Theory Dec. (2011) 70:475–500
DOI 10.1007/s11238-010-9199-2

Beliefs about overconfidence

Sandra Ludwig · Julia Nafziger

Abstract This experiment elicits beliefs about other people's overconfidence and abilities. We find that most people believe that others are unbiased, and only few think that others are overconfident. There is a remarkable heterogeneity between these groups. Those people who think others are underconfident or unbiased are overconfident themselves. Those who think others are overconfident are underconfident themselves. Despite this heterogeneity, people overestimate on average the abilities of others as they do their own ability. One driving force behind this result is the refusal to process information about oneself: not only does this lead to overestimation of one's own ability, but by means of social projection also to overestimation of others' abilities.

Keywords Beliefs · Belief elicitation · Bias · Overconfidence · Experimental economics

JEL Classification D83 · C91 · D01

1 Introduction

People have a hard time evaluating their own abilities objectively. Personal life is riddled with examples of overconfidence. Drivers overestimate their driving skills, students their scores in exams, couples the probability of not getting divorced. And

---

S. Ludwig (✉)
Department of Economics, University of Munich, Ludwigstrasse 28 (Rgb.), 80539 Munich, Germany
e-mail: sandra.ludwig@lrz.uni-muenchen.de

J. Nafziger
School of Economics and Management, Aarhus University, Building 1322, 8000 Aarhus C, Denmark
e-mail: jnafziger@econ.dk

also business life is riddled with examples of overconfidence. Managers overestimate the success of their merger strategies, traders are overly bullish on their investment strategies, employees overestimate their chances for a promotion. This has far reaching consequences for many economic situations.[1] For example, executive overconfidence leads to corporate investment distortions (Malmendier and Tate 2005), value-destroying mergers (Malmendier and Tate 2008), debt conservatism (Malmendier et al. 2007), or excess entry in a market (Camerer and Lovallo 1999). To limit the losses that arise from overconfidence, firms can, e.g., implement safety factors and buffer times in project planning schedules, or design incentive contracts that account for the overconfidence of its managers. But to design such features, a firm must be able to assess whether managers are overconfident. Moreover, the firm needs to know the extent of this overconfidence bias. For example, to implement appropriate buffer times, the firm must know by how much a manager underestimates the time it takes to produce a good relative to the time he announced.

In this experiment, we ask whether people actually know about other people's overconfidence and whether they are able to predict others' abilities correctly. While the persistent nature of overconfidence biases suggests that people are not capable in detecting biases in themselves, they might be able to detect them in others. For example, overconfidence about own abilities may arise because people refuse to process all available information (e.g., due to self-impression motives). When evaluating others' abilities, however, people might be more objective. Hence, we also ask, which information do people use when evaluating others? How do they form their beliefs about others?

Our baseline experiment analyzes in a very simple setting whether people know about the bias of others. In one treatment, subjects first answer multiple-choice questions. Then, they estimate their number of correct answers. We call a subject's number of correct answers her ability in the following. In another treatment, subjects first complete the same self-assessment task. Then, we elicit their belief on the bias of the subjects in the first treatment: are the others unbiased, underconfident, or overconfident? We extend this basic setup in several ways. We use words like "overestimation" in the instructions, and we show subjects the correct answers to the questions before evaluating the others. Furthermore, we elicit beliefs on the exact ability of the others, and ask subjects who they think is more likely biased (or unbiased)—they themselves or the others.

Our main result is that most subjects think that others are unbiased and only few think that others are overconfident. There is, however, a remarkable heterogeneity between these groups. Those subjects who think others are underconfident or unbiased are overconfident themselves. Those who think others are overconfident are underconfident themselves. Thus, subjects who know about biases try to correct them ("self-correction mechanism"). Those believing that underconfidence is the general bias, adjust the belief about the own ability upwards, ending up overconfident and vice versa. Whereas subjects who are unaware of biases cannot correct their own bias, and thus remain on average overconfident.

