My and Your Bias – What Do You Know About Them?

Sandra Ludwig and Julia Nafziger*

University of Bonn, Department of Economics, Adenauer Allee 24-42, 53113 Bonn, Germany.

sandra.ludwig@uni-bonn.de, jnafziger@uni-bonn.de.

May 3, 2007

Abstract

In this experiment, we test whether individuals are aware of other people's bias – like over- or underconfidence – and what they think about the relation between their own and other people's bias. We also try to infer whether people have some knowledge about their own bias and correct for it. More precisely, we consider people's self-assessment about their number of correct answers when answering a set of multiple choice questions. Our results confirm that people tend to overestimate their ability, i.e. the population on average is biased. Nevertheless, we find that most individuals do not think that others have a bias. Further, people seem not to be aware of their own bias.

Keywords: Bias, Knowledge, Overconfidence, Experimental Economics
JEL-Codes: C91, D01, D80

*We would like to thank Johannes Abeler, Ethan Cohen-Cole, Simon Gächter, Paul Heidhues, Denis Hilton, Philipp Kircher, Alexander Koch, David Laibson, Ulrike Malmendier, Felix Marklein, Burkhard Schipper, and Karl Schlag as well as participants at the Monday's Workshop in Bonn, the Brown Bag Seminar in Karlsruhe, the ESA meeting in Alessandria, the Spring Meeting of Young Economists in Seville, the ESEM meeting in Vienna, and the annual meeting of the Verein für Socialpolitik in Bayreuth for helpful comments. Financial support from the Bonn Graduate School of Economics is gratefully acknowledged. Nafziger also wants to thank the IDEI and the Nuffield College for their kind hospitality.

1 Introduction

Knowledge about other people's attributes is important in many economic situations. Imagine you hire a manager, give him a perfectly designed incentive contract, and after some time you wonder why things go wrong in your firm. The manager may have invested in too risky projects, made insensible acquisitions or hired wrong people. What went wrong – according to your incentive contract he shouldn't have done all these things! Well, maybe you did not know that your manager is overconfident. We are interested in whether people know that biases like overconfidence exist in the population.

Overconfidence can be defined and measured in different ways. On the one hand, one can define overconfidence in 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. An example of "optimistic overconfidence" is that people assess the likelihood that they get divorced too optimistically. 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 ability (like estimating their rank or percentile in a distribution). Much of the evidence for overconfidence comes from calibration studies by psychologists, in which subjects make probability judgements, e.g., that their answer to a question is correct. 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.

The fact that individuals are overconfident – in the sense that they overestimate their absolute or relative abilities – is confirmed by economists (see e.g. Camerer and Lovallo (1999) or Hoelzl and Rustichini (2005)), who also point out that the presence of overconfident individuals in economic settings has far going implications. For example, if you know that your opponent or employee is overconfident you should adjust your behavior in contests accordingly (Ando (2004) or Santos-Pinto (2005b)), write different incentive contracts (Santos-Pinto (2005a) or De La Rosa (2005)), or choose different strategies in Bertrand and Cournot Competition (Englmaier (2004) or Eichberger et al. (2005)). Malmendier and Tate (2005a, 2005b) observe that managers are indeed overconfident and that this characteristic is a disadvantage to the firm, whereas in Kyle and Wang (1997), for example, overconfidence is unilaterally beneficial.

The cornerstone of all these models is that people know that other people have a bias. In some models it is also important that people know whether they have a bias themselves and that they know about the relation of these biases (and that all this is common knowledge). Suppose you do not know that others are overconfident. Why should you behave differently in a contest if you have no idea that your opponent is overconfident? Why would you write a non-standard incentive contract if you have no idea that your manager is overconfident? Why do you hire overconfident managers in case this is a disadvantage for your firm – don't you know that they have a bias?

The aim of our experiment is to examine what people know about such biases: Are individuals aware that others have a bias in assessing their (absolute) abilities? What do they think about the relation of their and other people's biases? Are there some hints that they know about their own bias and correct for it?

Since overconfidence is a common phenomenon, we consider a bias that is one possible interpretation of overconfidence: over-or underestimation of one's absolute ability. Subjects assess their number of correct answers to a set of questions, which means they assess their absolute ability in this task. In contrast to calibration studies, we cannot claim that a subject is biased when her self-assessment is wrong, she might just make a mistake. If a group of subjects, however, tends to either over- or underestimate their abilities (i.e. their mistakes do not cancel out), we can say that the group is biased.
In order to avoid to influence people in their reasoning (and thus, their choice) by asking them the questions what they think about biases explicitly, we construct simple decision problems to elicit beliefs. Moreover, we conduct another treatment to see whether choices differ, when subjects face either these decision problems or the explicit questions. To analyze the effects of such framed instructions (i.e. asking the subjects explicitly) compared to the neutral way (i.e. the decision problems) is – besides the two questions above – another topic of our paper. So far, relatively little research in the overconfidence field considers whether asking subjects directly (as psychologist do it) changes behavior. For an overview on framing effects in other fields see Rabin (1998). Asking people directly whether others over- or underestimate or correctly estimate their abilities might cause that people become aware of problems like over- or underestimation. Therefore, subjects may adjust their beliefs or people may start to overrate the relevance of wrong estimates. This may lead to "over-adjustment" of beliefs.
The design of our basic experiment is as follows. At first, subjects in the reference treatment (R) answer seven general knowledge questions (multiple choice) – we refer to these subjects as Rs in the following. Then, the Rs choose an action, where the optimal choice depends on R*'s* belief about her number of correctly answered questions. Subjects in *another* treatment (T) are informed about the questions (not the correct answers) the Rs had to answer and the 'average action' the Rs have chosen. This average action reflects the Rs*'* average assessment of their number of correctly answered questions. Given this information, subjects in T have a choice between three actions. The chosen action reveals whether subjects in T think the Rs are either underconfident, rational or overconfident. Further, subjects in T choose a number reflecting their belief about the true average number of correct answers of the Rs.
Besides this baseline treatment we explore several extensions. In the first one, subjects in

T answer the questions themselves and assess their own number of correct answers before evaluating the Rs. This does not only give subjects a better feeling for the plausibility of the estimate of others, but also enables us to compare the own bias of a subject and the belief about the bias of others (the Rs): Do people, who are more biased themselves, also think that others are more likely to be biased or is it just the other way round? In another extension we test (as mentioned above) the impact of using a non-neutral language in the instructions. Furthermore, we consider whether subjects could be forced to recognize that the Rs are biased. To analyze this issue, we let the Rs answer very tricky questions instead of the hard ones, and subjects in T also see the correct answers to the tricky questions before they judge the Rs. These tricky questions are designed in a way to increase subject's confidence that they answered correctly but are in fact wrong with their answer (i.e. the correct answer is rather surprising). Lastly, we confront subjects in T not with the average guess of the Rs, but with single Rs. By doing so, we can infer whether subjects know that others make mistakes (these mistakes need not be systematic as they need to be to form a bias), even though they do not know that others are biased. In this treatment, we apply the strategy method to elicit the beliefs of subjects given any possible belief R can have.
Concerning relative biases, we add in several of the above treatments an additional decision problem. Here, subjects evaluate the relation of their own bias or mistakes and the average bias of the subjects in the reference treatment.

We observe that there are different types of subjects: Subjects who overestimate their number of correct answers as well as subjects who underestimate or correctly estimate it. The largest group is – with more than 50 and up to 90 percent – the group that overestimates the own ability. Our first result is that even if overestimation frequently occurs in the population (like previous studies have shown), a majority of subjects does not know that others have (on average) a bias. This result is striking as overestimation of one's own ability seems to be such a prevalent phenomenon in our experiment (and in the real world) that it should be self-evident that people are also aware of it. The more familiar subjects are with a task, however, the more subjects learn that others are on average biased. We cause this familiarity in our experiment by letting the subjects answer the questions themselves, by framed instructions (asking the subjects explicitly as explained above) or by letting them evaluate Rs who answered tricky questions and showing them the correct answers to these questions. We observe that asking subjects explicitly whether they think that others estimate their ability correctly gives subjects a hint about the existence of erroneous self-assessments: in contrast to the setting where subjects are confronted with the neutrally framed decision problem, more of them recognize that others are biased. Moreover, subjects in the framed session are less biased – indicating that the wording does not only make them recognize that others are biased, but also that they are biased themselves (for which they then correct).

Finally, when confronted with single Rs, subjects recognize that those might make mistakes. Combining our observations indicates that subjects think that Rs make unsystematic mistakes (which cancel out on average), but not that these mistakes are systematic (implying that the Rs are really biased).

An important question is how subjects make their judgement of the Rs (or a single one). In those treatments, where subjects answer the questions themselves, we see that they think that others are similar to them: if subjects think, for example, 2 is a good guess for their own ability, they also guess that 2 is the (average) number of correct answers of a single R (the group of Rs). This result can be interpreted in the way that subjects show a "false consensus bias" (see Mullen et al. (1985)): subjects' estimates of others are biased in the direction of their own belief about themselves. Even more interestingly, subjects think that similar[1] Rs are very likely to be correct with their choice. One possible interpretation of this finding is that a similar R is just a projection of the own self, i.e. subjects think about themselves that they are correct.

The largest group of subjects thinks that they are themselves more likely to judge their ability correctly than is the average population.[2] This assessment of relative biases is consistent with observations that people are overconfident in the sense that they think they are better than the average, where "better" in our case means to be less biased. Although this finding can be explained by the "better-than-average" effect – or more precisely by a self-serving bias[3] – it is surprising, since "the others" represent an average here. For this average, mistakes should cancel out (in case mistakes were just random), while for a single subject they do not.[4] Furthermore, we relate this "better-than-average" bias with the bias when assessing the own number of correct answers. The result is that subjects, who are biased in the question task, also have a "better-than-average" bias.

The main question of our paper – what people know about about themselves and others – is

[1]"Similar" subjects in the sense that R has the same belief about the number of correct answers as the subject in T has about herself. This can be seen in the treatment, where subjects are faced with single Rs and where we applied the strategy method.

[2]Some might be surprised that the largest groups thinks that others are biased, while oneself is not, did we state before that the majority of subjects does not know that others are biased. One should be careful here with "largest" and "majority". A minority (35 percent) states that they are more likely to be correct than others, but this minority still forms the largest group compared to those subjects who think that others are rational and they are biased themselves (33 percent) or who think that others and they themselves are rational (32 percent).

[3]For a general discussion on self-serving biases see Rabin (1995).

[4]Svenson (1981) conducted the well-known study showing that people think they are better drivers than is an average person. Here, it is not clear what the reference group and reference ability of agents is. In contrast to our study, the average in Svenson is not likely to be better as there are no mistakes that could cancel out.

also prominent in other fields in economics and psychology like the hyperbolic discounting model, game theory (where we especially mention beauty-contest experiments) and divorce statistics.

The hyperbolic discounting model was developed to explain time-inconsistent preferences (see, e.g., Laibson (1997) and O'Donoghue and Rabin (1999)). It is usually distinguished between people who are sophisticated, which means that they know that they have a bias[5], and people who are (partially) naive, i.e. they are (partially) not aware of their bias. Empirical evidence suggests that people are (partially) naive and not sophisticated (see, e.g., Della Vigna and Malmendier (2005)). It seems to be an open topic for future research to examine further the degree of "partiality". Although, the overconfidence bias and the hyperbolic discounting bias have many conceptual differences, we think that we contribute to this debate with our observations. Our results show that people do not even partially know that others are biased and suggest that people are not aware of their own bias.

The assumption that rationality of players is common knowledge is crucial for game theory and has been tested, for example, in so called beauty-contest experiments (see, e.g., Nagel (1995), Bosch-Domenech et al. (2002) or Ho et al. (1998)). In these experiments subjects play a game that is solvable by iterated deletion of strictly dominated strategies. Here it is interesting to observe how many iteration steps subjects are typically perform. The number of steps depends on a subject's own depth of thinking, what she knows about the depth of her opponents ("their bias"), the relation between the two ("relative bias") and that all of this is common knowledge. From observing the choice of a subject, however, it cannot be fully disentangled for which reasons this choice is made in beauty-contests: Is it her own limited depth of reasoning or that she thinks the others do – on average – not think as many steps ahead as she does or that the others do not know that she thinks so many steps ahead (and she either knows this or not)? Thus, we cannot unambiguously conclude from beauty-contests what people think about other people's reasoning (or "bias") or about the relation between one's own and the others' reasoning. As aforementioned, we consider a much more simple decision problem without strategic interaction. We are able to certainly identify individuals who are aware of other people's bias and those who are not and what people think about the relation between the own and other people's bias.

A study by the psychologists Baker and Emery (1993) suggests that people may be better at detecting "biases" of other people than biases of themselves. While individuals know quite accurately the likelihood of divorces (about 50 percent of U.S. couples who marry), they have extremely optimistic expectations assessing the likelihood that they get divorced themselves. People think that a divorce is rather unlikely to happen to them. Although our subjects

[5]They could not perfectly correct for their bias in these models because the player today and tomorrow are typically modelled as two different players. With biases like overconfidence one typically assumes in case people "know their bias" that they are uncertain about the exact size and direction of their bias and could hence not always perfectly correct for it.

have to go one step further in their reasoning, i.e. we ask subjects whether they think that others know their likelihood of divorce correctly (translated to the divorce example), some of our results are related. The finding by Baker and Emery indicates that many people think that they are "better" than – or different from – the average. This is related to our result that subjects say that similar subjects are unbiased, while other subjects might be biased. The phenomenon that people think they are better than others also arises in the second part of our study, where we examine relative biases. Here, people think that they are "better than the average" in the sense that they are less biased. Again, one can explain this result by a self-serving bias.

In a study by Frederick (2005), subjects face questions that induce "intuitive mistakes". This means that the answer that comes first to one's mind is wrong. Frederick does not aim at analyzing what subjects think about others, but on the influence of cognitive ability on decision making. Nevertheless, there is one similarity to our experiment. Subjects judge the difficulty of the questions by estimating the proportion of others who answered them correctly. Those who correctly answered the questions state that they are more difficult (as they are aware of the possible "intuitive mistakes") than do those who failed to answer correctly. This result is in line with our result that more information helps subjects to realize that others are wrong. The information in Frederick is, however, endogenous: it is only available to subjects who solved the questions correctly. Concerning our tricky questions, subjects cannot realize the trickiness (or the find correct answer) just by thinking a bit longer about the question.

The issue that the type of questions that subjects answer 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, but that 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. We do not want to add to this discussion. The tricky questions that we use are just 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.

The paper is structured as follows. In Section 2, we describe the experimental design for the treatments that deal with the question whether people know that others have a bias (or make mistakes). We first present the basic and the reference treatment, before we explain the extensions. In Section 3, we derive and discuss the theoretical predictions and present the results in Section 4. Afterwards, in Section 5, we analyze the question whether subjects are aware that others make mistakes. In Section 6, we consider the question what people think about their relative bias - first presenting the design, then the predictions and finally the results. In the last section, we conclude.

