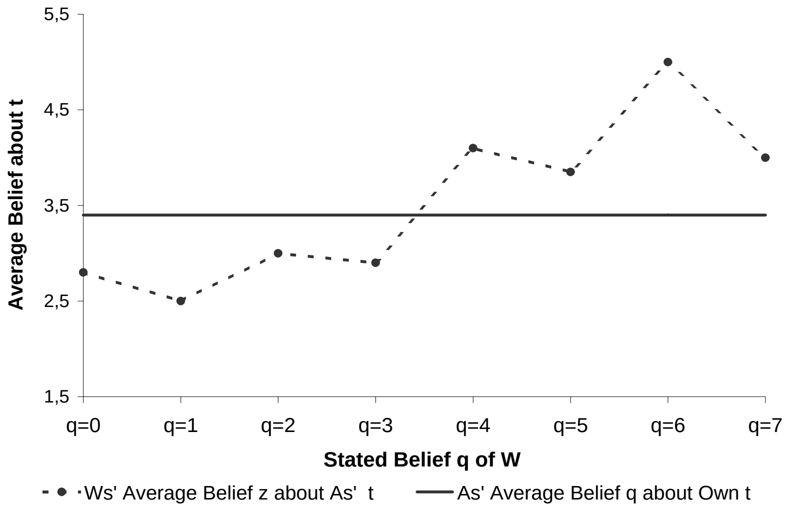

Figure 3: Average belief z about the *A subjects* $\bar{t}$ given the belief q of a *W subject*.

3.2.2 Tricky Questions

The second stage of Treatment W is analogue to the first, except that the *W subjects* answered different multiple-choice questions. Therefore, we also confronted them with a different reference treatment (Treatment AT) in this stage. In addition, we showed them the correct answers to the questions after they answered them and assessed their own number of correct answers, but before evaluating the *AT subjects*. Treatment AT is exactly like Treatment A, but 20 new subjects answer tricky multiple-choice questions instead of difficult ones. These questions look very simple (as one answer seems to be the correct one), but are in fact not. Thus, subjects are quite certain that they choose the right answer, but actually select the wrong one. Showing the correct answers to these questions, therefore should help subjects to recognize that the others overestimate their number of correct answers.

Unsurprisingly, the tricky questions lead to more pronounced overestimation: The average difference between the estimated and true number of correct answers is now 3.4 (Wilcoxon signed-rank test, p=0.000).[8] More importantly, *W subjects* now recognize this problem. Almost all (91%) say that *AT subjects* overestimate their number of correct answers. They still do, however, not fully recognize to which extent overestimation is a problem. They think that the true number of correct answers $\bar{t}$ of the *AT subjects* is around 2.9, whereas it

[8] We discuss the selection of questions and whether one can say that subjects are "more overconfident" with tricky questions in Section 6.

is indeed only 1.2 and the *AT subjects* believe it is 4.6.[9] Moreover, their guess is not much smaller than their guess for the *A subjects* – who answered the difficult and not the tricky questions – which is 3.4.

4 Extension II: Facing a Single Subject (Treatment S)

So far, we investigated whether subjects know about the bias of the group of *A subjects* with the result that a majority does not. We have seen that greater familiarity only alleviates this ignorance. In real world situations, often not the bias of a group, but the one of a single individual matters (e.g. the bias of one's manager or of one's opponent in a tournament). In the following, we want to ask whether subjects know that a single *A subject* can make a mistake when estimating her number of correct answers and whether they think that this mistake is just random or caused by a bias.

4.1 Experimental Design

Treatment S is like Treatment Q, except that subjects ("*S subjects*") now state beliefs about a single *A subject* and not the complete group of *A subjects*. Subjects additionally choose the number that mirrors their belief about another subject's true number of correct answers t. To elicit beliefs about a single *A subject*, we implemented the strategy method: 20 subjects state for every possible estimate $q \in \{0, 1, \ldots, 7\}$ of an *A subject* whether they think she is under-, overconfident or unbiased (strictly speaking: makes mistakes or not). For the numbers 0 and 7 on the boundary, subjects only choose between the two appropriate possibilities. In case he thinks the respective *A subject* is under- or overconfident, he has to choose a number $z \in \{0, 1, \ldots 7\}$, $z \neq q$, that mirrors his belief about her true number of correct answers. A subject in Treatment S was not paid for all his decisions, but for his decision when facing the particular estimate q of a randomly selected *A subject*. When making their decisions, *S subjects* did not know which decision is paid.