---

[1] Dunning et al. (2004) provide a discussion of the implications of overconfidence at the workplace, for education and health.

Despite this heterogeneity, people overestimate on average the abilities of others, as they do their own ability. Similar to a phenomenon known as social projection in psychology (see Allport 1924, and Krueger and Acevedo 2005), we observe that subjects project the belief they have about their own ability onto others. Thus, overconfidence about own abilities results in overconfidence about others' abilities. However, we observe that those subjects who are biased are on average slightly less biased regarding the ability of others than regarding their own ability: when evaluating themselves, people may be driven by self-impression motives and thus are less objective.

Does information which indicates the existence of an overconfidence bias help to reduce overconfidence about (others') abilities? Using words like "overestimation" in the instructions induces more people to recognize that others are biased, but not necessarily that they are overconfident. But even showing the subjects the correct answers to the questions before evaluating the others does not help much to detect others' biases. This information allows subjects to calculate their own ability, and thereby detect their own bias. Yet, subjects seem to be driven by self-impression motives: they use the information to make a very favorable guess about their own ability, and then project this guess onto others. As before, the result is overconfidence about others' abilities.

Concerning the subjects' belief about the relation between their own bias and the bias of the others, we find that the largest group of subjects thinks that they are themselves more likely to judge their ability correctly than are the others. This result is consistent with a strong self-serving bias. People do not only think that they have better abilities relative to others. But they also think that they are better in estimating their abilities than are others.

The article is structured as follows. Next, we review the related literature. In Sect. 2, we explain the experimental design and the results. After providing a brief overview of the experimental procedure and the treatments (Sects. 2.1 and 2.2), we describe our baseline experiment (Sect. 2.3). In Sect. 2.4, we describe the extension of our baseline experiment. In Sect. 3, we discuss possible drawbacks of our experiment (the selection of questions and possible hedging problems). We conclude in Sect. 4.

Related literature

Overconfidence captures unrealistic (i.e., too positive) perceptions of an individual.[2] A prominent example are "unrealistic positive views of the self", which encompasses overestimation of *absolute* abilities, as well as so called "self-serving biases" (see, e.g., Miller and Ross 1975). People who are prone to a self-serving bias, tend, for example, to ascribe their success to their own ability, while it is rather due to luck; or they overestimate their *relative* abilities (the "better-than-average effect"). For example, Svenson (1981) reports that about 80% of drivers believe to have better driving skills than the median driver. Camerer and Lovallo (1999) show how a better-than-average effect can lead to excess entry in a market. Overconfidence can also arise in the form of exaggerated perceptions of personal control ("illusion of control"). Langer

---

[2] For an overview of the psychological literature on overconfidence see, e.g., Taylor and Brown (1988), or Yates (1990).

(1975), for example, shows that people believe to have more control if they personally throw a dice (or choose a lottery ticket) than if someone else does it. Weinstein (1980) demonstrates that CEOs underestimate the failure probability of projects that they have chosen themselves. Finally, overconfidence can take the form of overoptimism about the future (see, e.g., Weinstein 1980).

In our experiment, we focus on the (over-)estimation of absolute abilities. There are different ways to measure overconfidence about the absolute ability. One way is the calibration method (see, e.g., Fischhoff et al. 1977; Alpert and Raiffa 1982; Lichtenstein et al. 1982). Subjects estimate for each (multiple-choice) question, the probability of a correct answer. A subject is unbiased, or "well calibrated", if his percentage of correct answers equals the average-stated probability. For difficult questions (see the debate outlined in Sect. 3.1), these studies observe that subjects are not well-calibrated or, more precisely, overconfident.