2 Experimental Design

The experiment was conducted at the University of Bonn. A total of 116 subjects participated in six sessions (with 18 to 22 participants each) - one session for each treatment (*T Average, T AveragePlus, T Frame, T Individual*) and one for each reference treatment (*R Hard, R Tricky*). Each subject participated in only one of the treatments. Note that we refer to a subject, who participated in one of the four T treatments, as "he" in the following and to one in the R treatments as "she". The experiment was programmed with the software z-Tree (Fischbacher (1999)). Subjects have been recruited via the internet by using the software ORSEE (Online Recruitment System for Economic Experiments) developed by Greiner (2004). The instructions[6] have been read out loudly before the experiment started and the subjects answered clarifying questions to make sure that they understand the experimental procedure. The wording of all but one instructions (see later) was kept neutral to avoid framing effects. We did not use terms like self-assessment, type, overconfidence, etc. which we use in the following to describe the design. Subjects could earn Tokens during the experiment, where 210 Tokens = 1 Euro. Average hourly earnings were 8 Euros.

2.1 Treatment Design - The Basics (*T Average* and *R Hard*)

In our baseline treatment, *T Average*, 20 subjects have to state whether they think "others" are on average overconfident, underconfident or rational. The "others" are 20 subjects (we call them *Rs* or she in the following) from the reference treatment *R Hard*, who answered seven very hard multiple choice questions from different fields of general knowledge. They were paid 190 Tokens for each correct answer.[7] After having answered these questions, R had to estimate her number of correctly answered questions – we denote this estimate by $q \in \{0, 1 \ldots 7\}$ – without knowing her true number of correctly answered questions $t \in \{0, 1 \ldots 7\}$. The resulting payoff, $\pi(t, q)$, from her estimate q depends on whether her guess is correct, i.e. equal to her true number of correctly answered questions questions, or not correct:

$$\pi(t, q) = 525 - 495\ \mathbf{1(t \neq q)}$$

where $\mathbf{1}(\cdot) = 1$ if and only if $t \neq q$ and 0 otherwise. This means that R is punished if she over- or underestimates her number of correctly answered questions t.[8] As we will show

[6]Instructions are in the appendix (translated from German).

[7]Subjects are free not to give an answer at all, which leads to a payoff of zero for this question.

[8]One might wonder why we did not punish more, the larger the deviation of an estimate q from t is. The answer is that with risk averse subjects, one can then no longer be sure whether they choose the number q

	Correct Answers (t)							
	0	1	2	3	4	5	6	7
Action 0	525	30	30	30	30	30	30	30
Action 1	30	525	30	30	30	30	30	30
Action 2	30	30	525	30	30	30	30	30
Action 3	30	30	30	525	30	30	30	30
Action 4	30	30	30	30	525	30	30	30
Action 5	30	30	30	30	30	525	30	30
Action6	30	30	30	30	30	30	525	30
Action 7	30	30	30	30	30	30	30	525

Table 1: Payoffs - How Many Questions Do You Think You Have Correct?

later, her estimate q should be equal to her belief about t (i.e. the t she considers as most probable). When answering the questions, she knows that she has to make a decision later on, where her payoff depends on her number of correctly answered questions. She does not know yet, however, the task and the relevant payoff table. With this procedure, we avoid that *Rs* try to game the experiment by deliberately giving wrong answers (e.g. by giving no answer at all) to be able to make the correct guess.[9]

We did not ask the *Rs* explicitly what they think how many questions they have answered correctly for not influencing their choice. Instead we let them choose between eight actions and show them the corresponding payoffs in a payoff table (see Table 1). From this payoff table one can easily infer that it is optimal, for example, to choose "Action 3" if one thinks it is most probable that one answered three questions correctly.

In the instructions for a subject in *T Average* (he), we explained him what the *Rs* had to do, how they were paid for this and we also showed him the multiple choice questions (without indicating the correct answers). In order to elicit whether he thinks she is over-, underconfident or rational on average, we told him the average q (average estimate of number of correct answers) of the *Rs* rounded to one decimal place, which is denoted by $\bar{q}$ in the following. Similarly, we denote by $\bar{t}$ the average t (average true number of correct answers), however, he is not told $\bar{t}$. Then, he has to state whether he thinks that $\bar{t}$ is smaller than $\bar{q}-0.5$ (which means thinking the *Rs* are overconfident), that $\bar{t}$ is between $\bar{q}-0.5$ and $\bar{q}+0.5$ (*Rs* are rational) or that $\bar{t}$ is larger than $\bar{q}+0.5$ (*Rs* are underconfident). By adding/subtracting 0.5 we capture rounding effects and small mistakes which remain on, even though the *Rs*

that equals t if and only if they think their true number is t.

[9] As we pay subjects for each correctly answered question and for their estimate, this problem should be alleviated, but we wanted to avoid such motivations completely. In fact, all subjects gave an answer to all questions.

	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: Payoffs - Are The Others Biased?

are rational on average. A subject in *T Average* receives 1680 Tokens if he is correct, which means he states that the *Rs* are overconfident (underconfident or rational, resp.) when they are indeed, otherwise he earns 315 Tokens. Note that we did not ask a subject in *T Average* explicitly whether he thinks that the *Rs* are underconfident, overconfident or rational, but gave him the choice between three actions $left$, *middle* and *right*. It could be inferred from the payoff table (see Table 2) that it is optimal to choose, for example, action *middle* if one thinks that the *Rs* estimated the number of questions correctly.
Moreover, a subject in *T Average* states how many questions he thinks the *Rs* answered on average correctly, i.e. how large he thinks $\bar{t}$ roughly is. For this statement, he chooses a number z out of the set $\{0, 0.1, 0.2, \ldots, 6.9, 7\}$. Of course, he could only choose a number smaller than $\bar{q} - 0.5$ if he stated before that the *Rs* are underconfident and correspondingly if he stated that they are overconfident or rational, respectively. He receives 105 Tokens in case his guess z of the average number of correct answers $\bar{t}$ is almost perfect – which means that the distance between his guess z and the true average $\bar{t}$ is smaller than 0.5 – and 20 Tokens otherwise.
For the estimation of $\bar{t}$ we implemented a similar procedure than before: We did not ask "How large do you think is $\bar{t}$?", but let subjects choose a number and let them infer from the payoffs what this choice means.

2.2 Treatment Design - The Extensions

In this section, we describe all other treatments that extend the baseline treatment *T Average* in various ways.

2.2.1 Impact of Answering The Questions Oneself (*T AveragePlus*)

We are interested in the question whether a subject's belief about the *Rs* being underconfident, overconfident or rational is influenced, when he answers the questions himself and estimates his own true number of correct answers. By completing these tasks he might get a better feeling for the difficulty of the questions and whether the average guess of the *Rs* is realistic. Therefore, in the treatment *T AveragePlus*, 17 subjects answered the same multiple choice questions, estimated their number of correct answers and stated whether they think that others are under-, overconfident or rational.[10]

2.2.2 Hard versus Tricky Questions (***R Tricky*** and ***T Frame***)

With another treatment, we want to test whether some form of feedback helps subjects to recognize that others are biased. To test this, we first conducted the treatment *R Tricky*, which is identical to *R Hard*, except that these new 20 *Rs* answered different multiple choice questions. Instead of the hard ones, we selected "tricky" ones, i.e. questions that look very simple, but are in fact very difficult: subjects are quite certain that they choose the right answer, but actually select the wrong one.
Subjects in the treatment *T Frame* (where we use non-neutral wording in the instructions; see the next subsection) answered both the hard and the tricky questions and performed all the tasks as subjects in *T AveragePlus*. They had, however, not only to judge the *Rs* in *R Hard* but as well those in *R Tricky*. In addition, they state how many questions they think the *Rs* (in *R Hard* and *R Tricky*, respectively) have on average correct by choosing a number z for subjects in both *R* treatments. Note that we want to highlight the trickiness of the questions to see whether subjects are forced by this kind of information to recognize problems like overestimation. Therefore, we showed subjects in *T Frame* the correct answers to the tricky questions before they assessed whether the *Rs* are under-, overconfident or rational (but of course *after* they answered the questions). In order to avoid hedging effects, we randomly selected for the payment one block of questions (either the hard or the tricky ones) and one block of decisions (corresponding to the hard or tricky questions) after subjects finished all decisions.

2.2.3 Impact of Framing (***T Frame***)

Psychologists generally use a non-neutral language in their experiments. We want to see whether such framing[11] has some impact on our results. Subjects might think differently about a problem when they read the word "overestimate" instead of "action *right*". By reading the word "overestimate" a subject might get an idea that overestimation is a problem (why else should he read this word in the instructions?). To analyse the impact of the wording, we "framed" the instructions in *T Frame*. The main differences are as follows: in *T Frame* we explicitly asked subjects "How many questions do you think you have correct?",

[10] These 17 subjects did, however, not state z, i.e. how many questions they think R answered on average correctly.

[11] The term framing should not be misleading here. We just mean by it that we use a non-neutral language which hints at problems like a wrong self-assessment.

while in all other treatments we let them choose between eight actions.[12] Furthermore, subjects had to state whether they think that the *Rs* under- or overestimate the true number of correct answers or estimate it correctly. In the treatments with neutral instructions, however, subjects choose between three corresponding actions $left$, $right$ and $middle$. Similarly, for the statement of the belief about the *Rs'* average number of correct answers, we explicitly asked in *T Frame* "How many questions do you think the others answered on average correctly?", while in the neutral treatments we let subjects choose a number z and they have to infer the meaning from the payoffs.

2.2.4 Single Subject versus The Average (*T Individual*)

Does it make a difference whether the "others" represent the average of the *Rs* or a single *R*? As we explain more precisely in the next section, in theory it does: For a single subject one cannot distinguish by observing the guess q and the true number t of correct answers whether she makes just an unsystematic mistake or is really biased if the numbers differ. By observing the average numbers $\bar{q}$ and $\bar{t}$ of a group of subjects, however, one can conclude that these subjects are on average biased if the averages differ.

From the treatments described above, in which subjects face the average of the *Rs*, we can infer whether subjects think others are on average biased or not, but we cannot infer whether they think others make unsystematic mistakes. Yet, we are also interested in whether subjects think that *Rs* just make mistakes but are not biased, do not even make mistakes, or that they are biased (i.e. mistakes are systematic). Regarding the first case, for example, subjects would be aware that a single subject might make mistakes, but that these cancel out for a group of subjects. Thus, the group is unbiased.

To analyze the issue, we additionally conducted the treatment *T Individual*, in which subjects state beliefs about single *Rs* and not the complete group. In *T Individual*, 20 subjects have to perform all the tasks subjects in *T AveragePlus* have to. A difference to *T AveragePlus* is that we implemented the strategy method in *T Individual*. Thus, subjects do not receive specific information about a single *R*, but they state for every possible estimate $q \in \{0, 1, \ldots, 7\}$ of *R* whether they think she is under-, overconfident or rational (strictly speaking: makes mistakes or not – see Section 3.1). For the numbers 0 and 7 on the boundary, subjects only choose between the two appropriate possibilities. In case he thinks *R* 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 *T Individual* was not paid for all his decisions, but for his decision when facing a particular estimate q of an *R*. For

[12] Note that subjects in the *R Treatments* faced a different decision problem when estimating their number of correct answers as they face a payoff table and have to infer the meaning. This decision problem was explained to the subjects in *T Frame* and we made clear that it meant the same as the question "How many questions do you think you have correct?".

	R Hard	*R Tricky*	*T Average*	*T AveragePlus*	*T Individual*	*T Frame*
Hard Questions	√			√	√	√
Tricky Questions		√				√
Estimate Own t	√	√		√	√	√
Others Biased?			√	√	√	√
Relative Bias?				√	√	√
Guess z of $\bar{t}$ (t)			√		√	√
Info about $\bar{q}$			√	√		√

Table 3: Overview of the Tasks in the Treatments and in which Treatment Subjects Receive Information about the *Rs'* Average Belief $\bar{q}$. "Hard/Tricky Questions" means that subjects answer the hard/tricky questions.

$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$
Questions	Estimate Own t	Relative Bias?	Info about $\bar{q}$	Others Biased? & Guess z

Table 4: Timing

his payment, one R was randomly selected. Her q and t – together with his decision when facing her estimate q – determined his payment. Again, we did not ask all these questions directly, but confronted subjects with simple decision problems to infer their beliefs.

In Table 3, we provide an overview of all the tasks subjects have to complete in each treatment. We also indicate, in which treatment we inform subjects about the average estimate $\bar{q}$ of the *Rs*. In Table 4, we list the timing of the single tasks and when we inform subjects about $\bar{q}$. Since not all stages are present in all treatments, the corresponding timing of a treatment follows by skipping the missing stages. In these tables, we show already the "relative bias" task that is explained in more detail in Section 6.

3 Predictions

3.1 Definitions and Assumptions

For the theoretical predictions of our experiment, we need some weak assumptions and definitions about the players' behavior. We assume that individuals are subjective expected utility maximizers, with a strictly increasing utility function, i.e. they prefer more money compared to less.[13]

[13]This seems reasonable since here are no concerns for concepts like fairness. Note that we make no assumptions regarding the curvature or differentiability of the utility function. Thus, we could – by an

Next, we define what we mean by an under- or overconfident individual – i.e. a *biased* individual – or by a rational individual. Biased means that an agent's self-assessment is wrong – he *systematically* under- or overestimates his number of correct answers and is thus under- or overconfident, respectively. By systematically we mean that the mistakes an individual makes when estimating her ability are not random, in the sense that they do not cancel out on average. A rational agent in contrast makes on average no mistakes.[14] Thus, we can identify whether a population of individuals is rational – if they were rational, then $\bar{t} = \bar{q}$ (roughly) holds.[15] If they were not rational, then $\bar{t} \neq \bar{q}$. In the latter case, we define the bias as $b := \bar{t} - \bar{q}$ and say that a population with $b < 0$ is overconfident (or overestimates its ability, i.e. the true number of correct answers) and one with $b > 0$ is underconfident (or underestimates its ability). For a single individual, however, we cannot infer from observing her t and q that she is biased or not, since she could have made only an unsystematic mistake ($b < 0$: negative mistake, $b > 0$: positive mistake).

Note that in *T Individual*, we ask subjects whether they think that a single subject R is right with her self-assessment. Thus, we can in general not conclude from *T Individual* whether subjects think that a single R is biased or not, but only whether it makes (systematic or unsystematic) mistakes or not.[16]

Our main interest is whether subjects think others are biased or not. Hence, in most treatments, we consider averages over the beliefs q. Nevertheless, we want to know whether subjects think others make mistakes and thus discuss most results of *T Individual* separately in Section 5.

3.2 Eliciting Beliefs

In the following, we look at individuals' choices and explain how these mirror a subject's beliefs. In our experiment, all decision problems the individuals face 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. Of course, an individual might be uncertain which alternative is true, and hence

appropriate definition of the reference point – also think of the utility function as a value function in the spirit of Kahneman and Tversky (1979) to capture concepts like gain-loss utility.

[14]Statistical: A *rational* individual estimates that her type is $\mathrm{E}[t|\xi]$, when her true type is t with $\mathrm{E}[t|\xi] = t - \varepsilon$ and ξ is the available information of the individual. Hence, $\mathrm{E}(\varepsilon) = 0$, i.e. a rational individual makes on average no mistakes. Assuming that across individuals ($i \in I$) the ε_i's are uncorrelated random variables, one could apply the weak law of large numbers to see that $\lim_{I \to \infty} \frac{1}{I} \sum_i \varepsilon_i = E(\varepsilon) = 0$. For a biased individual $E(\varepsilon) = b \neq 0$, i.e. on average it makes mistakes.