4.2 Results Treatment S

The results of Treatment S confirm that the *S subjects* are aware that single *A subjects* make mistakes when estimating their number of correct answers: No *S subject* states for all possible estimates q of an *A subject* that she makes the correct choice. For each value of an *A subject's* belief q, 50 – 95 percent of *S subjects* state that the *A subject* is wrong.
We cannot directly conclude from the above observations that *S subjects* think single *A subjects* are biased or just make a random mistake. To investigate this issue, we compare the

[9] A binomial test indicates that the subjects' guesses are significantly smaller than the *AT subjects'* average belief of 4.6 (p=0.00) and larger than the true average of 1.2 (p=0.00).

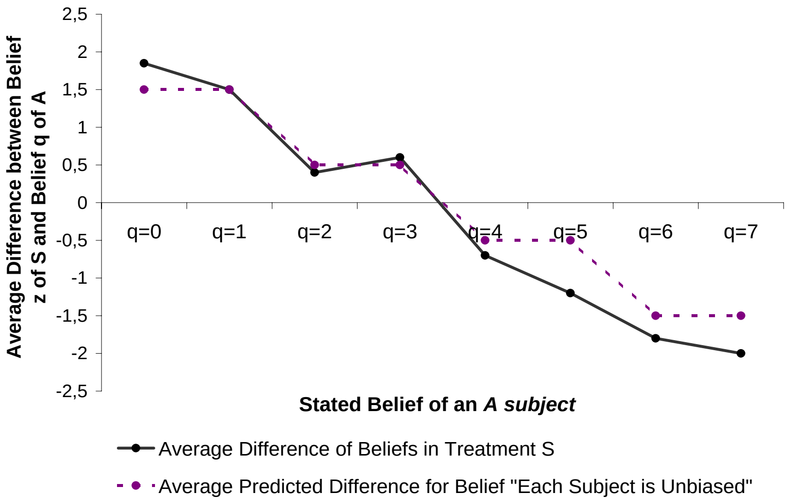

Figure 4: Average value of the difference between the *S subject's* belief z about an *A subject's* t for each possible belief q of an *A subject*.

average choice of subjects in Treatment S with the predicted choice that would be consistent with *S subjects* thinking that an *A subject* is unbiased. For this, we have to specify which values of t can be compatible with a specific estimate q of an *A subject* given the assumption that she is unbiased. We assume that the distribution of the mistake an unbiased *A subject* makes is symmetric around zero and uniform. This implies that e.g. an *A subject*, who states $q = 1$ could actually have a $t \in \{1, 2, 3, 4\}$.[10] In expectation, she thus has 2.5 correct – since all four possibilities are equally likely. Since 2.5 is larger than her stated belief of 1, she underestimates her number of correct answers. We plot the predicted choices of an *S subject* for all possible beliefs of an *A subject* in Figure 4 (the dashed line). Figure 4 also shows the average actual choices of the *S subjects*. One sees that the actual choices are close to the ones predicted for unbiased *A subjects* (especially for low stated beliefs). Hence, *S subjects* consider the mistakes they identify as random and do – as for the average of subjects – not think that a single subject has a bias.

In Section 3.2, we already analyzed the relation between a subject's own bias and his exact belief about the average number of correct answers of a group of other subjects. Now, we want to investigate the relation between a subject's own bias and his belief about the number of correct answers of a *single* other subject. For this we conduct a standard OLS regression.

[10] The drawback of this approach is that the precision of an *A subject* that states extreme beliefs q is higher – for example, someone who says "I answered seven questions correctly" is always right. To circumvent this, one can define unbiasedness such that the expected value of the mistake not for every single subject, but for the population, is zero. The results we show differ, however, only slightly if we use this assumption instead.

We regress the estimates z on the own belief q, the own number of correct answers t, and the belief q^A an *A subject* has about herself, yielding (p-values are in brackets below the coefficients, $R^2 = 0.3438$):

$$\begin{array}{rccccccc} \hat{z} = & 0.161\,q & + & 0.513\,t & + & 0.633\,q^A & + & 0.633. \\ & (0.044) & & (0.00) & & (0.00) & & (0.106) \end{array}$$

Thus, all three variables have a significant and positive effect on the guess about other subjects. Clearly, this guess is related to the belief of an *A subject* – as this is the information an *S subject* receives. Yet, also the belief about himself influences his guess. As in Section 3.2, the latter hints at a false consensus effect. To further investigate this, we distinguish between similar and non-similar individuals: A similar *A subject* has exactly the same belief q about her number of correct answers than an *S subject* has about his number of correct answers. We observe that the average distance between the estimate z for a similar subject and this subject's q is 0.75. But for non-similar subjects it is 1.87 (according to a Wilcoxon signed-rank test the difference is significant, $p < 0.02$). Moreover, the belief of an *S subject* about a similar *A subject* and this *A subject's* own belief are positively correlated (Spearman rank-order correlation coefficient: 0.59, $p = 0.003$). These observations together indicate that a subject thinks that a similar subject is rather correct with her self-assessment. For non-similar individuals, there is no correlation (Spearman rank-order correlation coefficient: -0.089, $p = 0.296$).[11] Thus, for a non-similar individual, the own belief is not crucial for one's estimate.