Economists refined this procedure (see, e.g., Hoelzl and Rustichini 2005; Blavatskyy 2009; Urbig et al. 2009). To use a neutral decision frame and to reward correct decisions, they let subjects choose between being paid according to their performance, or according to some lottery. For example, a subject who thinks her answer to a question is correct with a probability (strictly) larger than 60%, should prefer being paid according to her performance to being paid according to some lottery which pays her a high payoff (which she would get for a correct answer) with a probability of 60%. As shown, such a procedure reduces overconfidence. A second possibility to measure overconfidence is the confidence interval estimation method. Subjects state confidence intervals that should contain the true value in, e.g., 90% of the cases (see, e.g., Lichtenstein et al. 1982; Russo and Schoemaker 1992). The aforementioned studies find that most people state too narrow confidence intervals. A study by Cesarini et al. (2006), however, points out that the confidence interval method might artificially induce overconfidence and awareness of the own bias. When subjects are asked how many of their stated 90% confidence intervals actually contain the true value, they do not answer nine out of ten, but roughly six. Similarly, subjects notice that their peers state too narrow confidence intervals. Our study extends theirs by investigating whether subjects are aware of their peers' overconfidence when awareness of the own bias is not induced by the confidence interval method and by investigating possible driving forces behind the subjects' beliefs.

In contrast to the aforementioned studies, in our experiment, subjects make a point estimation, i.e., they estimate their number of correct answers. We measure overconfidence as the average deviation of the true and estimated number of correct answers (where the average is taken over all subjects), and incentivize a right estimate.[3] We believe the concept of an "average value" is considerably easier to explain and understand for the subjects who have to assess the bias of others than the concept of a "probability" or a "confidence interval".

---

[3] Note that our procedure does not allow to test for overconfidence at the individual level. A one-time divergence of the true and estimated number of correct answers could just be a mistake. Yet, if subjects are unbiased, mistakes should cancel out if one considers the average number of truly and estimated correct answers for a group of subjects. Thus, at the aggregate level, our procedure measures overconfidence in a similar vein than the probability or confidence interval estimation methods.

By eliciting the belief that people have about others' overconfidence bias, our article is closely related to the fast growing literature on belief and preference elicitation. Most of the literature focuses either on the elicitation of *own* preferences or characteristics (for an overview of this literature see, e.g., Manski 2004); or, in the context of strategic interactions, on a player's beliefs about his opponent's behavior. Here the main question is not whether beliefs are unbiased, but whether beliefs can explain observed behavior. Costa-Gomes and Weizsäcker (2008) or Nyarko and Schotter (2002), for example, analyze in normal form games whether subjects play a best response to the beliefs they state about their opponents' behavior. Similar issues are of interest in games where social preferences matter, like public good games (see, e.g., Offerman et al. 1996; Croson 2000), or trust games (see, e.g., Dufwenberg and Gneezy 2000; Guerra and Zizzo 2004; Bacharach et al. 2007). With so called beauty contests or guessing games, researchers examine how many steps of iterated elimination of dominated strategies players are able to apply (see, e.g., Bosch-Domenech et al. 2002 or Ho et al. 1998). The number of steps a player applies depends on his understanding of the concept of iterated dominance, and on his belief about the number of steps his opponents apply. When observing a player's guess, it is, however, difficult to disentangle to which level a player understands the concept of iterated dominance, and how well he predicts the ability of others (i.e., their depth of reasoning). Therefore, some studies ask participants to explain their guess to better understand their reasoning about others (see Bosch-Domenech et al. 2002 for an overview on newspaper studies).

Our approach is complementary to these studies. We elicit beliefs about others' characteristics, rather than beliefs about others' behavior. Such characteristics are usually the primitives of an economic model. Therefore, understanding people's beliefs about others' characteristics helps to better understand their beliefs about others' behavior. There are, however, only few studies that elicit beliefs about others' preferences, attitudes or characteristics. A literature strand in psychology asks how beliefs about others' personality traits (such as the look, age, gender, or social class of the other person) influence the behavior (for a review see Miller and Turnbull 1986). These studies adopt a different methodological approach than we do. They induce beliefs about others by providing subjects with information about those traits. Then they investigate, for example, whether people react to stereotypes, or whether beliefs are self-fulfilling.