[15]In the experiment, we allow for small deviations from $\bar{t} = \bar{q}$ for a rational group.

[16]In Section 5 we also try to infer whether subjects think mistakes are systematic or not.

forms beliefs about the probabilities of the different alternatives being true. We show in the appendix that *an individual chooses the alternative on which she puts the largest probability to be the correct one* (Proposition 1 in the appendix).[17]

This proposition implies that subjects should state the number of questions they think they answered most likely correctly. As explained above, the *average* values of stated and true number should not differ much in case individuals are roughly rational and only make random mistakes. When individuals tend to be biased in a certain direction (i.e. either over- or underconfidence), however, these numbers differ even on average.

3.3 Hypotheses on the Beliefs About the Bias of the Average

Based on previous studies by psychologists and economists (see introduction), we predict the following.

Hypothesis 1 *Subjects overestimate their abilities on average – more so with the tricky questions.*

The fact that subjects overestimate their abilities and that the degree of overestimation depends on the type of questions is a well known result from psychology. In the psychological literature on overconfidence the so-called "hard-easy" effect arises: People have been found to be underconfident for "easy" questions and overconfident for "hard" ones (see, e.g., Juslin (1994)). Our tricky questions are designed in a way that provokes more negative mistakes: Subjects are more sure that they selected the right answer but, in fact, this answer turns out to be wrong. This means that confidence rises and the number of correct answers decreases compared to the hard questions. This effect of such "surprising" questions is also addressed in Juslin (1994).

Whether one can say that subjects are more overconfident with the tricky questions depends on the way one defines overconfidence. On the one hand, one can simply say that a population is more biased (here: overconfident) if and only if the absolute value of their bias is larger – i.e. $\bar{t} - \bar{b}$ is smaller – the statement is correct. On the other hand, one can argue that subjects are not more biased because they really *are* more biased, but because the tricky questions *make* them more biased, i.e. subjects only seem more biased (see Brenner et al. (1996)). We do not deepen this discussion as our main point is not the influence of the tricky questions on the degree of overconfidence – instead, we want to see whether subjects can be induced by these questions to recognize that *Rs* are overconfident. Whenever we say in this

[17]So called "probability matching" (see e.g. Shanks et al. (2002)) could occur in our decision problem. Suppose the majority of subjects chooses the action "the others are biased". Then also if probability matching happens the results should imply that subjects put the largest probability on this action (similar for the other tasks). Shanks et al. 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.

context that overestimation is more pronounced with the tricky questions, we do not want to claim that these subjects have a stronger bias.
Under the assumption of a symmetric distribution of mistakes, one could for instance use a Wilcoxon Test (to test the hypothesis that the difference between q and t has median value zero given the pairs (q_i, t_i) of individuals i) for testing whether subjects are rational or biased.

Proposition 1 (see appendix) also implies that whenever a subject believes that the *Rs* are more likely to be either over-, underconfident or unbiased, he also states this when asked for his assessment. Further, his guess of $\bar{t}$ (the average of the true number of correct answers) should be the number that he thinks mirrors $\bar{t}$ most likely. Therefore, for agents who are uncertain between positive and negative biases or between different sizes of biases (including positive, negative and zero biases), we interpret their choice as reflecting what they think to be most likely true (and say sometimes for simplicity "they think", without the most likely). Obviously, we cannot distinguish between agents who are certain or uncertain about their statement being true.
We predict that at least some subjects know that the population is biased. A priori, it is not clear whether more subjects think that *Rs* are biased or more of them think that they are rational. From experiments and field evidence about hyperbolic discounting we know that some individuals are only "partially naive" and not fully naive (see e.g. Della Vigna and Malmendier (2006)). Partially naive means that they know their *own* bias to some extent. In case people know that they are biased themselves, there is some chance that they also know that others are biased. Note that in our experiment a wrong guess could just be a mistake and not a bias, while hyperbolic discounters are always biased. Thus, subjects in our experiment might be aware that people make mistakes (un-, or systematic), but many might expect mistakes to cancel out on average.
Concerning our different treatments we make the following prediction:

Hypothesis 2 *The more information subjects receive about the problem (no information in T Average, answering questions themselves in T AveragePlus, seeing the correct answers and framed instructions in T Frame), the more subjects state that others are biased.*

This hypothesis seems evident in the (theoretical) sense that subjects, who can use more information, can update their beliefs and thus, make better decisions. Experimental studies on whether subjects update information according to Bayes' rule, however, rather provide evidence that subjects are not "perfect Bayesians" (e.g. Kahneman and Tversky (1972) or Zizzo et al. (2000)). Nevertheless, we think that in our experiment, more information works in the stated direction. Subjects are forced to reason better how realistic it is that the *Rs* have on average $\bar{q}$ questions correct, once they answered the questions themselves and recognize that it is indeed very hard to give so many correct answers. This effect is reinforced when they see the correct answers of the tricky questions – here they could recognize that

these tricky questions induce overestimation.
Effects of better reasoning on decisions are for instance explored by Croson (2000). She finds that the frequency of equilibrium play in prisoner's dilemma and public good games increases when first subject's beliefs about the actions of others are elicited before the game is played. It is a priori not clear whether the effect of framing is stronger or weaker than the one of answering the questions oneself. Nevertheless, we think that there is an effect – reading words like "overestimation" gives subjects a hint that such things could occur. Hence, we predict that more subjects state that the *Rs* are biased.

When thinking about others, individuals often tend to conclude from their own behavior or own beliefs on others. This is the so-called false consensus effect, see e.g. Mullen et al. (1985).[18] We expect this effect to be crucial, when subjects judge the others. Thus, we have the following hypothesis.

Hypothesis 3 *When making statements about the* Rs *(about their bias or about their average number* $\bar{t}$ *of correct answers), this statement tends in the direction of the own behavior (own bias or guess of own number of correct answers).*

4 Results

In the following, we first discuss the results on the own bias of subjects. This refers to the first part of the experiment, the question task and the self-assessment, which is present in all treatments but *T Average*. Note that in *T Frame* we pose both types of questions and subjects have to evaluate both reference groups *R Hard* and *R Tricky*, respectively. When presenting the results, we therefore split this treatment into *T Frame Hard* and *T Frame Tricky*, where each part refers to either the hard or tricky questions and the corresponding decisions. Since overconfidence has already been extensively investigated in the psychological literature, our discussion is very brief.
We then turn to our results on the new issues – the knowledge about other people's bias and the belief about the relation between own and other people's biases.

4.1 The Own Bias (Hypothesis 1)

Table 5 shows the average type $\bar{t}$ and the average estimate $\bar{q}$ for each treatment, the difference between the two (bias) and the p-values from a Wilcoxon Test. With the hard questions the bias ranges from -0.5 to -1.4, with the tricky ones from -1.6 to -3.4. The p-values indicate

[18]Note that doing this need not be suboptimal, especially when subjects have no further information on the identity or characteristics of others. Therefore, the term "false" can be misleading.

	$\bar{t}$	$\bar{q}$	Bias	p
R Hard	2.3	3.4	-1.1	0.006
R Tricky	1.2	4.6	-3.4	0.000
T AveragePlus	2.1	3.5	-1.4	0.005
T Individual	1.7	2.7	-1.0	0.068
T Frame Hard	2.6	3.1	-0.5	0.207
T Frame Tricky	1.6	3.2	-1.6	0.001

Table 5: Reported and True Number of Correct Answers

significant differences between $\bar{t}$ and $\bar{q}$ for all treatments except for *T Individual* and *T Frame Hard*. Although not significant in *T Individual*, the average bias is -1 which is quite large. Even if the average bias of the whole group indicates overestimation, different types of individuals exist. Figure 1 illustrates the percentage of subjects that are under-, overconfident or rational (make positive, negative or none mistakes) in the different treatments. The fraction of subjects, who are overconfident (make negative mistakes), ranges from 53 to 90 percent in the different treatments. As intended, with the tricky questions the percentage of those, who overestimate their number of correct answers, is higher. Moreover, it is interesting to see that on average 75 percent of subjects in *R Tricky* focused on one and the same answer for each question. With the hard questions, in contrast, the answer that has been chosen most often for a question, has on average only chosen by 44 percent of the subjects in *R Hard*. One could take this as a vague hint that subject's confidence in an answer also increased with the tricky questions. How does the type of questions influence the true number of correct answers and the belief about it? As discussed in Section 3.3, the size of the bias cannot necessarily be interpreted as stronger overconfidence. Nevertheless, the effect of the type of questions on the true and believed number of correct answers is important. It might be that subjects recognized that these questions are tricky and adjusted their beliefs accordingly. As argued in Section 3.3, we find that in *R Tricky* the true numbers of correct answers t are significantly smaller than in *R Hard*, whereas the estimated numbers q and thus the mistakes are significantly larger in *R Tricky* (Mann-Whitney U test: $p = 0.009$, $p = 0.001$ and $p = 0.000$, respectively). This indicates that people in *R Tricky* seem *not* to recognize the trickiness of the questions. Similarly, within the two parts in *T Frame* (hard and tricky questions), true numbers of correct answers $t's$ are significantly larger and mistakes are significantly smaller for the hard questions (Mann-Whitney U test: $p = 0.026$ and $p = 0.03$, respectively).

Moreover, we are interested in the effects of framing. We observe that the estimates q are larger and that overestimation is much more pronounced in *R Tricky* compared to the framed treatment *T Frame Tricky*. The $q's$ and also the mistakes of subjects are significantly differ-

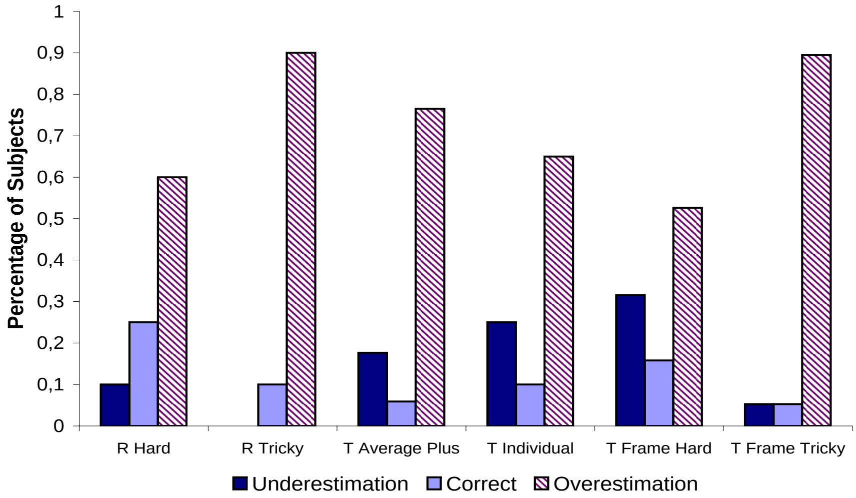

Figure 1: Percentage of Subjects Being (Not) Correct in the Treatments

ent across these treatments (Mann-Whitney U test: $p = 0.001$ and $p = 0.002$, respectively). This result is (to our knowledge) new in this context. A possible explanation for it is what psychologists call self-impression management: "[This concept] suggests that a person acts to show himself in a positive light, even when he is the only observer of his own behavior." (Murnighana et al. (2001)). Comparing the neutral with the framed treatment, subjects in the neutral treatment do not have as strong emotions when their decision turns out *not* to be optimal as subjects in the framed treatment who are forced to think of terms like self-assessment. The latter subjects feel ashamed or more stupid when they are wrong or they even do not want to appear themselves arrogant. Therefore, in the framed treatment, subjects are reluctant to make overly optimistic guesses – instead they make more realistic guesses such that overestimation is reduced.

4.2 The Bias of Others

4.2.1 What Do You Think about the Bias of Others? (Hypothesis 2)

In this subsection we analyze the subjects' perception of the *Rs'* bias.

Result 1 *Without further information, a majority of the subjects thinks that others estimate their ability correctly. The more familiar subjects are with the task or the more information they receive (answering the questions themselves, framed instructions, seeing the correct answers of the questions), the less subjects think that others estimate their ability correctly.*

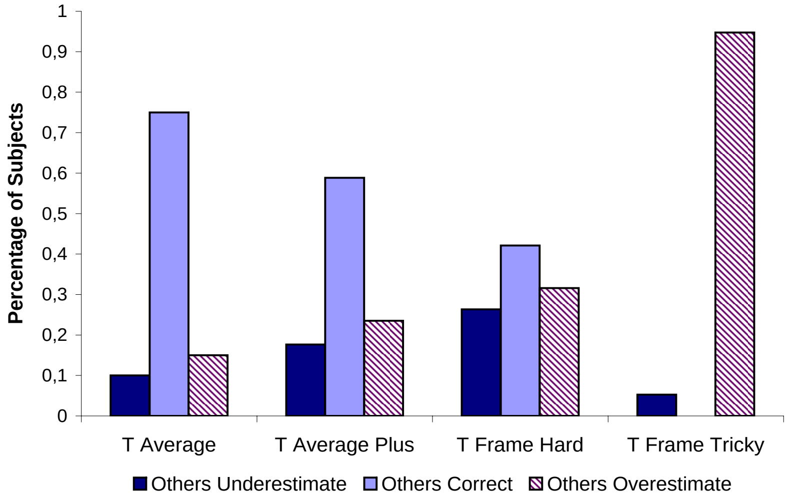

Figure 2: Beliefs About Others' Self-Assessment

This result is illustrated in Figure 2. The figure shows the percentage of subjects in the different treatments believing that the *Rs* are underconfident, rational or overconfident. Except for the second part in the framed treatment, where subjects saw the correct answers before evaluating the *Rs'* average estimate, a majority of subjects states that the *Rs* are rational. Being asked for their choice in a questionnaire after the experiment, subjects say that they made this choice because they either think the mistakes the *Rs* make cancel out on average or that the *Rs* have better information about their own number of correct answers, or that the *Rs* are simply able to make the correct choices.

Next, we explore the impact of a single piece of information. First, we ask about the effect of answering the questions oneself. Does this induce more subjects to recognize that others are biased? As aforementioned, answering the questions oneself gives subjects a better feeling for the difficulty of the task. Hence, subjects get a better impression how realistic it is that the *Rs* indeed answered $\bar{q}$ questions correctly on average as these estimate. In Figure 2, we see that the percentage of subjects, who think that the *Rs* are biased, is slightly higher in *T AveragePlus* compared to *T Average*. We cannot reject, however, the hypothesis that there is no relation between the number of subjects in the two treatments, who think that the *Rs* are rational or biased, according to a Fisher's exact test ($p = 0.234$ one-sided).