A second implication of these observations is that a similar subject is considered to be more likely correct than are non-similar subjects: The "adjustment" of a similar *A subject's* belief about herself is smaller than of non-similar ones (see above), and there is no adjustment at all for similar *A subjects* in 60 percent of the observations, whereas for non-similar subjects this is only true for 22 percent. Even more remarkably, subjects also think that similar *A subjects* are likely to be correct if these *A subjects* hold "extreme" beliefs, for which most other subjects say that this extreme belief must be mistaken.[12]

5 Extension III: Relative Bias

In many strategic interactions, it is not only important to know whether other subjects are biased, but also to know about the relation between the own bias and the bias of others: Who is more likely to be correct or biased – the other subjects or oneself? We address this issue in the third extension.

[11] Moreover, if we do the above regression again, but only for the non-similar individuals, the own belief is no longer significant.

[12] For *A subjects* that are *not* similar and who have beliefs $q \in \{0, 4, 5, 6, 7\}$, the distances between q and the estimate z of an *S subject* are significantly larger than they are for similar *A subjects* ($p \leq 0.02$).

	Alternative I	Alternative II
$q = t$ and $\lvert\bar{q} - \bar{t}\rvert < 0.5$	800	800
$q = t$ and $\lvert\bar{q} - \bar{t}\rvert > 0.5$	500	300
$q \neq t$ and $\lvert\bar{q} - \bar{t}\rvert < 0.5$	300	500
$q \neq t$ and $\lvert\bar{q} - \bar{t}\rvert > 0.5$	210	210

Table 3: Payoffs to elicit beliefs about relative biases.
$q = t$ ($q \neq t$) refers to the self-assessment of a subject in Treatment Q or S, respectively, while $|\bar{q} - \bar{t}| \gtrless 0.5$ refers to the average self-assessment of the *A subjects*. In Treatment S, subjects evaluate a single *A subjects* and not the group average, thus, $|\bar{q} - \bar{t}| \gtrless 0.5$ has to be replaced by $q = t$ and $q \neq t$, respectively.

5.1 Experimental Design

To examine the aforementioned question, we include an additional task in Treatments Q, S and W. In Treatment W, we explicitly ask subjects whether they think that "I and others made the correct choice" (or "both are wrong", "others right and I am wrong", "others wrong and I am right"). If subjects are right with their statement, they receive 400 Tokens, otherwise 50. Hence, we can see from a subject's choice in Treatment W whether he thinks that he is rather biased himself than are the others or whether he thinks both are equally likely (un-)biased.
In order to elicit such beliefs in Treatments Q and S, subjects choose between two alternatives (I and II), whose payoffs for the four possible events are shown in Table 3. Payoffs of the two alternatives only differ for the event that a subject is correct himself, while the *A subjects* are wrong (on average), or when the subject is wrong but the *A subjects* are correct (on average). In case that both are correct or both are wrong the two alternatives lead to identical payoffs. Combining the choice between these two alternatives with the statement whether a subject thinks that *A subjects* are under-, overconfident or unbiased one can, nevertheless, conclude whether he thinks that both make the right or wrong decision or only one is wrong while the other one is right.

5.2 Results

Figure 5 shows the percentage of subjects in Treatments Q and W (for the difficult questions) thinking that oneself does not make a mistake and the *A subjects* are biased, that oneself and the *A subjects* are correct, that oneself makes a mistake while the *A subjects* are unbiased or that both are wrong. In both treatments, it is the largest group of subjects that thinks it is more likely that they do not make a mistake themselves, while the others are biased. About two thirds of the subjects think that they and the others or only they themselves are unbiased. Thus, familiarity – the framing – does not make subjects recognize that overestimation is a problem for themselves.

The result is striking in the following respect: If one thinks that the population is unbiased on average, one should rather state that the others are correct because it is more likely that oneself – as a single subject – is wrong than the whole population (as random mistakes cancel out for the population but not for a single subject). As we have seen in Section 4, subjects are aware that single subjects make random mistakes. Thus, subjects do not believe that they – compared to others – make a mistake at all. This indicates a strong better-than-average effect (or self-serving bias): Individuals do not only think that they have better relative abilities (as psychologist observe), but also that they are better in judging their own ability than are others in this task (even if for others random mistakes cancel out). This strong self-serving bias can also be the driving force behind the observation that familiarity does not help subjects to recognize that they are biased themselves: Subjects might simply ignore the information included in the framing that indicated that they – and not only the others – might be wrong.