Few studies elicit beliefs about others as we do. Huck and Weizsäcker (2002) investigate whether subjects are able to forecast others' choices over lotteries (i.e., their risk preferences) without bias. They show that subjects are not able to do so. In contrast to our experiment, they find no evidence that the result is driven by social projection. The beliefs of the subjects in their experiment are rather distorted by conservatism. Baker and Emery (1993) analyze people's predictions of divorce likelihoods—not only their own likelihood of getting divorced, but also that of others. They show that individuals have extremely optimistic expectations about the likelihood that they do not get divorced themselves. They, however, know quite accurately the likelihood of divorces in the population. This result is similar to one of ours: people are less biased when evaluating others. Similarly, MacDonald and Ross (1999) show that students assess the longevity of their own relationship more optimistically and less accurately than their peers and parents.

2 Design and results

2.1 Experimental procedure

We conducted the computerized experiment at the University of Bonn. We programmed the experiment with the software z-Tree (Fischbacher 2007), and recruited subjects via the internet by using the software ORSEE developed by Greiner (2004). A total of 96 subjects participated in five treatments. During the experiment, all subjects were seated at individual computer terminals. Before an experiment started, we read out loudly the instructions and the subjects answered control questions.[4] We kept the wording in the instructions neutral except for one treatment. Hence, we did not use terms like ability or overconfidence—although we use them in the following to describe the design and the results. Subjects earned tokens during the experiment, where 210 tokens = 1 Euro. Average earnings per hour were 8 Euros.

2.2 Overview of treatments

Our baseline experiment (see Sect. 2.3) consists of two treatments, A and B (strictly speaking these are two experiments), and a control treatment C. In Treatment A, the subjects (“*A subject*” or “she”) first answer several difficult multiple-choice questions. A subject’s number of correctly answered questions determines her “ability”. Then the subjects estimate their ability. In Treatment B, subjects (“*B subject*” or “he”) first complete the same tasks as the *A subjects*. We then elicit their belief about the direction of the *A subjects’* bias: do they think that the *A subjects’* ability is *on average* higher, lower or equal than the *A subjects’* estimate of their ability? In Treatment C, we elicit the subjects’ belief about the direction of the *A subjects’* bias as in Treatment B. However, subjects do not complete the self-assessment task: they only read through the multiple-choice questions.

Treatments W and AT extend the baseline experiment (see Sect. 2.4). First, we ask the *W subjects* not only about the direction of the *A subjects’* bias, but also about the true ability of the *A subjects*. Second, we elicit their beliefs about the relation of the own and the others’ (potential) bias. Third, we use value-laden terms in the instructions. Fourth, Treatment W additionally includes decision tasks that refer to a different type of questions (tricky ones). The *W subjects* receive information about the correct answers to these tricky questions before evaluating other subjects who answered these tricky questions (the *AT subjects* in Treatment AT). Treatment AT uses the tricky questions—it is otherwise identical to Treatment A.

2.3 Eliciting beliefs about biases

2.3.1 Treatment A

Treatment A consists of two stages. In the first stage, all 20 *A subjects* answer seven multiple-choice questions from different fields of general knowledge. Questions are

---

[4] The instructions are given in the Appendix.

Table 1 Payoff table to elicit a subject's belief about her number of correct answers in Treatment A

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

difficult. Subjects should either know an answer more or less for sure, or not at all. For each correct answer, a subject earns 190 tokens.

In the second stage, the *A subjects* estimate their number of correct answers (their ability). The procedure to elicit their estimate is as follows. Subjects get a payoff table (see Table 1). Each subject then chooses one out of eight actions $q \in \{0, 1 \dots 7\}$. As illustrated in Table 1, for any number of correct answers $t \in \{0, 1 \dots 7\}$, a subject receives a high payoff if $q = t$ and a low one otherwise. Thus, a subject who believes she has (most likely) t answers correct, should choose action $q = t$. In the sequel, we call the chosen action q a subject's *estimate* (or *belief*).[5]

We want to ensure that subjects do not use a hedging strategy. For example, a subject might deliberately answer no question correctly in the first stage to make for sure the correct estimate in the second stage. Thus, we implement the following two-step procedure. When answering the questions, an *A subject* knows that she has to make a decision afterwards. She also knows that her payoff for this decision depends on her number of correctly answered questions. But she does not know the exact decision and the relevant payoff table.