Framing and answering the questions together, in contrast, (i.e. comparing *T Frame Hard* and *T Average*) has a significant effect (Fisher's exact test: $p = 0.038$ one-sided). It increases (decreases) the percentage of subjects who think the others are biased (unbiased). Nevertheless, no significant difference arises, when only framing is added given that subjects

	Feedback $\bar{q}$	Guess z	p-value (Mann-Whitney U test)
T Average	3.4	3.2	0.002
T Frame Hard	3.4	3.4	0.138
T Frame Tricky	4.6	2.9	0.000

Table 6: Belief z about the Others' Average Number of Correct Answers ($\bar{t}$) versus the Others' Average Belief $\bar{q}$ about Their Own Number of Correct Answers (t)

answer the questions themselves: Comparing *T AveragePlus* and *T Frame* yields no significant effect ($p = 0.252$). In Figure 2, we can see, however, that the percentage of subjects thinking that the *Rs* are biased is larger in *T Frame Hard* than in *T Average*/*T AveragePlus*. This increase in the percentage of people thinking that the *Rs* are biased might be caused by the frame – by reading words like "overestimate" and "underestimate" subjects get forced to recognize that people's self-assessment might be wrong.
Does it have an effect when subjects see the correct answers to the tricky questions? This kind of feedback has a significant effect – provided with this information, almost all subjects believed that the *Rs* are overconfident.[19] Psychologists have shown (for an overview see Pulford and Colman (1997)) that feedback in form of giving the correct answers has the greatest impact on a subject's *own* bias when feedback contradicts a subject's belief most. Our result indicates that this also holds for giving feedback when evaluating others and not oneself. This is interesting since here the adjustment has to proceed in two steps as subjects conclude from their own bias on the bias of others: At first, subjects recognize that it is impossible for themselves to have as many questions correct as the *Rs* think they have on average in *R Tricky*, i.e. $\bar{q}$. In a second step, subjects conclude from their own ability that it must also be impossible for the *Rs* to answer that many questions correctly.

What do the subjects guess is the average true number of correct answers $\bar{t}$ of the *Rs* given the feedback $\bar{q}$ they receive about the others' average belief about $\bar{t}$? This means, we consider the subjects' estimate z given information $\bar{q}$ of the reference treatments (this information thus differs whether the hard or tricky questions are considered). The result is summarized in Table 6. The figure shows the average estimate $\bar{q}$ chosen by the *Rs* for the tricky and hard questions, respectively, the estimates z and the p-values from a Mann-Whitney U test – testing whether z and $\bar{q}$ are different from each other.

For *T Average*, we see that although subjects think that the *Rs* are roughly correct when evaluating their abilities, they think that the *Rs* are a little bit overconfident ($\bar{q} > z$). With

[19]Comparing *T Frame Tricky* to *T Frame Hard*/*T AveragePlus*/*T Average*, there are more (less) subjects who think that the *Rs* are biased (rational) according to a Fisher's exact test ($p = 0.0015/0.0001/0.00$ one-sided).

the tricky questions, where subjects recognized after seeing the correct answers that the *Rs* are overconfident, they adjust their estimate z of $\bar{t}$ downward to 2.9. Although this is significantly smaller than $\bar{q} = 4.6$, the estimate is still higher than the true average $\bar{t} = 1.2$. Interestingly, the estimate z is not that much smaller than 3.4, which was the subject's guess for the hard questions in *T Frame*. Thus, subjects recognize that the *Rs* overestimate their abilities, but are still not aware that overestimation is such a severe problem.

4.2.2 Why Do You Think What You Think About the Bias of Others? (Hypothesis 3)

In the last section, we already got some hints that subjects conclude from their own behavior on the behavior of others. Now, we want to investigate the reasons for the subjects' choices in more detail.

First, we have a closer look at the relationship between the bias a subject in *T AveragePlus* has himself and the belief he has about the bias of the *Rs*. The cumulative distribution functions of the average value of the bias of subjects who either think that the *Rs* are correct, overestimate or underestimate their ability is shown in Figure 3. The cumulative distribution function of those subjects who think that the *Rs* are on average correct is always below the other two functions. Hence, those subjects have less extreme (negative) biases. From the average biases, we see that in *T AveragePlus* those who think the *Rs* are overconfident have on average a bias of -2, those who think that the *Rs* are roughly unbiased have a bias of -0.7 and the rest has a bias of -3. The difference between the biases of subjects who say that others are biased and those who say they are rational are according to a Mann-Whitney U test significant (p=0.033). Moreover, in *T AveragePlus* 85 percent of the subjects having a "small" bias (larger or equal to -1) think that others are rational, while 60 percent of those who have a bias smaller than -1 (i.e. who overestimate more heavily) say that others are biased. This result is striking since we cannot directly explain it by a false consensus effect. Recall that subjects do not know how good their own self-assessment is. The result can, however, be taken as evidence that subjects have some[20] knowledge about the degree of their own bias. Subjects may conclude from their bias onto the bias of others. For instance, a subject may reason as follows: "I am rational and I know this, so the others are rational, too". These findings are summarized more generally in the following:

Result 2 *Those who think that others make on average the correct choice, make on average better choices themselves; while those who think that others are biased, make on average more biased choices. Moreover, the other way around, most subjects that are unbiased also think that others are unbiased.*

[20]If we sometimes say a subjects "knows about his bias", we mean the following: The subject knows that he is, e.g., overconfident to some extent, but he does not know the exact magnitude of this bias. Would he know the magnitude, he could perfectly correct for the bias.

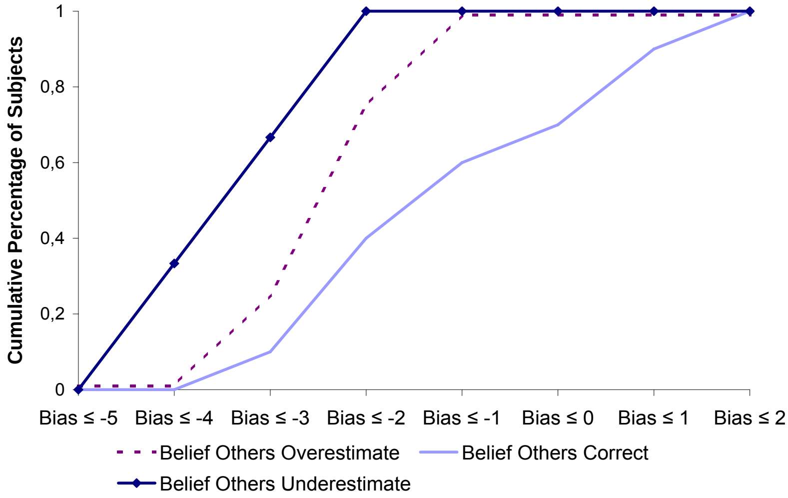

Figure 3: Own Bias given Belief About Others' Self-Assessment

This result is also striking in another aspect. It gives us some hint that the choice of subjects is not driven by a "better than the average effect" (or self-serving bias), but by their implicit self-knowledge as described above. What do we mean by "better than the average effect" in this context? After subjects estimated their own number of correct answers to be q, they learn the estimate $\bar{q}$ of the *Rs*. Thus, subjects can see whether – according to their own and the *Rs'* beliefs – they are better or worse than the average. If they think that they are better than the average but $q \leq \bar{q}$, they can simply state that the others are overconfident in order to sustain their self-image of being better than the average, as this means the others are actually worse than $\bar{q}$ and thus maybe even worse than q. There is not a large difference, however, between the percentage of those stating that others are overconfident or rational: $q \leq \bar{q}$ holds for 40 percent of the subjects stating that others are rational and for 50 (33) percent of those, who say that others are overconfident (underconfident). Hence, we find no clear evidence that subjects try to fool themselves to make them better than the average by stating others are overconfident.

In *T Frame Hard* Result 2 is slightly different. Those saying that the *Rs* underestimate their ability, have on average a bias of -1.2. Those saying that the *Rs* are roughly correct, have a bias of -0.75. And those, who say that the *Rs* are overconfident, are in fact (on average) underconfident with a bias of 0.5.[21] In *T Frame Hard* subjects, who say that the

[21]The estimate q of those subjects, who say that others are over- or underconfident, is significantly smaller than the estimate q of those, who say that others are rational (according to a Mann-Whitney U test $p = 0.004$ and $p = 0.011$, respectively).

Rs are overconfident, could be aware (and this awareness could be caused by the frame) that overconfidence not only exists in the population but also for themselves. Hence, they might adjust their choice accordingly, which leads to a less severe bias (and even underconfidence). With those, who say that the *Rs* tend to be underconfident, it is exactly the other way around (as well as in *T AveragePlus*). These subjects have the most severe bias. Thinking that they are underconfident themselves might induce them to choose an estimate q that is too high such that overconfidence arises.

How does the belief z about the (average) true number of correct answers t ($\bar{t}$) of the *Rs* relate to a subject's own stated number of correct answers q? We can analyze this issue in the treatments *T Average*, *T Frame* and *T Individual.* We find that subjects think that others have a similar (average) number of correct answers than they have themselves and – assuming some implicit self-knowledge when subjects make their choice – they think that the others are rational (or do not make mistakes). Hence, we can also see the following result as a strengthening of our interpretation of Result 2: Subjects may not only think that others have a similar bias, but also that they have a similar (average) number of correct answers.

Result 3 *Subjects think that others are similar to them, i.e. they have a belief z about the others' average ability $\bar{t}$ that is close to the belief q about their own ability. Moreover, they think that similar subjects are likely to be correct when estimating their ability.*

The first part of this result can be explained by the false consensus effect, which says that people tend to overestimate the degree to which, for example, their own behavior or beliefs are shared by other people (compare our prediction). Hence, by the false consensus effect people overestimate the frequency with which their own estimate q is present in the population. Therefore, it is likely that subjects in our experiment adjust their estimate z of $\bar{t}$ in the direction of their own estimate q – under the restriction that they think that the *Rs* are roughly rational. This is illustrated in Figure 4, which shows the average estimate z chosen by the subjects in *T Frame* given their belief q of the own number of correct answers and the information the subjects receive (i.e. the *Rs'* average belief $\bar{q}$ which is 3.4). It can be seen that subjects with lower q's (up to 3) choose on average an estimate z that is lower than 3.4, whereas subjects with higher beliefs about the own ability t (from 4 on) choose on average a higher z.
The first part of Result 3 is further supported by the following observations. The average difference between a subject's q and the chosen z is only -0.09 in *T Frame Hard* (and for the tricky questions it is still only 0.41). According to a Wilcoxon Test there is no significant difference in the median of the chosen number and the chosen action ($p = 0.647$ and $p = 0.155$). Furthermore, in *T Frame Hard* the estimates q and z are correlated: The Spearman rank order correlation coefficient is 0.737 (with $p = 0.0002$).
The second part of the result can be derived from *T Individual.* Recall that a subject in *T*

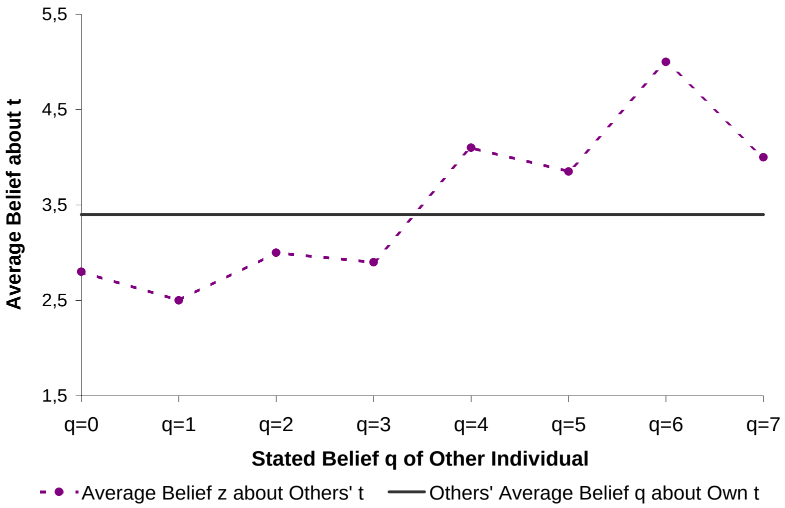

Figure 4: Average Belief z about the Others' Ability given the Belief q about the Own Ability

Individual states a belief z about an *R's* true number of correct answers t for each possible belief R can have about her t. This means that a subject in *T Individual*, which states a belief z that equals *R's* belief q, believes that this R is correct.

We compare a subject's estimate z of *R's* true number of correct answers t with this *R's* estimate q of the own number of correct answers given that R is *similar*. A similar R has exactly the same belief q about her number of correct answers than the subject in *T Individual* has about his number of correct answers. For example, an R who thinks she has three questions correct is similar to a subject in *T Individual* that thinks he has three questions correct himself. Given such a similar R, we consider the estimate z a subject in *T Individual* has about this *R's* number of correct answers. The average difference between the estimate z for a similar subject and this similar subject's belief q about his own number of correct answers is -0.05. The medians of these numbers do not differ significantly ($p = 0.476$, Wilcoxon test). Moreover, the belief about a similar individual and the *Rs* own belief are correlated (Spearman rank order correlation coefficient is 0.59 with $p = 0.003$[22]). This implies that a subject thinks that the similar subject is correct with her self-assessment. Subjects even think that similar *Rs* are likely to be correct if these *Rs* hold "extreme" beliefs, for which most other subjects say that this extreme belief must be mistaken.[23]

[22]If we exclude one subject (that always chose 0 for high q's of the other person), the numbers are even more similar to *T Frame Hard*.

[23]For *Rs* that are *not* similar and who have belief $q \in \{0, 4, 5, 6, 7\}$, the absolute values of the differences between q and z are significantly larger than than they are for similar *Rs* ($p \leq 0.02$). Dissimilar *Rs* with

What can we learn from this? Baker and Emery (1993) showed that individuals know that the average married person in their country gets divorced, but state at the same time that they themselves will not get divorced. If we replace "getting divorced" by "being biased", we get a similar result in our experiment. In case subjects make their statement because they think similar subjects are like them – not only with respect to their ability, but also with respect to their bias (see Result 2 and the first part of Result 3) – we can conclude that subjects also think about themselves that they are unbiased or do not make mistakes as they think that others are unbiased. Note again that a similar R has the same belief about her number of correct answers than has the subject in *T Individual* about himself. If this is true, it also implies that we can explain the second part of Result 3 by a false consensus bias: Subjects conclude from their own beliefs on others. If they think they are correct, then also a similar individual is correct.
As the estimates z are significantly different from the beliefs $q's$ of a single R for all her estimates q besides 2,3 and 4 ($p \leq 0.002$, Mann-Whitney U test), subjects in *T Individual* know that the *Rs* make mistakes.

Furthermore, we are interested in the question whether subjects make the same mistakes (or have the same bias) when evaluating themselves and when evaluating the *Rs*. On the one hand, subjects have better information about themselves than about the *Rs* and this should make it easier to judge themselves. On the other hand, individuals often reject information about themselves, such that they could see themselves in a good light (see, e.g., Bénabou and Tirole (2002)). This should make evaluating the *Rs* easier since a subject does not care about the implications of his choice (which the *Rs* will never get to know) on *R's* self-image. In *T Frame*, we find no significant difference between the own bias and the bias in assessing the average number of correct answers $\bar{t}$ of the *Rs* by choosing z. On average, the own bias in *T Frame Hard* is about 0.26 larger in absolute terms (it is more negative) than the bias in assessing $\bar{t}$ and in *T Frame Tricky* the own bias is about 0.24 larger in absolute terms (it is more negative again). The latter finding is surprising as with the tricky questions the subjects see the correct answers to the questions after having assessed the own ability t, but *before* assessing $\bar{t}$ of the *Rs*. This additional information seems not to improve the subjects' assessment about the others.
Finally, we compare the own bias (mistake) and the bias in guessing the ability of a similar R in *T Individual*.[24] We find that the own bias is significantly larger – in the sense that overestimation is more pronounced – according to a Wilcoxon test ($p = 0.024$). The average own bias is -1.8, whereas the average bias when assessing a single R is -0.87.

belief $q \in \{1, 2, 3\}$, however, are most often considered to be correct, too.