When subjects face a single *A subject* in Treatment S, the result is similar: A majority of subjects states that rather they are correct themselves than is a single *A subject*, i.e. they choose Alternative I. We plot the percentages of *S subjects'* choices of Alternatives I and II for every single belief of an *A subject* in Figure 6. More striking in Treatment S, however, is the observation that for low stated beliefs q of an *A subject*, *S subjects* tend even more to Alternative I (indicating that they rather think they are correct themselves) than for high beliefs of an *A subject*.[13] Thus, *S subjects* are more confident in those subjects who appear to be good types (i.e. who state a high belief about themselves).

Earlier, we analyzed the relation between a subject's own bias and the belief about the bias of others. We discussed, how a self-correction mechanism explains the results. Now, we can look at the relation between a subject's own bias and the belief about the relative bias. The analysis of the relation provides again evidence that subjects try to correct their choices if they are aware of biases: In Treatment Q, subjects who choose Alternative I – which indicates that they think they are rather correct themselves – have a significantly stronger bias than those choosing Alternative II (Mann-Whitney U test, $p = 0.007$). The pattern for Treatment W is similar. Here, those who say that they are more likely correct than are the *A subjects* have on average a higher bias.

[13] According to a Wilcoxon signed-rank test, the medians of the number of subjects choosing Alternative I/Alternative II when q is lower than 4 or at least 4 differ significantly ($p = 0.05$, two-sided).

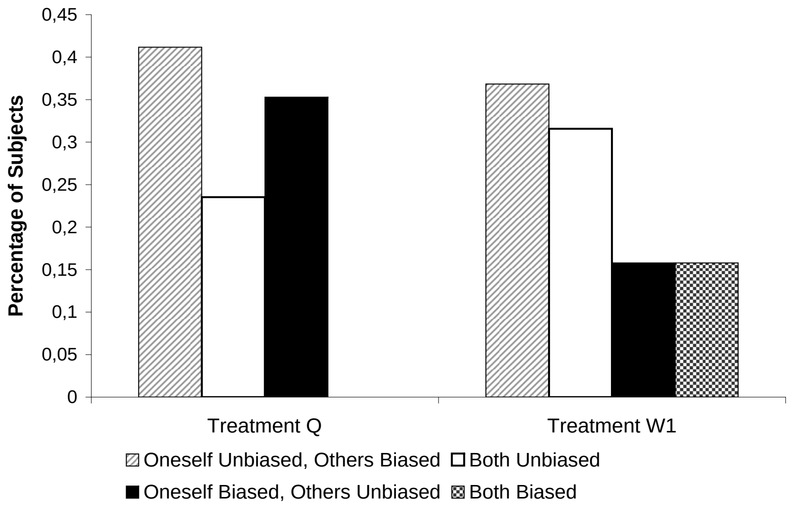

Figure 5: Beliefs about relative biases. Treatment W1 refers to the part of Treatment W with the difficult questions.

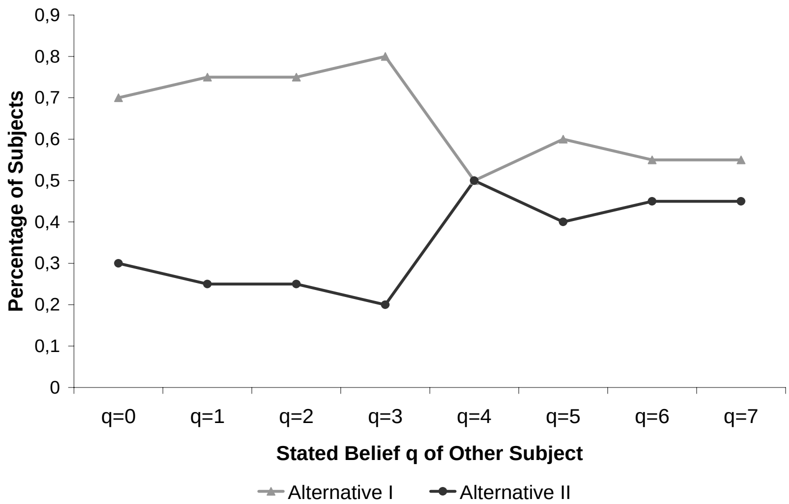

Figure 6: Average choice of alternatives I and II in Treatment S. Alternative I/ Alternative II indicates that a subject thinks it is rather correct himself/rather the others are correct.

6 Discussion

In the following, we discuss the selection of questions in our experiment, which is a possible drawback, the procedure to measure the overconfidence bias, and applications of our results to organizations and markets.