2.3.2 Treatment B

Treatment B is also divided into two stages. First, 17 *B subjects* answer the same multiple-choice questions as the subjects in Treatment A. Then, at the second stage, they estimate their ability, receive information about Treatment A, and assess the bias of the *A subjects*. Thus, the procedure is the same as in Treatment A. Subjects do not get to know the exact task of the second stage when answering the questions.

We now describe the second stage in more detail. First, we explain the *B subjects* the full procedure of Treatment A. Second, we inform them (after they estimated their ability) about the *A subjects'* average estimate q, denoted by $\bar{q}$ in the following. This information makes the task of evaluating the *A subjects* similar to the task of evaluating oneself.

---

[5] We implement this neutral procedure to not influence the beliefs of *B subjects* (who get the information about Treatment A) by words like "estimate your ability". A possible concern may be that this procedure is somewhat complex. To make sure that subjects understand the payoff table, they have to answer control questions. In addition, in Treatment W (cf. Sect. 2.4), we simply ask subjects how many questions they think they answered correctly. Stated beliefs do not differ compared to Treatment A (Mann–Whitney U test, $p = 0.688$, two-tailed).

Table 2 Payoff table to elicit the *B subjects'* beliefs on the direction of the *A subjects'* bias

	Action		
	Left	Middle	Right
$\bar{t} < \bar{q} - 0.5$	315	315	1,680
$\bar{q} - 0.5 \leq \bar{t} \leq \bar{q} + 0.5$	315	1,680	315
$\bar{t} > \bar{q} + 0.5$	1,680	315	315

Afterwards, the *B subjects* evaluate whether the *A subjects* overestimate their average ability ($\bar{t}$) by at least 0.5 ($\bar{t} - \bar{q} < -0.5$), underestimate it ($\bar{t} - \bar{q} > 0.5$), or estimate it correctly (i.e., $-0.5 \leq \bar{t} - \bar{q} \leq 0.5$). We refer to the former two alternatives as a belief that the *A subjects* are biased (namely over-, or underconfident). The latter mirrors a belief in unbiasedness. The procedure to elicit the belief of a *B subject* about the *A subjects'* bias is similar to before. Subjects get a payoff table (cf. Table 2). They can choose between three neutrally named actions. For any of the three possible classifications of the *A subjects'* bias, exactly one action yields a high payoff.[6] Hence, to achieve the high payoff, a subject who puts the highest belief on, e.g., the possibility that the others overestimate their ability, should choose the action *left* in Table 2.

2.3.3 Results baseline treatments

Only 25% of the *A subjects* correctly estimate their ability, while 60% overestimate and 15% underestimate it. Moreover, the *A subjects* overestimate their ability on average, i.e., they are overconfident in the given task and for the given questions. Their bias is -1.1, where we measure the bias by $\bar{t} - \bar{q}$ ($\bar{t} = 2.25$ and $\bar{q} = 3.35$). According to a Wilcoxon signed rank test, the medians of the values t and q differ significantly ($p = 0.006$, two-tailed). The distribution of the values of t and q in Fig. 1 illustrates the difference. The t-distribution is left skewed, whereas the q-distribution is right skewed.[7]

Do the *B subjects* know about this bias? The answer is no. 59% say that the *A subjects'* estimates are correct (i.e., others are "unbiased") and only 23% think that the *A subjects* overestimate their ability on average.

In the control treatment (Treatment C), 20 subjects evaluate the *A subjects'* bias, but do not answer the questions and estimate their own ability (we just show them the questions). The result is more pronounced, but not significantly different[8]; 75% think that others are unbiased, and only 15% think that the *A subjects* overestimate their ability.