[24] As in *R Hard* no one has belief zero or one, we have to skip those subjects in *T Individual* who have a belief of zero or one.

5 *T Individual*

For *T Individual* one should be aware that one should replace "over-, underconfident or unbiased", by "negative, none or positive mistake", as we explained in Section 3.1. Even for a rational subject the true and stated number of correct answers can differ – the subject might simply make (unsystematic) mistakes. In the following, we try to disentangle what subjects think about the bias or mistake of single subjects.

5.1 Beliefs About the Bias/Mistake of Single Subjects

Consider a subject that assesses the *Rs*. He might think that "each R is rational" (and only makes unsystematic mistakes). If he thinks each R is rational, this is consistent with the belief that the *Rs* are unbiased on average. Yet, if he thinks (some) single *Rs* are biased this can still be consistent with a belief that the whole population of *Rs* is on average unbiased: He might think that the biases cancel out for the population, i.e. "the population is rational".[25] Under some assumptions, we can calculate, which choices of subjects in *T Individual* are consistent with a belief that *Rs* are rational on average. We present two alternative ways to do this. The first alternative corresponds to the possibility "each subject is rational", the second one to "the population is rational" as just explained.
Regarding the first alternative, we assume that if an individual is rational the distribution of the mistake is uniform and symmetric around zero. This implies that the precision of *Rs* that state extreme q's is higher - for example, someone who says "I answered zero questions correctly" is always right (since he e.g. did not mark any answer). Denote the possible beliefs of R about her number of correct answers by $\{zero, \ldots, seven\}$. The choices of the subject in *T Individual* are $left$, *middle* and $right$, which mean that a subject thinks the other subject overestimates, correctly estimates or underestimates her correct answers. Given a specific belief of R, which (rough) choices of a subject in *T Individual* are consistent if he believes R is rational? We can infer that the following (rough) choices[26] are consistent given a belief $\{zero, \ldots, seven\}$: for belief *zero* and *one* action $right$ (i.e. underestimation), for *two*, *three*, $four$ and $five$ action *middle* (i.e. correct estimation) and for *six* and *seven* action $left$ (i.e. overestimation). For example, an R, who states she has one question correct, how many questions could it actually have correct? Under the assumption that mistakes are uniformly and symmetrically distributed around zero, a subject that has actually one, two, three or four questions correct could state that it has one correct (i.e. have the belief *one*). Then the average of actually correct questions is 2.5 (remember that we assume different mistakes are

[25]Note that a subject in *T Individual* does not know the distribution of types of the *Rs*, i.e. he does not know how many *Rs* think they have one question correct etc. Therefore, we make assumptions on a subject's belief about this distribution in the following.

[26]These are "rough" choices in the sense that we calculate averages and round these to receive the choice.

equally likely). This is by 1.5 larger than 1 what the subject guessed herself. Hence, one should choose for such an *R right* (i.e. underestimation).
Concerning the second alternative, we assume that single subjects can be biased, but that their (systematic) mistakes cancel out in the population ("the population is rational"). Thus, a subject that states $q = 0$ can make a mistake. We assume that the absolute value of mistakes is at most three and that all mistakes have the same probability. We can derive (similar to above) that the prediction, when a subject in *T Individual* should choose $left$, *middle* or $right$, is exactly the same as for the first alternative.[27]
Similarly, one can see, which choice of subjects in *T Individual* is consistent with his belief that *Rs* are unbiased on average, when they are asked to guess the number of correct answers of R (making the same assumptions as above). Namely, if a subject believes that each R is rational but makes mistakes, he should choose the following numbers for each of *R's* beliefs $\{zero, \ldots, seven\}$: *zero*: 2, *one*: 2.5, *two*: 2.5, *three*: 3.5, $four$: 3.5, $five$: 4.5, *six*: 4.5, and *seven*: 5.5. If a subject believes that the population of *Rs* is rational but single *Rs* are biased, he should choose the following numbers for each of *R's* beliefs $\{zero, \ldots, seven\}$: *zero*: 1.5, *one*: 2, *two*: 2.5, *three*: 3, $four$: 3, $five$: 4.5, *six*: 5, and *seven*: 5.5.
Hence, given the assumption that subjects think the others are rational on average[28], we can compare how the two alternative approaches (assumptions see above) fit the experimental data: Given subjects think that others are rational on average, is it rather the case that they think that each individual is rational or that individuals might be biased but the population is, nevertheless, rational?

5.2 Results on Knowledge About Mistakes

When analyzing *T Individual* in more detail, we are interested in the question whether subjects are aware that others make mistakes and whether (and when) they think such mistakes are unsystematic or systematic. In order to investigate these questions, we compare the two different approaches explained in the preceding section: We assume that subjects either think that each R is rational or that the population of *Rs* is rational. Regarding the belief in *T Individual* about the goodness of the *Rs'* guess, we can infer the following from the solid line in Figure 5[29]: For low stated q's of R, subjects think that she has more likely a higher q than she stated, while *Rs* with high q's are expected to have more likely a lower q than stated. This means that especially *Rs* with extreme beliefs are considered to be wrong. Remarkably, *no* subject states for all possible estimates q's of R that she makes the correct

[27]To be precise, only rough choices are the same, i.e. when we consider the rounded values. The unrounded values differ.

[28]We only consider this case here as it turned out that a majority of subjects thinks that the *Rs* are rational on average.

[29]The solid line is the same in Figure 6

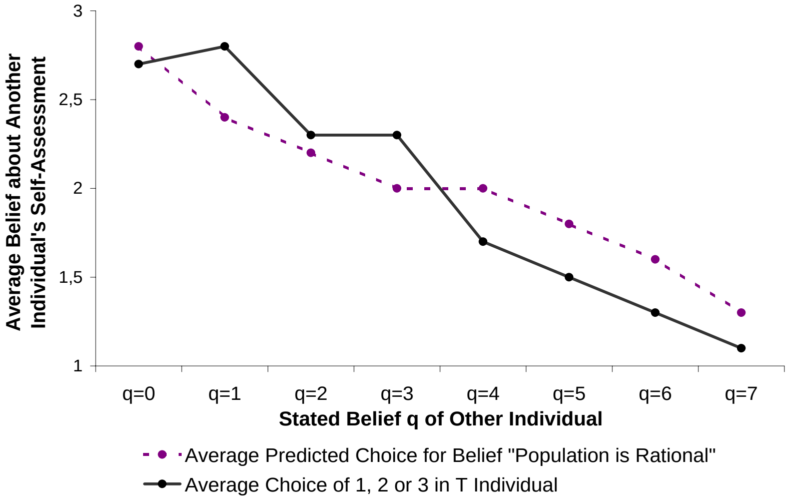

Figure 5: Average Assessment of *R's* Self-Assessment for Each of Her Possible Beliefs q about Her Ability t: 1=Overestimate, 2=Right, 3=Underestimate. Assumption: "Population is Rational"

choice. For each value of the belief q, $50-95$ percent of subjects state that R is wrong. Thus, subjects know that *Rs* make mistakes when these assess their q. Furthermore, we compare the average choice of subjects in *T Individual* with the predicted choice that would be consistent with subjects in *T Individual* thinking that the *Rs* are rational. In Figure 5, we derive the predicted choice under the assumption that the population is rational and in Figure 6, we derive it under the assumption that each single individual is rational and just make mistakes. We see in both figures that the curves for the predicted and actual choice are close. The second prediction fits, however, better for small qs.

Figures 7 and 8 again differ only in the predicted choices, but the actual considered choice is the same (i.e. the solid line). We see from the solid line that for more extreme beliefs of R, the average distance between the estimate z about R and the stated belief q of this R is increasing. Consistently with the choice above, subjects think that mistakes are more severe for *Rs* with extreme beliefs: The higher q, the more overestimation is pronounced and similarly, the lower q, the more underestimation is pronounced. We compare these actual choices to the predicted choice that would be consistent with subjects thinking that the *Rs* are rational but make mistakes. We see that the predicted curve in Figure 8 – where the assumption is that a subject thinks "each subject is rational but makes mistakes" – and the true curve are quite close for stated beliefs smaller than 4. In Figure 7 – where the assumption is that subjects think the population is rational – this is not true. For higher

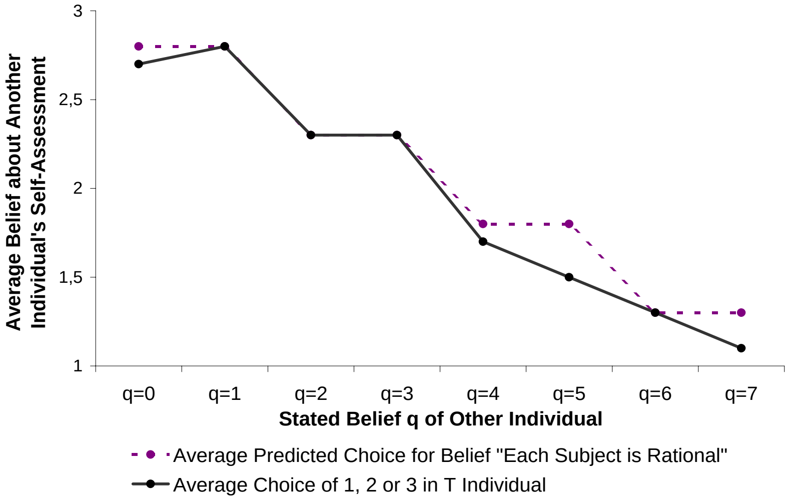

Figure 6: Average Assessment of *R's* Self-Assessment for Each of Her Possible Beliefs q about Her Ability t: 1=Overestimate, 2=Right, 3=Underestimate. Assumption: "Each Subject is Rational"

beliefs, the true curve and the predicted one diverge (under both assumptions). One possible interpretation of this is that subjects rather think that "pessimists" (those with low qs) are rational and just making mistakes instead of being biased. Whereas for "optimists" (those with high qs), we cannot infer which assumption fits observed behavior better.

6 Relative Bias

Finally, we want to analyze what subjects think about the relation of their own possible bias or mistake when assessing the number of correct questions and the bias or mistake of the *Rs*.

6.1 Experimental Design

An additional task is included in the following treatments: *T AveragePlus, T Frame, T Individual*. In *T Frame*, we explicitly asked subjects whether they think that "I and others made the correct choice" (or both are wrong/ others right/ I am right). If subjects are right with their statement, they receive 400 Tokens, otherwise they receive 50 Tokens. In *T AveragePlus* and *T Individual*, subjects choose between two alternatives (in *T Individual*, with the strategy method, they choose for each of the eight possible estimates q of R between the two alternatives). Subjects in *T AveragePlus* made this decision based on a payoff table

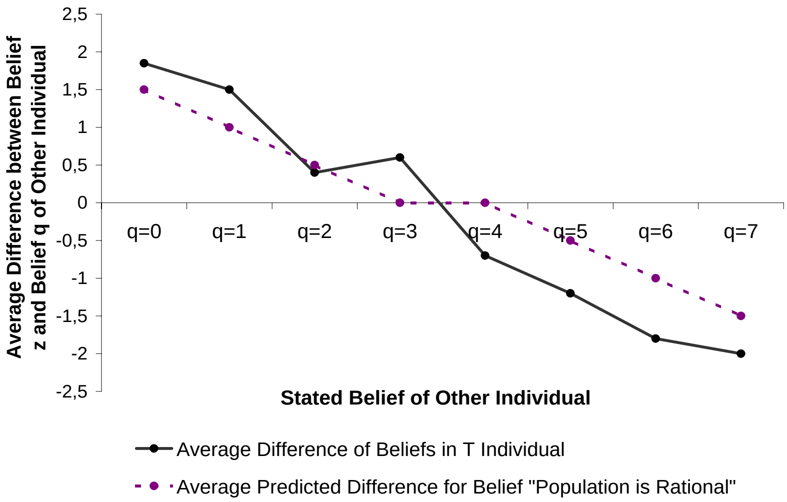

Figure 7: Average Value of Absolute Distance Between a Subject's Belief z about R's Ability t for Each of R's Possible Beliefs q about Her Own Ability. Assumption: "Population is Rational"

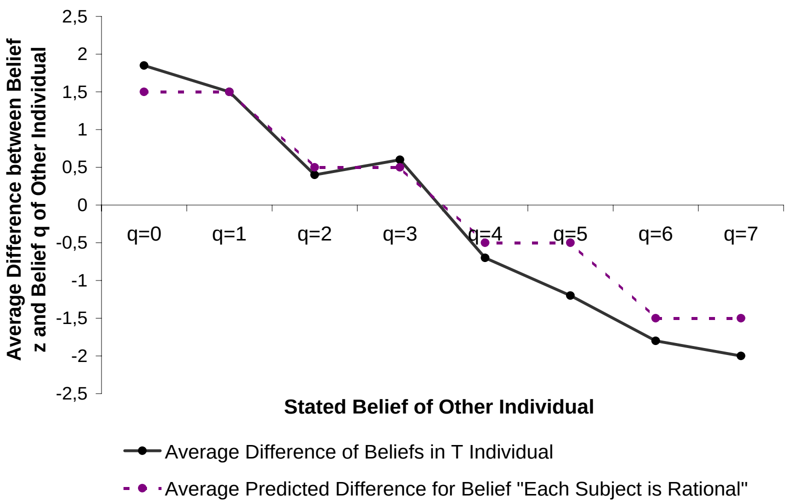

Figure 8: Average Value of Absolute Distance Between a Subject's Belief z about R's Ability t for Each of R's Possible Beliefs q about Her Own Ability. Assumption: "Each Subject is Rational"

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

Table 7: Payoffs - Relative Biases

(see Table 7). This payoff table shows that the payoffs of the alternatives I and II for the four possible events, where $q = t$ ($q \neq t$) refers to the self-assessment of a subject in *T AveragePlus* and $|\bar{q} - \bar{t}| \gtrless 0.5$ refers to the average self-assessment of the *Rs*. We see in Table 7 that payoffs of the two alternatives only differ for the second and third case, i.e. for the event that the subject is correct himself while the *Rs* are wrong on average or when the subject is wrong but the *Rs* are correct on average. In case that both are correct or both are wrong the alternatives lead to identical payoffs. Combining this with the statement whether he thinks that *Rs* are under-, overconfident or rational, one can see whether he thinks that both make the right/wrong decision or only one is wrong while the other is right (we will explain this in more detail below).

The decision of subjects in *T Individual* between the two alternatives is exactly the same as in *T AveragePlus* besides a difference in the payoff table. In *T Individual*, $|\bar{q} - \bar{t}| \gtrless 0.5$ is replaced by $q = t$ and $q \neq t$, i.e. whether the self-assessment of single *R* is correct or wrong instead of the average self-assessment.