6.1 Choice of Questions

The issue that the type of questions matters for overconfidence has been intensively investigated in the literature with different results. A well-known result is the hard-easy effect. Lichtenstein and Fischhoff (1977), for instance, show that with easy questions overconfidence vanishes and even turns into underconfidence. Gigerenzer (1993) claims that the type of questions does not matter, whereas it matters whether questions are randomly selected or not. If they were selected randomly, overconfidence would vanish. Among others, Brenner et al. (1996) show that this is not true.

What is the implication of this discussion for our experiment? As we have seen overestimation is more pronounced with the tricky questions. Obviously, this is caused by the questions and thus, one cannot claim that subjects have a stronger bias with the tricky questions. In the text, we did not deepen this discussion as our focus is not whether subjects are more or less overconfident with tricky questions. Instead, we use the tricky questions just as a means to be able to provide subjects with a strong signal (by showing them the correct answers) that others might be wrong with their assessment. Thus, in this respect the selection of question discussion has no implication for our experiment.

From another perspective, however, the discussion matters. If Gigerenzer (1993) is right with his critique, then the B, Q, W and S subjects' beliefs about how we selected the questions are important: Do they think we selected them in a special way or randomly? Thinking that we randomly selected the questions could cause their belief that others are unbiased (being aware that overconfidence vanishes in case of a random selection of questions). A study by Frederick (2005) (which deals, however, with a different topic than ours) makes clear that this can be an issue. In this study, subjects face questions that induce "intuitive mistakes". This means that the answer that comes first to ones mind is wrong. Those subjects, who correctly answered the questions, recognized that the questions were not as easy as they looked like. Hence, the latter subjects think more often that others are wrong than do those who did not answer the questions correctly. However, such reasoning is harder with our questions as they do not induce intuitive mistakes: As we have seen, there is no significant difference between Treatments B and Q – showing that when subjects think longer about the questions this does not make them significantly more often recognize that others are biased. Moreover, the significant differences between Treatments B and W speak against the explanation that the subjects' beliefs about the rationality of others is shaped by their belief about how we selected the questions: If subjects believed that we selected the questions randomly and this causes unbiasedness, we should see no difference between these treatments.

6.2 Overconfidence

The literature applies several definitions of overconfidence. On the one hand, one can define overconfidence about own knowledge or ability or one can define it as being too optimistic regarding the own performance ("optimistic overconfidence"), which does not necessarily depend on own knowledge. On the other hand, overconfidence can refer to absolute abilities as well as to relative abilities, i.e. people make assessments either regarding their own ability or regarding their ability compared to other people's abilities (like estimating their rank or percentile in a distribution). In our experiment, we deal with overconfidence about the own absolute ability. We choose this for the following two reasons. First, optimistic overconfidence would have been harder to measure in the laboratory. Second, we choose absolute and not relative abilities because if people were even not aware of this simple bias, the awareness of other notions of overconfidence (like relative biases) is even less likely.

Besides several definitions of overconfidence, there are also different ways to measure overconfidence. Much of the evidence for overconfidence comes from calibration studies by psychologists, in which subjects make probability judgements, e.g., about their answer to a question being correct. In contrast, we let people make a point estimation. First, we think that this is easier to understand than a calibration task (especially for the subjects in Treatments B, Q, W and S). Second, and more importantly, with this task, we can distinguish between random mistakes and real biases.

6.3 Applications and Directions for Future Research

The finding by psychologists that people are overconfident inspired economists to include such a bias also in their models: what is the optimal incentive contract for an overconfident agent (de la Rosa, 2007; Santos-Pinto, 2007*a*), how do such agents behave in tournaments (Santos-Pinto, 2007*b*) and team production (Gervais and Goldstein, 2004), or how do they compete in an oligopolistic market (Eichberger et al., 2007; Englmaier, 2004). All these models involve strategic interactions – something we have not considered in our experiment. The reason for this design is simple: Without strategic interaction, subjects' decisions are fully driven by their belief about the bias of others and not by strategic considerations. While adding strategic interaction makes this impossible, it adds other interesting questions: Do subjects exploit the bias of others strategically (given they know their opponents are biased), or do they violate the "no agree to disagree assumption" (I think I am better than you, you think you are better than me).

Furthermore, our results gave quite pessimistic predictions regarding the knowledge about other people's (e.g. an opponents') bias – an assumption that is, however, met in the theoretical models above. Thus, it seems an important direction for future research to examine whether our results would be replicated in a strategic context.

In Section 3.1 and 3.2, we have seen the presence of a self-correction mechanism: e.g., those subjects, who state that others are overconfident, are underconfident themselves. This indicates that at least some people are to some extent aware that they have a bias – for which they try to correct (as they might not know the size of their bias, they might overreact). Our experimental design does not allow us to examine what people really know about their own bias, which is, however, an interesting question for future research.

Finally, it seems important to examine further the implications of un-/awareness of other people's bias for organizations and markets. For example, in firms, overconfident managers can have a positive (Goel and Thakor, 2005) or negative (Malmendier and Tate, 2005; 2006; Malmendier et al., 2006) impact on the firm's performance depending on the dimension one considers. Thus, the question arises, whether managers are selected systematically with respect to their overconfidence bias or whether shareholders are simply not aware of this bias and its possible positive or negative implications for the firm.
Another area where the knowledge about other people's bias matters are university admissions or career centers. It is, for example, well-known that men are more overconfident than women (see e.g. Barber and Odean, 2001). In many countries significantly less women hold a university degree than men and overestimation of abilities can be one explanation for this observation. Hence, an interesting issue is whether people responsible for university admissions and teachers are aware of the fact that men are more overconfident and whether they try to correct for it.

7 Conclusion

Overconfidence is an everyday life phenomenon: People overestimate their driving abilities, students their scores in exams or their position in the distribution or ranking, couples the likelihood of not getting divorced, and portfolio managers their prediction abilities. As overconfidence is such a common characteristic of people's behavior and is observed so frequently in real life, one expects that people are also aware of this bias. While we confirm that overestimation of the own ability is a prominent phenomenon, we nevertheless find that a majority of subjects does not think or know that others have a bias. We examine some explanations for this behavior: the lack of familiarity with the task, the distinction between real biases and random mistakes, and the relation between the belief about one's own ability and bias and the belief about other people's biases. Concerning familiarity, we observe that making individuals more familiar with the task increases the awareness, but still they do not recognize the full extent of the overconfidence bias. A better explanation might thus be that subjects are well aware that others make mistakes when evaluating their ability – they think, however, that these mistakes cancel out on average, i.e. are random. Finally, we saw that also the belief about oneself matters for the belief about others: We discussed how the results are driven by a better-than-average effect (i.e. a self-serving bias), a self-correction mechanism and a false consensus effect.

Appendix

We show that the payoffs in our experiment induce a subject that maximizes some increasing utility function $u(\cdot)$ to choose the alternative which he thinks is most likely to be correct. Before we show this, note that all decision problems in our experiment have the same structure: A subject has the choice between several alternatives ($J = \{2, 3, 4, 8, 70\}$). For example, a subject has eight alternatives for the statement how many questions she thinks she answered correctly. If a subject makes the "right" choice (e.g., she states the right number of correctly answered questions), she receives a high payoff and if her choice is not correct, she receives a low payoff.

An individual might be uncertain which alternative is true, and hence forms beliefs about the probabilities of the different alternatives being true. To show that a subject chooses the alternative on which she puts the largest probability to be the correct one, we first define beliefs and strategies of a subject. We let μ_j be the individual's belief that alternative j is true $j \in \{0, \ldots J\}$ with $\sum_j^J \mu_j = 1$. Given (with some abuse of notation) a pure strategy $j \in J = \{0, \ldots J\}$ of an individual, a mixed strategy, $\sigma : J \rightarrow [0, 1]$, assigns to each pure strategy j a probability $\sigma_j \geq 0$ with which j is played where $\sum_j \sigma_j = 1$. Further, we denote by c the high payoff (525, 1680, 105, 400, 500 Tokens) and by $c - \kappa$ the low one (20, 30, 315, 50, 300 Tokens).[14] We assume, without loss of generality, that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_J$.

Then, one can show the following:

Proposition 1 *Unless $\mu_1 = \mu_2$, an individual plays a pure strategy. More precisely, the individual sets $\sigma_1 = 1$ if $\mu_1 > \mu_j \ \forall j \neq 1$. If $\mu_1 = \ldots = \mu_n > \mu_{n+1} \geq \ldots \geq \mu_J$ with $J \geq n \geq 2$, then any σ with $\sigma_1 + \ldots + \sigma_n = 1$ can be optimal.*

This result implies that a mixed strategy is not optimal as long as an individual that is uncertain about the right action attaches a higher probability to one possible action than to all other actions.

Since all the decision problems in our experiment have this structure, we can apply this proposition to all of them.

Proof.

The subjectively expected utility of an individual from strategy σ is

$$\mu_1[\sigma_1 u(c) + \sigma_2 u(c-\kappa) + \ldots + \sigma_J u(c-\kappa)] + \mu_2[\sigma_1 u(c-\kappa) + \sigma_2 u(c) + \ldots + \sigma_J u(c-\kappa)] + \ldots + \mu_J[\sigma_1 u(c-\kappa) + \sigma_2 u(c-\kappa) \ldots \sigma_J u(c)].$$

[14] So called "probability matching" (see e.g. Shanks et al., 2002) could occur in our decision problem. Shanks et al. (2002) show that this anomaly occurs less often in case financial incentives are provided. Thus, "probability matching" should not be a severe problem for our experiment.