6.2 Predictions

In this context, Proposition 1 (see appendix) is to be interpreted as follows. Define two states of the world: state 1 is the state in which a subject guesses his number of right answers correctly and the *Rs* are biased ($|\bar{q} - \bar{t}| > 0.5$). State 2 is the state in which the subject guesses his number of right answers not correctly and the *Rs* are (roughly) unbiased ($|\bar{q} - \bar{t}| < 0.5$). According to our Proposition, this subject should choose alternative I if he believes that state 1 occurs with a strictly larger probability than state 2 and otherwise alternative II. Combining the choice of alternative I or II with the statement that others are over-, underconfident or rational, we can deduce what individuals in *T AveragePlus* think about their mistakes or biases and others' biases. If, for example, a subject says that others are biased and chooses alternative I, this means that he thinks that he makes more likely the right decision himself, while the *Rs* do not, i.e. he is rational (or does not make a mistake) but the *Rs* are biased. If he says others are rational and chooses alternative I, this can be translated into the statement that the subject thinks that it is more likely that both are unbiased (or he makes no mistake). Saying that others are biased and choosing alternative

II implies that a subject thinks that both, himself and the *Rs* are biased (he may only make a mistake). And finally, saying that others are rational and choosing alternative II suggests that the *Rs* are unbiased, while oneself might be more likely biased (or make more likely a mistake).

In *T Frame* the inference about relative biases is easier as subjects immediately choose between the four alternatives "both are right/wrong", "only oneself is right", "only the others are right". Given these four alternatives, our proposition implies that a subject chooses the alternative with the statement, he believes most likely to occur.

As it is known from psychologists (see, e.g., Svenson (1981)), people tend to say that they are better than the average in ability tasks. As mentioned earlier, one can explain this observation by a self-serving bias. The tasks in our experiment do not require to say who is better in answering questions, but who is better in estimating the own ability or who is more rational. This is not the same but similar. Hence, we expect that subjects tend to indicate that they are rational (or do not make mistakes), while others are biased or that they at least state (by choosing alternative I) that this is more likely than the converse. Even though a subject may think that he makes better choices than a single R, this belief, however, seems surprising when he faces the complete group of *Rs* and their average estimate. The reason is that for him – even though he is rational – mistakes do not cancel out while for a rational average mistakes should (roughly) cancel out. Thus, we predict the following:

Hypothesis 4 *When facing a single R a majority of subjects tends to say that she is biased, while oneself is not (does more likely (not) make a mistake). When facing the average of the* Rs *it is the other way round.*

Applying again the assumption that "each subject is rational", it can further be seen that if a subject believes that an R is rational, then for the *Rs* that indicate that they have $q \in \{2, 3, 4, 5\}$ questions correct, the average deviation between stated and true number of correct answers is 0.5, while for the remaining q's it is strictly larger than 1. Thus, if one thinks that oneself and R are rational but make/s mistakes, then it could be plausible[30] to state that it is more likely that oneself is wrong and R is correct than is the converse (i.e. choosing alternative II) for $q \in \{2, 3, 4, 5\}$ and to state that it is more likely that oneself is correct and she is wrong than is the converse (i.e. alternative I) for the remaining actions q.

6.3 Results (Hypothesis 4)

In this section, we present the results on the question what individuals think about the relation between their own bias and others' biases. Who is more likely to be biased or make mistakes?

[30] One could not say what is implied here, since it depends on the belief about the size of the own mistake relative to the other one.

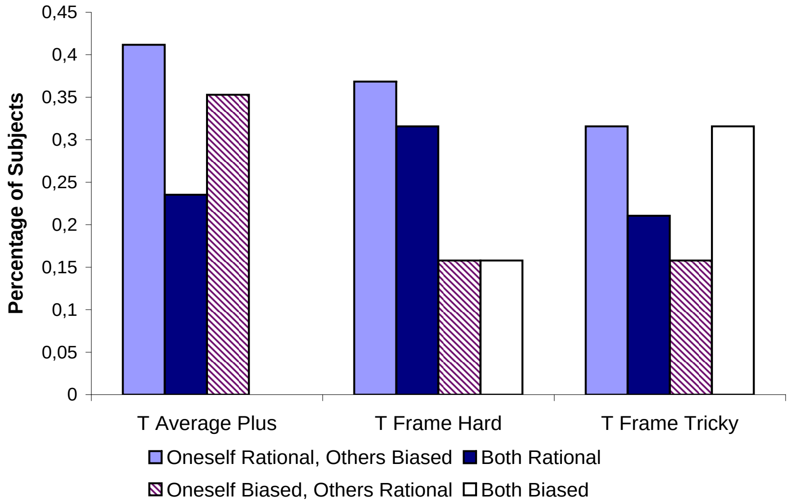

Figure 9: Relative Beliefs

Figure 9 shows the percentages of subjects in *T AveragePlus* and *T Frame* (*Hard* and *Tricky*) thinking that oneself does not make a mistake and the *Rs* are biased, that oneself and the *Rs* are correct, that oneself makes a mistake while the *Rs* are rational or that both are wrong (make a mistake/have a bias).

In *T AveragePlus*, we observe that 65 percent of the subjects choose alternative I – i.e. they rather think that they are correct themselves and the *Rs* are wrong – and 35 percent choose alternative II. Combining this choice with their statement about the rationality of *Rs*, we get the percentages in Figure 9. In the part of *T Frame* with the hard questions the percentage of those, who think that their self-assessment is better, is lower. In both treatments roughly the same percentage of subjects thinks that they are rather wrong themselves. In *T AveragePlus*, however, 35 percent think that the *Rs* are rather better, while in *T Frame Hard* 16 percent think that the *Rs* are better or that both are biased/make mistakes, respectively. With the tricky questions, less subjects think that both are correct (21%), slightly less think that they are better themselves (32%), and many more think that both are wrong (32%). Thus, with the tricky questions, we find that the percentage of those, who think that oneself (not neccessarily the *Rs*) is right, decreases.[31] We summarize these findings in the following result:

Result 4 *The majority of subjects thinks that it is more likely that they do not make a*

[31]We can, however, not reject the hypothesis that there is no relation between the number of subjects who think that oneself is correct or wrong according to Fisher's exact tests ($p > 0.05$ one-sided). If we consider only those, who think that the *Rs* are wrong, there are significantly less (more) who think that they are correct (wrong) themselves than in *T AveragePlus* (Fisher's exact test, p=0.034 one-sided).

mistake, while the others are biased. This percentage decreases as subjects receive more information (i.e. framing and seeing the correct answers).

This result is somehow surprising. If one thinks that all subjects (oneself and the others) are rational, one should tend to choose alternative II since a single rational individual makes mistakes, while for the average they cancel out. The choice of alternative I is only consistent with the beliefs "I do not make mistakes at all" or "the average is very likely biased and I am unlikely to make a mistake". The first belief is surprising as it implies that subjects are not aware that they might make mistakes. Also the second belief is surprising as subjects seem to be aware that there is something like a bias in the population, but are not aware that they are biased (or make a mistake).[32] From the analysis of *T Individual* we already know, that subjects are aware that others make mistakes. Although subjects are aware of it, they do not think that they make mistakes themselves. We can explain this again by a self-serving bias. Subjects think that they are different from the others or better than these are.

Next, we consider how the own bias is related to the belief about the relative bias. When evaluating the relative bias, subjects may be reluctant to say that they are better than others in their self-assessment. This behavior would reveal overconfidence of the type "I think I am better than the average". It is hence interesting to see whether subjects, who have a bias when evaluating absolute abilities (here: answering questions), also have a bias when assessing relative abilities (here: evaluation of relative bias). In *T AveragePlus*, for those, who choose alternative I (i.e. they rather think that they are correct themselves and the *Rs* are wrong), the difference between true type and believed type is on average -2.18, while for those, who choose alternative II it is 0 (meaning that they are unbiased).[33] According to a Mann-Whitney U test, subjects who think that they are more likely to be correct, are significantly more biased than those who think that rather the *Rs* are correct whilst oneself is wrong ($p = 0.007$). Remarkably, *all* subjects who choose alternative I – i.e. who rather think that they are correct themselves and the *Rs* are wrong – are overconfident themselves. The pattern in *T Frame* is similar. Those subjects, who think that they are right, while the *Rs* have a bias, have the largest bias (average bias is -1 with the hard questions, -2 with the tricky ones). Those, who say that they as well as the *Rs* are likely to be wrong, are either underconfident (with the hard questions their bias is 1) or have the smallest bias with the tricky questions. This may be due to the fact that subjects, who know that individuals are biased, try to behave accordingly and adjust their guess of the number of correct answers

[32]They might think that the only possible state of the world is that both are wrong. Hence, they are indifferent between the two alternatives. We think, however, that subjects put some (small) positive probability on the other states of the world such that the choice of alternative I reveals that they think they are correct, while the others are not.

[33]Since positive and negative biases cancel out, also the average *absolute values* of the biases are interesting. These are 2.18 given alternative I and 1.33 given alternative II, i.e. they are still lower for alternative II.

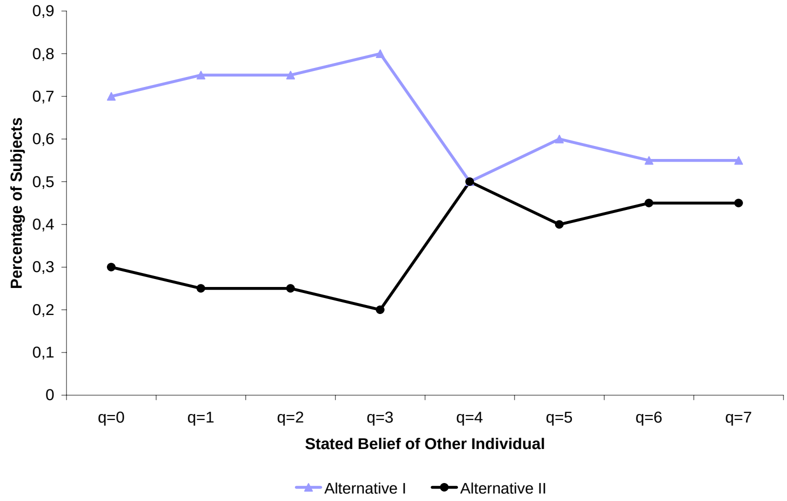

Figure 10: Average Choice of Alternatives in *T Individual*

downward. Those, who say that both are correct, have roughly the same bias as those, who say that the Rs may make better guesses (-0.5 (-1.75) versus -0.67 (-1.7) for hard (tricky) questions).

Finally, we turn to the analysis of *T Individual*. The average choice of alternatives I or II of the subjects for every single belief of an R is shown in Figure 10. Here we see again, that subjects tend to think that they are more likely to make the correct self-assessment themselves, i.e. they tend to choose more often alternative I (confirming our Hypothesis). Moreover, for low beliefs of an R about her type (i.e. pessimistic Rs), subjects in *T Individual* tend even more to alternative I than for high beliefs of R (i.e. optimistic Rs). Thus, it seems that they trust an optimist more to make the right decision than a pessimist. This observation is interesting since subjects think that both, optimists as well as pessimists make mistakes as we have also seen in Section 5.2. This means that although subjects realize that Rs with high qs might just appear to be a good type (since they might be actually worse then they think), subjects still seem to believe that these "high types" are somehow better than others.[34]

As we have mentioned before, if one thinks that oneself and R are rational (but make/s mistakes), one should choose for $q \in \{2, 3, 4, 5\}$ alternative II. We see in the figure that the proportion of subjects that chooses alternative I is always larger than the proportion

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

choosing alternative II. For $q \in \{4, 5, 6, 7\}$ the proportions are close. Thus, there is again a tendency that subjects think that they are better (more rational) than others (here: better than single individuals in *R Hard*).

7 Conclusion

Empirical studies show that overconfidence occurs in various settings: People overestimate their driving abilities, students their scores in exams or their rank in the distribution, couples the likelihood of not getting divorced, and portfolio managers their prediction abilities. In our experiment, subjects estimated how many out of seven multiple choice questions they answered correctly. Our observations confirm that overestimation of the own ability is a prominent phenomenon in the population. As overconfidence is such a common characteristic of people's behavior and is observed so frequently in real life, it seems obvious that people are also aware of this bias. Remarkably, we find that a majority of subjects does not think or know that others have a bias.

What are the consequences of this ignorance? If we think of economic interactions, it is often important that agents are aware that others are biased in order to make optimal decisions. For instance, an agent who is not aware that his opponent in a contest is biased or who does not know at least, which belief the opponent has about his type, cannot adjust his effort optimally, as is assumed in Ando (2004). Malmendier and Tate (2005a, 2005b) observe that managers are overconfident and that this is disadvantageous for the firm. Thus, principals should be aware of overconfident managers in order to be able to counteract possible decision defects. Santos-Pintos (2005a) show that incentive contracts should be designed in a special way for overconfident agents. Given that a majority of subjects tends to be overconfident, ignorance of this bias leads to suboptimal contracts for a majority of agents.

Our results indicate that the more information subjects receive on the task the others have to complete – i.e. the more familiar they are with the task – the more subjects learn that others are on average biased regarding this task. Hence, more familiarity with the task others have to complete might help subjects to recognize that others are biased. Therefore, it helps to make better decisions when subjects face biased individuals. This highlights the importance of information and feedback to make better economic decisions.

Moreover, we observe that when subjects are confronted with the question what they think about the relation between their own and others' biases, a majority states that they are more likely able to estimate their ability correctly than is the population. This has, for example, implications for decision making in firms: Suppose a principal has to decide whether to delegate to subordinates or not. In real life, it is often observed that people do not want to delegate even though there is no incentive or verifiability problem. Our results indicate that one explanation for this phenomenon is that either the principal believes that other agents are not able to make the right decision as they have a bias while he has no bias himself or that his bias is smaller.

We have also seen that individuals think that similar subjects are very likely not to make mistakes, while they, nevertheless, know that mistakes occur. This indicates that subjects are not only unaware of the bias in the population, but also of their own bias (or mistakes). Although we think that this can be taken as some evidence, we think it is a topic for future research to investigate the knowledge about own biases (like hyperbolic discounting or overconfidence) further.

Appendix

Before we prove that subjects play a pure strategy, we first want to define beliefs and strategies of a subject. We let μ_j be the individual's belief that alternative j is true, where $j \in \{0, \dots J\}$, $J \in \{2, 3, 4, 8, 70\}$ with $\sum_j^J \mu_j = 1$. Next, we define a strategy of an individual. A pure strategy of an individual is an action or alternative a subject can choose, i.e. a pure strategy is $j \in \mathcal{J} = \{0, \dots J\}$. Given an individual's pure strategy set $\mathcal{J}$, an individual's mixed strategy, $\sigma : \mathcal{J} \to [0, 1]$, assigns to each pure strategy j a probability $\sigma_j \geq 0$ with which j will be played, where $\sum_j \sigma_j = 1$. Further, we denote by c the high payoff (i.e. 525, 1680, 105, 400 or 500 Tokens) and by $c - \kappa$ the low one (i.e. 20, 30, 315, 50 or 300 Tokens). We assume, without loss of generality, that $\mu_1 \geq \mu_2 \geq \dots \geq \mu_J$.
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 $\mu_1 = \dots = \mu_n > \mu_{n+1} \geq \dots \geq \mu_J$ with $J \geq n \geq 2$, then any σ with $\sigma_1 + \dots + \sigma_n = 1$ 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 – subjects have the choice between $\{1, \dots, J\}$ alternatives with $J \in \{2, 4, 3, 8\}$, we can apply this proposition to all of them. If subjects make the "right" choice (the right guess for the interval, the right guess for the relative bias or the right guess for the number of correctly answered questions), they receive a high payoff, say c, and if the choice is not correct, they receive $c - \kappa$.