Rearranging yields

$$u(c)\sum_j \mu_j\sigma_j + u(c-\kappa)\underbrace{\left[\sigma_1\left(\sum_{j\neq 1}\mu_j\right)+\sigma_2\left(\sum_{j\neq 2}\mu_j\right)+\ldots+\sigma_J\left(\sum_{j\neq J}\mu_j\right)\right]}_{=\sum_j \sigma_j(\sum_{i\neq j}\mu_i)}. \tag{1}$$

Suppose now that subjects never put the same probability on alternatives. Without loss of generality $\mu_1 > \mu_2 > \ldots > \mu_J$. The expected utility under a strategy that sets $\sigma_1 = 1$ would be

$$u(c)\mu_1 + u(c-\kappa)\sum_{j\neq 1}\mu_j. \tag{2}$$

Compare this to a strategy σ' that puts some positive weight on *other* alternatives (i.e. $\sigma'_1 < 1$). This means, we subtract (1) from (2), where we, however, replace all σ_j by σ'_j in the latter. This yields

$$u(c)\underbrace{\left[(1-\sigma'_1)\mu_1 - \sum_{j\neq 1}\mu_j\sigma'_j\right]}_{(A)} + u(c-\kappa)\underbrace{\left[\sum_{j\neq 1}\mu_j - \sum_j \sigma'_j(\sum_{i\neq j}\mu_i)\right]}_{(B)}$$

As long as this difference is positive, the strategy that sets $\sigma_1 = 1$ is optimal. Consider term (A) using that $\sigma'_2 = 1 - \sum_{j\neq 2}\sigma'_j$:

$$(1-\sigma'_1)\mu_1 - \sum_{j\neq 1}\mu_j\sigma'_j = (1-\sigma'_1)(\mu_1-\mu_2) + \mu_2\left(\sum_{j>2}\sigma'_j\right) - \sum_{j>2}\mu_j\sigma'_j = (1-\sigma'_1)(\mu_1-\mu_2) + \sum_{j>2}(\mu_2-\mu_j)\sigma'_j.$$

This is strictly larger than zero since $\mu_1 > \mu_2 > \ldots > \mu_J$. The smallest value it can take is zero if and only if $\mu_1 = \mu_2 = \ldots = \mu_J$. Consider now term (B):

$$\sum_{j\neq 1}\mu_j - \sum_j \sigma'_j\left(\sum_{i\neq j}\mu_i\right) = (1-\sigma'_1)\sum_{j\neq 1}\mu_j - \sum_{j\neq 1}\sigma'_j\underbrace{\left(\sum_{i\neq j}\mu_i\right)}_{1-\mu_j} = (1-\sigma'_1)(\mu_2-\mu_1) + \sum_{j>2}\sigma'_j(\mu_j-\mu_2).$$

This term is (strictly) negative (the term equals zero if $\mu_1 = \mu_2 = \ldots = \mu_J$), but the absolute value is the same for the term (A) and (B). Since the first is weighted by $u(c) > u(c-\kappa)$, subjectively expected utility from the strategy setting $\sigma_1 = 1$ is larger than from σ' and hence, this is the optimal strategy.

It is easy to see that this result also holds true for $\mu_1 > \mu_2 \geq \ldots \geq \mu_J$, since $\sigma'_1 < 1$. If, however, $\mu_1 = \ldots = \mu_n > \mu_{n+1} \geq \ldots \geq \mu_J$ with $J \geq n \geq 2$, then any σ with $\sigma_1 + \ldots + \sigma_n = 1$ can be optimal. To see this, note that term (A) simplifies to

$$\sum_{j>n}(\mu_2-\mu_j)\sigma'_j$$

and (B) to

$$\sum_{j>n}\sigma'_j(\mu_j-\mu_2)$$

as $\mu_1 = \mu_n$. Consider a strategy σ' that sets $\sigma'_j = 0$ for all $j > n$ (i.e. all j for which $\mu_2 - \mu_j > 0$) and $\sum_{j \leq n} \sigma'_j = 1$. Then term (A) and term (B) would be both equal to zero under this strategy σ'. Hence, the strategy setting $\sigma_1 = 1$ yields the same expected payoff than σ'. Thus, any strategy that sets $\sigma_1 + \ldots + \sigma_n = 1$ can be optimal. ■

References

Baker, L. A. and Emery, R. E. (1993), 'When every relationship is above average: Perceptions and expectations of divorce at the time of marriage', *Law and Human Behavior,* vol. 17 (4), pp. 439–450.

Barber, B. M. and Odean, T. (2001), 'Boys will be boys: Gender, overconfidence, and common stock investment', *The Quarterly Journal of Economics,* vol. 116 (1), pp. 261–292.

Bosch-Domenech, A., Montalvo, J. G., Nagel, R. and Satorra, A. (2002), 'One, two, (three), infinity, . . . : Newspaper and lab beauty-contest experiments', *American Economic Review,* vol. 92 (5), pp. 1687–1701.

Brenner, L. A., Koehler, D. J., Liberman, V. and Tversky, A. (1996), 'Overconfidence in probability and frequency judgements: A critical examination', *Organizational Behavior and Human Decision Processes,* vol. 65, pp. 212–219.

Camerer, C. and Lovallo, D. (1999), 'Overconfidence and excess entry: An experimental approach', *American Economic Review,* vol. 89 (1), pp. 306–318.

Croson, R. T. A. (2000), 'Thinking like a game theorist: Factors affecting the frequency of equilibrium play', *Journal of Economic Behavior & Organization,* vol. 41 (3), pp. 299–314.

de la Rosa, L. E. (2007), 'Overconfidence and moral hazard', mimeo, University of Aarhus.

Della Vigna, S. and Malmendier, U. (2006), 'Paying not to go to the gym', *American Economic Review,* vol. 96 (3), pp. 694–719.

Eichberger, J., Kelsey, D. and Schipper, B. (2007), 'Ambiguity and social intreaction', mimeo, University of California, Davis.

Englmaier, F. (2004), 'A strategic rationale for having overconfident managers', mimeo, University of Munich.

Fischbacher, U. (1999), 'z-tree-zurich toolbox for readymade economic experiments-experimenters manual', mimeo, University of Zurich.

Frederick, S. (2005), 'Cognitive reflection and decision making', *Journal of Economic Perspectives,* vol. 19 (4), pp. 25–42.

Gervais, S. and Goldstein, I. (2004), 'Overconfidence and team coordination', mimeo, Duke University.

Gigerenzer, G. (1993), *The Bounded Rationality of Probabilistic Mental Models*, in K. I. Manktelow and D. E. Over (Eds.), Rationality: Psychological and Philosophical Perspectives, London: Routledge.

Goel, A. M. and Thakor, A. V. (2005), 'Overconfidence, CEO selection and corporate governance', mimeo, Washington University, St. Louis.

Greiner, B. (2004), 'The online recruitment system orsee 2.0 - a guide for the organization of experiments in economics', mimeo, University of Cologne.

Ho, T.-H., Camerer, C. and Weigelt, K. (1998), 'Iterated dominance and iterated best response in experimental "p-beauty contests"', *American Economic Review,* vol. 88 (4), pp. 947–69.

Hoelzl, E. and Rustichini, A. (2005), 'Overconfident: Do you put your money on it?', *Economic Journal,* vol. 115 (503), pp. 305–318.

Holt, C. A. and Laury, S. K. (2002), 'Risk aversion and incentive effects', *American Economic Review,* vol. 92 (5), pp. 1644–1655.

Laibson, D. (1997), 'Golden eggs and hyperbolic discounting', *The Quarterly Journal of Economics,* vol. 112 (2), pp. 443–77.

Lichtenstein, S. and Fischhoff, B. (1977), 'Do those who know more also know more about how much they know?', *Organizational Behavior and Human Performance,* vol. 20, pp. 159–183.

Malmendier, U. and Tate, G. (2005), 'CEO overconfidence and corporate investment', *The Journal of Finance,* vol. 60 (6), pp. 2661–2700.

Malmendier, U. and Tate, G. (2006), 'Who makes acquisitions? CEO overconfidence and the market's reaction', mimeo, University of California, Berkeley.

Malmendier, U., Tate, G. and Yan, J. (2006), 'Corporate financial policies with overconfident managers', mimeo, University of California, Berkeley.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E. and Venderklok, M. (1985), 'The false consensus effect: A meta analysis of 115 hypothesis tests', *Journal of Experimental Social Psychology,* vol. 21, pp. 263–283.

O'Donoghue, T. and Rabin, M. (1999), 'Doing it now or later', *American Economic Review,* vol. 89 (1), pp. 103–124.

Rabin, M. (2000), 'Risk aversion and expected-utility theory: A calibration theorem', *Econometrica,* vol. 68 (5), pp. 1281–1292.

Santos-Pinto, L. (2007*a*), ‘Positive self-image and incentives in organizations’, mimeo, Universidade Nova de Lisboa.

Santos-Pinto, L. (2007*b*), ‘Positive self-image in tournaments’, mimeo, Universidade Nova de Lisboa.

Shanks, D., Tonney, R. and McCharthy, J. (2002), ‘A re-examination of probability matching and rational choice’, *Journal of Behavioral Decision Making,* vol. 15, pp. 233–250.

Yates, J. F. (1990), *Judgement and Decision Making*, Englewood Cliffs, NJ: Prentice Hall.