Proof.
The subjectively expected utility of an individual from strategy σ is

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

Rearranging yields

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

Suppose now that subjects never put the same probability on alternatives. Without loss of generality $\mu_1 > \mu_2 > \dots > \mu_J$. 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

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

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

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

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

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

This term is (strictly) negative (the term equals zero if $\mu_1 = \mu_2 = \cdots = \mu_J$), 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 $\mu_1 > \mu_2 \geq \cdots \geq \mu_J$, since $\sigma'_1 < 1$. If, however, $\mu_1 = \cdots = \mu_n > \mu_{n+1} \geq \cdots \geq \mu_J$ with $J \geq n \geq 2$, then any σ with $\sigma_1 + \cdots + \sigma_n = 1$ 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 $\sigma_1 + \cdots + \sigma_n = 1$ can be optimal. ■

<u>Instructions (translated from German)</u>

Instructions *R Hard* – Part 1

In this scientific experiment you can earn money with your decisions. During the experiment your payoffs are given in tokens.
After the experiment this amount of tokens will be converted into euros according to the exchange rate of **1 euro for 210 tokens** and paid cash to you.

<u>Course of the Experiment:</u>

The experiment consists of two stages. In stage 1 you answer 7 multiple-choice questions. In stage 2 you make a decision. **The payoff for this decision depends among other things on the number of multiple-choice questions you answered correctly**. You get the instructions for stage 2 after having answered the 7 questions.

Stage 1:

- **7 multiple-choice questions** are posed. For each question you get **4 possible answers** to choose from. At a time, only **one** of these possible answers is correct.
  You choose your answer to a question by clicking on the circle in front of the corresponding answer and then clicking "OK". As soon as you click OK, you cannot change your answer any more and the next question appears.

- You have at most 45 seconds to give your answer to each of the questions. During these 45 seconds you can give your answer at any time. The time that is left for a question is shown on the screen. When time has run out, the computer automatically shows the following question.
- **Please note:** If you do **not** click on one answer or **not** click OK before the time has run out, this means the same as if you give a wrong answer.
- Once you have answered all questions, the computer determines how many questions you have answered correctly. **You receive the information how many correct answers you have after the experiment, i.e. after stage 2.**

<u>Payoff for stage 1:</u>

For each **correct** answer you receive **190 tokens** and for each **wrong one you receive 10 tokens**.

Instructions *R Hard* – Part 2

Stage 2:

In stage 2 you choose one out of eight possible **actions 0, 1, 2, 3, 4, 5, 6, 7**. This is done by entering one of these numbers in the corresponding cell on the computer screen and you confirm your choice by clicking on "OK".

Payoff stage 2:

The following table shows the payoffs in tokens, which you receive depending on you choice and how many questions you answered correctly in stage 1. **You are not told until after the experiment how many questions you answered correctly.**

		Number of correct questions							
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Action 0	525	30	30	30	30	30	30	30
A	**Action 1**	30	525	30	30	30	30	30	30
c	**Action 2**	30	30	525	30	30	30	30	30
t	**Action 3**	30	30	30	525	30	30	30	30
i	**Action 4**	30	30	30	30	525	30	30	30
o	**Action 5**	30	30	30	30	30	525	30	30
n	**Action 6**	30	30	30	30	30	30	525	30
	Action 7	30	30	30	30	30	30	30	525

Calculation of your total payoff:

- Your total payoff from the experiment is given by the **number of all your correctly answered questions** multiplied by 190 tokens and the **number of wrong answers** multiplied by 10 tokens (your payoff in stage 1) and the payoff from your chosen action (your payoff in stage 2). In addition you receive a payment of 525 tokens.
- This total payoff is converted into euros according to the exchange rate 1 Euro = 210 tokens.

Instructions *R Tricky*

The instructions for *R Tricky* are identical to *R Hard*. The difference is that subjects answer the tricky instead of the hard questions.

Instructions *T Average*

Course of the experiment:

You make a decision between three actions and and a decision about a number. In order to make this decisions, you receive some information on another experiment *(Experiment I)*, which has been conducted a week before.

Description of *Experiment I*

Experiment I had **20 participants**. The experiment consisted of two stages.

Stage 1

- In the first stage, the participants answered 7 multiple-choice questions. For each of the questions there have been 4 possible answers. At a time, only **one** of these possible answers was correct. For each of the questions, the participants had at most 45 seconds to give their answer. When time had run out, the computer automatically showed the next question. In case no answer had been clicked on during this time, this was equivalent to giving a wrong answer.
- The questions are attached to these instructions and you can look at them later on.

Stage 2

- In the second stage the participants have chosen one out of eight possible **actions 0, 1, 2, 3, 4, 5, 6, 7**.
- The payoffs (in tokens) for every possible combination of the "number of correctly answered questions" and the "chosen action" have been determined according to the payoff table below. The participants of *Experiment I* had this table in stage 2 in order to make their decision.

		Number of correct questions							
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Action 0	525	30	30	30	30	30	30	30
A	**Action 1**	30	525	30	30	30	30	30	30
c	**Action 2**	30	30	525	30	30	30	30	30
t	**Action 3**	30	30	30	525	30	30	30	30
i	**Action 4**	30	30	30	30	525	30	30	30
o	**Action 5**	30	30	30	30	30	525	30	30
n	**Action 6**	30	30	30	30	30	30	525	30
	Action 7	30	30	30	30	30	30	30	525

Further relevant information

- The participants knew in stage 1 (when answering the 7 questions) that they make a decision in stage 2 and that the payoff depends on the number of correctly answered questions. The payoff table and detailed instructions for stage 2 have not been handed to the participants until the beginning of stage 2.
- The number of correct questions has been determined for each participant by the computer. At the end of the experiment, the participants received 190 tokens for each correct answer and 10 tokens for each wrong one. **Each participant has not been told the number of correctly answered questions and the payoff until he/she has chosen his/her action in stage 2.**
- At the end of the experiment, the payoff of the participants from answering the questions and from their decision, as well as an additional payment of 525 tokens has been converted into euros according to the exchange rate **210 tokens = 1 euro** and paid cash to the participants.

Description of today's Experiment

Relevant results from *Experiment I*:

Based on the answers and the decisions of the participants of *Experiment I*, two averages have been calculated **after** the experiment:

1. The **average number of correct answers "R"** of all participants:

   The average is calculated as follows: the number of correct answers of all participants is added and then divided by the number of participants (20). The resulting value is rounded on one **decimal place**. Thus, the average can take values from 0 to 7 in steps of 0.1.

2. The **average action "A"** chosen by the participants:

   The average is calculated as follows: each participant chooses an action whereat the actions are assigned numbers from 0 to 7 (see table). The **numbers** of the chosen action of each participant are added and then divided by the number of participants (20). The resulting value is rounded on one **decimal place**. Thus, the average action can also take values from 0 to 7 in steps of 0.1.

Your decision:

Before you make your decision, you are **told the value of the average action** (**A**) chosen by the participants of *Experiment I*.

- You choose between three actions: **action 1**, **action 2 und action 3**. You select action 1, 2 or 3 by clicking on the corresponding action on the computer screen. In the following the actions are explained more detailed.
- After you have chosen one of the actions, you choose a number as described in the following:
  - If you have chosen **action 1**, you can choose a **number which is larger than or equal to A - 0.5 and smaller than A + 0.5**.
  - If you have chosen **action 2**, you can choose a **number which is larger than A + 0.5 and smaller than or equal to 7.**

- If you have chosen **action 3**, you can choose a **number which is larger than or equal to 0 and smaller than A - 0.5**.

A= average action of the participants of Experiment I

- You can give the number in **steps of 0.1**. You chose a number by entering the number you want to choose in the corresponding cell on the screen.
- When you have made all decisions, please confirm your choice by clicking on “OK”.

Your payoff consists of the following two components:

Payoff component 1:

		Actions		
		Action 1	**Action 2**	**Action 3**
Value of the Average number of correct questions (R)	**R is smaller than *A-0.5***	315	315	1680
	R is larger/equal *A - 0.5 and* smaller/equal *A+0.5*	1680	315	315
	R is larger than *A+0.5*	315	1680	315

A= Value of the average action of the participants of *Experiment I*

R= Value of the average number of correct questions in *Experiment I*

Payoff component 2:

If the distance (explanation see below) between the number you have chosen and the average number of correct questions **R** is smaller than or equal to 0.5 and your payoff from component 1 is *1680* tokens, then you receive in addition *105* tokens, if the distance is larger than 0.5 or your payoff from component 1 is *315* tokens, you receive *20* tokens.

Explanation „Distance“:

Consider two numbers X and Y. The distance between these two numbers is

X-Y if X is larger than Y and Y-X if X is smaller than Y.

Total payoff:

Your total payoff ist he sum of your payoffs from component 1 and 2 and an additional payment of 625 tokens.

Instructions *T AveragePlus* – Part 1

<u>Course of the experiment:</u>

The experiment consists of 4 stages. In stage 1 you answer 7 multiple-choice questions. In stage 2 you make a decision. **The payoff for this decision depends among other things on the number of multiple-choice questions you answered correctly**. After stage 2 you receive some information on another experiment (*Experiment I*). In *Experiment I* stage 1 and 2 have been played as well. Having received this information, you make a decision between two alternatives in stage 3. The payoff you get from the choice of an alternative depends on *Experiment I* and your decision in stage 2. In stage 4 you make a decision between three alternatives, whereat your payoff from the choice of an alternative depends on *Experiment I*. You get the instructions for stage 2, 3 and 4 after having answered the 7 questions.

Stage 1:

Exactly like in *R Hard.*

Instructions *T Average+* – Part 2

Stage 2:

Exactly like in *R Hard.*

Relevant Information on *Experiment I:*

In ***Experiment I*** there have been **20 participants**. The experiment consisted of exactly the same 2 stages as just described: Answering 7 multiple-choice questions in stage 1 and choice between actions 0, 1, 2, 3, 4, 5, 6, 7 in stage 2.

At the end of *Experiment I*, payoffs of the participants from answering the questions and from the decisions as well as an additional payment of 525 tokens have been converted into euros according to the exchange rate **1 Euro per 210** tokens and paid cash to the participants.

Based on the answers and the decisions of the participants of *Experiment I*, two averages have been calculated **after** the experiment:

The **average number of correct answers "R"** of all participants:

The average is calculated as follows: the number of correct answers of all participants is added and then divided by the number of participants (20). The resulting value is rounded on one **decimal place**. Thus, the average can take values from 0 to 7 in steps of 0.1.

The **average action “A” chosen** by the participants:

The average is calculated as follows: each participant chooses an action whereat the actions are assigned numbers from 0 to 7 (see table). The **numbers** of the action of each participant are added and then divided by the number of participants (20). The resulting value is rounded on one **decimal place**. Thus, the average action can take values from 0 to 7 in steps of 0.1.

Before you make your decision in stage 3 and 4, you are **told the value of the average action** (**A**) chosen by the participants of *Experiment I*.

Stage 3:

Decision stage 3:

In stage 3 you can choose between the following two alternatives. The choice is done by clicking on the alternative on the screen and confirming the choice with “OK”.

Payoff stage 3:

Your payoff in stage 3 depending on the distance between **R** and **A**, your payoff in stage 2 (which you are not told until the end of the experiment) and your choice between the two alternatives is:

	Alternative 1	**Alternative 2**
Your payoff in stage 2 is 525 and the distance between R and A is smaller than or equal to 0.5	800	800
Your payoff in stage 2 is 525 and the distance between R and A is larger than 0.5	500	300
Your payoff in stage 2 is 30 and the distance between R and A is smaller than or equal to 0.5	300	500
Your payoff in stage 2 is 30 and the distance between R and A is larger than 0.5	210	210

Explanation „Distance“:

Consider the two numbers R and A. The distance between these two numbers is

R-A if R is larger than A and A-R if R is smaller than A.

Stage 4:

Decision stage 4:

You choose between three alternatives: **Left**, **Middle and Right** by clicking on the corresponding alternative on the computer screen. Please confirm your choice by clicking on “OK”.

<u>Payoff stage 4:</u>

		Alternatives		
		Middle	**Right**	**Left**
Value of the Average number of correct questions (R)	**R is smaller than** ***A-0.5***	315	315	1680
	R is larger/equal ***A - 0.5 and*** **smaller/equal** ***A+0.5***	1680	315	315
	R is larger than ***A+0.5***	315	1680	315

A= Value of the average action of the participants of *Experiment I*

R= Value of the average number of correct questions in *Experiment I*

Calulation of your total payoff:

Your total payoff in the experiment is given by the sum of:

- The number of all your correctly answered questions multiplied by 190 tokens and the number of wrong answers multiplied by 10 tokens (your payoff in stage 1).
- Your payoff in stage 2.
- Your payoff in stage 3.
- Your payoff in stage 4.
- In addition you receive a payment of 725 tokens.

This total payoff is converted into euros according to the exchange rate 1 Euro = 210 tokens.

Instructions *T Frame* – Part I

Course of the experiment:

The experiment consists of two parts: In part I you answer **two blocks of questions A and B** each with 7 multiple-choice questions. In part II you make 8 decisions. The first four decisions **(1A-4A)** refer to question block **A**, the next four decisions **(1B-4B) to** question block **B**.

The payoff from decision 1A (1B) depends among other things on your number of correctly answered multiple-choice questions in block A (B). Afterwards you receive some information on another experiment (*Experiment I, II resp.*). In *Experiment I (II)* question block A (B) have been answered and decision 1A (1B) have been made, too. Having received the information, you make decision 2A (2B). The payoff for decision 2A (2B) depends on *Experiment I (II)* and on your decision 1A (1B). Subsequently, you make decision 3A (3B) and 4A (4B), whereat your payoffs depend on *Experiment I (II).*

Stage 1:

Exactly like in *R Hard* except that subjects answer two different blocks of 7 multiple-choice questions (the hard and the tricky questions). Subjects are paid like in *R Hard* but only for one block of questions that is randomly selected.

Instructions *T Frame* – Part II

Decision 1A:

Decision 1A:

You state how many of the 7 questions in question block A you think you have answered correctly. For this, you enter a whole number between 0 and 7 in the corresponding cell and then click on „OK".

Payoff decision 1A:

If your statement coincides with the actual number of correctly answered questions in block A („Your estimation is correct"), you receive 525 tokens, i fit does not coincide ("Your estimation is not correct"), you receive 30 tokens.

Relevant information on *Experiment I and II:*

In ***Experiment I and II*** respectively there have been **20 participants**. These experiments consisted of answering the questions of block A in *Experiment I* and block B in *Experiment II* and each time a statement, how many questions have been answered correctly. For this statement, the participants have chosen between eight actions 0, 1, 2, 3, 4, 5, 6, 7. In case the actual number of correctly answered questions coincides with the number of the action, a participant received 525 tokens, if there was no coincidence he/she received 30 tokens.

At the end of *Experiment I (II)* payoffs of the participants from answering the questions and from the decisions as well as an additional payment of 525 tokens have been converted into euros according to the exchange rate **1 Euro per 210** tokens and paid cash to the participants.

Based on the answers and the decisions of the participants of *Experiment I* and *II* respectively, two averages for each experiment have been calculated **after** the experiment:

The **average number of correct answers "R"** of all participants:

The average is calculated as follows: the number of correct answers of all participants is added and then divided by the number of participants (20).

The **average estimation "E"** of the participants:

The average is calculated as follows: the chosen statements about the number of correctly answered questions of each participant are added and then divided by the number of participants (20).

Both averages **E** and **R** of *Experiment I* and *II* are rounded on **one** decimal place. Thus, the averages can take values from 0 up to 7 in steps of 0.1.

Decisions 2A, 3A and 4A:

Before you make decisions 2A, 3A and 4A, you are told the **value of the average estimation (E)** of the participants of *Experiment I*.

Decision 2A:

You decide how good your estimation of the number of correct questions is and how good the average estimation (E) of the participants of *Experiment I* is. There are four alternatives:

- **"Both estimations are good**": your estimation is correct (see above) and the distance (explanation see below) between the average estimation (E) and the average number of correct questions (R) in *Experiment I* is smaller than or equal to 0.5.
- **"My own estimation is better"**: your estimation is correct and the distance between E and R in *Experiment I* is larger than 0.5.
- **„Average estimation is better"**: your estimation is not correct and the distance between E and R in *Experiment I* is smaller than or equal to 0.5.
- **"Both estimations are bad"**: your estimation is not correct and the distance between E and R in *Experiment I* is larger than 0.5.

Payoff decision 2A:

If you select the alternative that is actually true, you receive 400 tokens, otherwise you receive 50 tokens.

Explanation „distance":

Consider the two numbers R and E. The distance between these two numbers is R-E if R is larger than E and is E-R if R iss maller than E.

Decision 3A:

You state how well you think the participants in *Experiment I* assess themselves:

- The participants **overestimate** their actual number of correctly answered questions on average. This means that the average number of correct (R) in *Experiment I* is by more than **0.5 smaller** than the average estimation (E).
- The participants estimate their actual number of correctly answered questions on average almost **correct**. This means that the average number of correct (R) in *Experiment I* is larger than or equal to **E-0.5** and smaller than or equal to **E+0.5**.

- The participants **underestimate** their actual number of correctly answered questions on average. This means that the average number of correct (R) in *Experiment I* is by more than **0.5 larger** than the average estimation (E).

You choose between the three alternatives **(overestimate, correct, underestimate)** by clicking on the corresponding alternative and confirming with OK.

<u>Payoff decision 3A:</u>

When the alternative you have chosen is actually true, then you receive **1680 tokens**, when it is **not** true, you receive **315 tokens**.

<u>Decision 4A:</u>

You state, what you think **how large the average number of correctly answered questions (R)** of the participants in *Experiment I* is. This is done by entering a number between 0 and 7 in steps of 0.1 in the corresponding cell.

Take notice of the following conditions:

- If you have chosen "**correct"** in decision 3A, you can choose a **Number** that is **larger than or equal to E - 0.5 and smaller than or equal to E + 0.5**.
- If you have chosen **"underestimate"** in decision 3A, you can choose a **Number** that is **larger than E + 0.5 and smaller than or equal to 7.**
- If you have chosen **"overestimate"** in decision 3A, you can choose a **Number** that is **larger than or equal to 0 and smaller than E - 0.5**.

<u>Payoff decision 4A:</u>

If the distance between the number you have chosen and the average number of correct questions **(R)** is smaller than or equal to 0.5 and you selected in decision 3A the alternative that is actually true, then receive **105 tokens**, otherwise you receive **20 tokens**.

Decision 1B-4B

After decisions 1A-4A decisions 1B-4B regarding block B follow.

Here, the following decisions are equivalent 1A-1B, 2A-2B, 3A-3B, 4A-4B besides that they refer now to **block B** and ***Experiment II***.

After decision 2B you are told the correct answers to the questions of block B. Afterwards you make decision 3B and 4B.

Calculation of your total payoff:

Your total payoff from the experiment is the sum of:

- The number of your correctly answered questions in the block of questions randomly selected by the computer multiplied 190 tokens and the number of wrong answers in this block multiplied by 10 tokens.
- Your payoff from decisions 1A-4A or 1B-4B: For the payment the computer again randomly selects whether decisions 1A-4A or 1B-4B are paid.
- In addition you receive a payment of 420 tokens.

This total payoff is converted into euros according to the exchange rate 1 Euro = 210 tokens.

Instructions *T Individual* – Part 1

Course of the Experiment:

The experiment consists of 4 stages. In stage 1 you answer 7 multiple-choice questions. In stage 2 you make a decision. **The payoff for this decision depends among other things on the number of multiple-choice questions you answered correctly**. After stage 2 you receive some information on another experiment (*Experiment I*). In *Experiment I* stage 1 and 2 have been played as well. Having received this information, you make eight times a decision between two alternatives in stage 3. The payoff you get from the choice of an alternative depends on *Experiment I* and your decisions in stage 2. In stage 4 you make eight times a decision between three alternatives, whereat your payoff from the choice of an alternative depends on *Experiment I*. You get the instructions for stage 2, 3 and 4 after having answered the 7 questions.

Stage 1:

Exactly like stage 1 in *R Hard.*

Instructions *T Individual* – Part 2

Stage 2:

Decision stage 2:

In stage 2 you choose one out of eight possible **actions 0, 1, 2, 3, 4, 5, 6, 7**. This is done by entering one of these numbers in the corresponding cell on the computer screen and you confirm your choice by clicking on "OK".

Payoff stage 2:

The following table shows the payoffs in tokens, which you receive depending on you choice and how many questions you answered correctly in stage 1. **You are not told until after the experiment how many questions you answered correctly.**

		Number of correct questions							
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Action	**Action 0**	525	30	30	30	30	30	30	30
	Action 1	30	525	30	30	30	30	30	30
	Action 2	30	30	525	30	30	30	30	30
	Action 3	30	30	30	525	30	30	30	30
	Action 4	30	30	30	30	525	30	30	30
	Action 5	30	30	30	30	30	525	30	30
	Action 6	30	30	30	30	30	30	525	30
	Action 7	30	30	30	30	30	30	30	525

Relevant Information on *Experiment I:*

In ***Experiment I*** there have been **20 participants**. The experiment consisted of exactly the same 2 stages as just described: Answering 7 multiple-choice questions in stage 1 and choice between actions 0, 1, 2, 3, 4, 5, 6, 7 in stage 2.

At the end of *Experiment I*, payoffs of the participants from answering the questions and from the decisions as well as an additional payment of 525 tokens have been converted into euros according to the exchange rate **1 Euro per 210** tokens and paid cash to the participants.

Based on the answers and the decisions of the participants of *Experiment I*, two values have been identified after the experiment:

1. The **number of correct answers "R"** of a participant.
2. The **action "A" chosen** by a participant. The value A of an action is a number between 0 and 7 next to an action (see table).

You are randomly assigned to one participant of *Experiment I*. When you make your decisions, you do not know which participant it is. Therefore, you make your decisions in stage 3 and 4 for all possible values of A, i.e. 0,1,2,3,4,5,6,7. For none of these values of **A** you get to know the value of **R**.

Stage 3:

Decision stage 3:

In stage 3 you choose for every possible **A** (0,1,2,3,4,5,6,7) between two alternatives – i.e. you make **eight** times a decision between the two alternatives. The choice is done by clicking on the alternative on the screen and confirming the choice with "OK" when you finished **all eight** decisions.

Payoff stage 3:

The following table shows your payoff in stage 3 depending on the values of R **and** A of the participant of ***Experiment I*** that is assigned to you**, your payoff in stage 2 (which you are not told until the end of the experiment) and your choice between the two alternatives:**

	Alternative 1	**Alternative 2**
Your payoff in stage 2 is 525 and R equals A	800	800
Your payoff in stage 2 is 525 and R is larger or smaller than A but not equal to A.	500	300
Your payoff in stage 2 is 30 and R equals A.	300	500
Your payoff in stage 2 is 30 and R is larger or smaller than A but not equal to A.	210	210

Stage 4:

Your decision:

- You make for each possible **A** (0,1,2,3,4,5,6,7) a decision between three alternatives – i.e. you make eight times a decision between three alternatives: **Left**, **Middle and Right** by clicking on the corresponding alternative on the computer screen. Please confirm your choice by clicking on "OK".

- Then, you choose a number for each possible **A** (i.e. you choose eight times a number):
    - When you chose **Left**, you choose a whole number between **0** and **A-1**
    - When you chose **Right**, you choose a whole number between **A+1** and **7**
    - When you chose **Middle,** you choose exactly the number **A**

  A **table** that shows you all possible numbers you can choose in stage 4 for each possible choice of alternatives in stage 4 and all values of A is attached to the instructions.
- After you have made **all sixteen** decisions, please confirm your choice by clicking on "OK".

<u>Payoff stage 4:</u>

1. You receive 105 tokens, when the number you have chosen coincides with the value **R** of the participant that is assigned to you. If there is no coincidence, you receive 20 tokens.

2. Based on your decision and the values R and A **of the participant that is assigned to you**, you receive the following payoff:

		Alternatives		
		Middle	**Right**	**Left**
Number of correct questions (R) of the selected participant	**R is smaller than *A***	315	315	1680
	R equal to *A*	1680	315	315
	R is larger than A	315	1680	315

A= Action chosen by a participant of *Experiment I*

R= Number of correct questions of a participant in *Experiment I*

Calulation of your total payoff:

Your total payoff in the experiment is given by the sum of:

- The number of all your correctly answered questions multiplied by 190 tokens and the number of wrong answers multiplied by 10 tokens (your payoff in stage 1).
- Your payoff in stage 2.
- Your payoff in stage 3.
- Your payoff in stage 4.
- In addition you receive a payment of 400 tokens.

This total payoff is converted into euros according to the exchange rate 1 Euro = 210 tokens.

References

[1] Ando, M. (2004), "Overconfidence in Economic Contests," mimeo, Nihon University.

[2] Baker, L. and Emery, R. (1993), "When Every Relationship is Above Average: Perceptions and Expectations of Divorce at the Time of Marriage," *Law and Human Behavior* **17**, 439–450.

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

[4] Bosch-Domenech, A. Montalvo, J. Nagel, R., and Satorra, A. (2002), "One, Two, (Three), Infinity, ... , Newspaper and Lab Beauty-Contest Experiments," *American Economic Review* **92** 5, 1687–1701.

[5] Brenner, L. A. , Koehler, D. J., Liberman, V., and A. Tversky (1996), "Overconfidence in Probability and Frequency Judgements: A Critical Examination," *Organizational Behavior and Human Decision Processes*, **65** 3, 212–219.

[6] Camerer, C. and Lovallo, D. (1999), "Overconfidence and Excess Entry, An Experimental Approach," *American Economic Review* **89** 1, 306–329.

[7] Croson, R. (2000), "Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play," *Journal of Economic Behavior and Organization* **41** 3, 299–314.

[8] Della Vigna, S. and Malmendier, U. (2006), "Paying Not to Go to the Gym," *American Economic Review* **93** 3, 694–719.

[9] De La Rosa, L. (2005), "Overconfidence and Moral Hazard," mimeo, University of California, Berkeley.

[10] Eichberger, J. Kelsey, D., and Schipper, B. (2005), "Ambiguity and Social Interactions," mimeo, University of California, Davis.

[11] Englmaier, F. (2004), "A Strategic Rationale for Having Overconfident Managers," mimeo, University of Munich.

[12] Fischbacher, U. (1999), "z-Tree. Toolbox for Readymade Economic Experiments," IEW Working Paper 21, University of Zurich.

[13] Frederick, S. (2005), "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives* **19** 4, 25–42.

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

[15] Greiner, B. (2004), “The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics,” Papers on Strategic Interaction, University of Cologne, Working Paper Series in Economics 10.

[16] Ho, T. Camerer, C. and K. Weigelt (1998), “Iterated Dominance and Iterated Best Response in Experimental “P-Beauty Contests,” *American Economic Review* **88** 4, 947–968.

[17] Hoelzl, E. and A. Rustichini (2005), “Overconfident: Do You Put Your Money On It?,” *The Economic Journal* **115**, 305-318.

[18] Juslin, P. (1994), “The Overconfidence Phenomenon as a Consequence of Informal Experimenter- Guided Selection of Almanac Items,” *Organizational Behavior and Human Decision Processes,* **57**, 226–246.

[19] Kahneman, D. and A. Tversky (1979), “Prospect Theory. An Analysis of Decision under Risk,” *Econometrica* **47** 2, 263–292.

[20] Kahneman, D. and A. Tversky (1972), “Subjective Probability: A Judgment of Representativeness, *Cognitive Psychology*, **3**, 430–454.

[21] Kyle, A. S. and F. A. Wang (1997), “Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?,” *Journal of Finance* **52** 5, 2073–2090.

[22] Laibson, D. (1997), “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics* **62**, 443–77.

[23] Lichtenstein, S. and B. Fischhoff (1977), “Do Those Who Know More also Know More About How Much They Know?” *Organizational Behavior and Human Performance* **20**, 159–183.

[24] Malmendier, U. and G. Tate (2005a) “Who Makes Acquisitions? CEO Overconfidence and the Market’s Reaction,” mimeo, Stanford University.

[25] Malmendier, U. and G. Tate (2005b), “CEO Overconfidence and Corporate Investment,” *Journal of Finance* **60** 6, 2661–2700.

[26] Mullen, B., J. L. Atkins, D. S. Champion, C. Edwards, D. Hardy, J. E. Story and M. Venderklok (1985), “The False Consensus Effect: A Meta Analysis of 115 Hypothesis Tests,” *Journal of Experimental Social Psychology* **21**, 263–83.

[27] Murnighana, K., Oeschb, J. and M. Pillutlac (2001), “Player Types and Self-Impression Management in Dictatorship Games: Two Experiments,” *Games and Economic Behavior* **37** 2, 388–414.

[28] Nagel, R. (1995), "Unraveling in Guessing Games: an Experimental Study," *American Economic Review* **85** 5, 1313–1326.

[29] O'Donoghue, T. and M. Rabin (1999), "Doing It Now or Later," *American Economic Review* **89** 1, 103–124.

[30] Pulford, B. and A. Colman (1997), "Overconfidence: Feedback and Item Difficulty Effects," *Personality and Individual Differences* **23** 1, 125–133.

[31] Rabin, M. (1998), "Psychology and Economics," *Journal of Economic Literature* **36**, 11–46.

[32] Rabin, M. (1995), "Moral Preferences, Moral Constraints, and Self-Serving Biases," mimeo, University of California, Berkeley.

[33] Santos-Pinto, L. (2005a), "Positive Self-Image and Incentives in Organizations," mimeo, Universidade Nova de Lisboa.

[34] Santos-Pinto, L. (2005b), "Positive Self-Image in Tournaments," mimeo, Universidade Nova de Lisboa.

[35] Shanks, D. Tonney, R. and J. McCharthy (2002), "A Re-examination of Probability Matching and Rational Choice," *Journal of Behavioral Decision Making* **15**, 233-250.

[36] Svenson, O., (1981) "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica* **47**, 143–148.

[37] Yates, J. F. (1990), "Judgement and Decision Making," Englewood Cliffs, NJ: Prentice Hall.

[38] Zizzo, D, S. Stolarz-Fantino, J. Wen and E. Fantino (2000), "A Violation of the Monotonicity Axiom: Experimental Evidence on the Conjunction Fallacy," *Journal of Economic Behavior and Organization*, **41**, 263–276.