---

[6] Again, subjects have to answer control questions to show that they understand the procedure.

[7] Treatment A is not an outlier. In all treatments, in which subjects estimated their own t for the difficult questions, subjects are on average overconfident (the bias varies from -1.4 to -0.5).

[8] According to a Fisher exact test, the fractions of subjects in Treatments B and C, who think that the *A subjects* are unbiased or biased, do not differ significantly ($p = 0.234$, one-tailed).

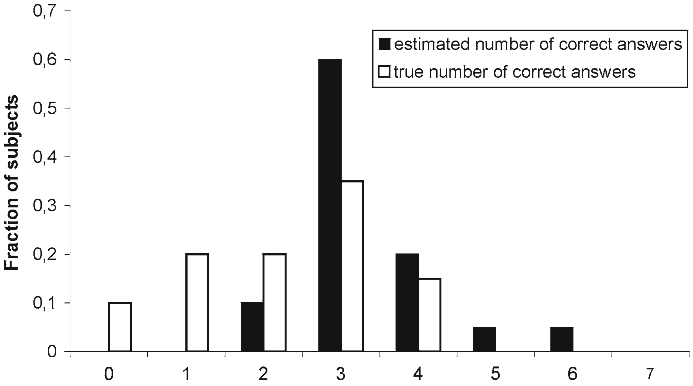

Fig. 1 Distribution of true and estimated number of correct answers in Treatment A

2.4 Information and belief formation: Treatment W

In the following, we ask how subjects form their beliefs about others (and themselves): how do subjects use information to form their beliefs? And how does their belief about others relate to their belief about themselves and their own bias? For this, we consider an extension of our baseline experiment—Treatment W with 19 subjects.

Treatment W consists of two stages. Each stage is divided into two parts, where each part is associated with a different block of questions. One block consists of the seven difficult questions, we used in Treatment A, B, and C. The other block consists of seven tricky questions. At the first stage, subjects answer both blocks of questions. Like in the other treatments, they do not know the exact tasks of the second stage when they answer these questions. At the second stage, they make their decisions (like estimating their ability and the one of the others), for each block of questions. Subjects do not receive any information on results or payoffs of the first decision part before the second part is finished. To avoid hedging, we randomly select one block of questions and one block of decisions for the payment.

The first decision part of Treatment W is associated with the difficult multiple choice questions and is analogous to Treatment B. However, we change the information content of the instructions by using a non-neutral wording. For example, we explicitly ask the *W subjects* "Do you think that others over-, underestimate their number of correct answers, or estimate it correctly?". In addition, we elicit the subjects' belief about the exact ability of the *A subjects*. We ask subjects "How many questions do you think the others answered on average correctly?". They state a number between 0 and 7 with one decimal place. Their number has to be consistent with their statement whether they think others are over-, underconfident or unbiased. If their statement is roughly correct (by ±0.5), they receive a high payoff, and otherwise a low one. Moreover, we elicit each subjects' belief on who is more likely correct or biased: we ask the subjects whether (i) their own estimate and the average estimate of the others are correct, (ii) their own estimate is better (i.e., their own estimate is correct while the others are wrong), (iii) the average estimate of the others is better, or (iv) both estimates are wrong. If their statement is correct, they receive a high payoff, and otherwise a low one.

Table 3 Overview of subjects' self-assessment (in %) with difficult/tricky questions in the different treatments

	Treatment				
	Difficult questions			Tricky questions	
	A	B	W	AT	W
Correct	25	6	16	10	5
Overestimation	60	76	53	90	90
Underestimation	15	18	31	0	5

The second decision part of Treatment W is almost identical to the first, except for two changes. First, it refers to a different set of questions. We thus confront the *W subjects* with a different reference treatment (Treatment AT). In Treatment AT, 20 subjects (*AT subjects*) answer tricky multiple-choice questions, else Treatment AT is exactly like Treatment A. Second, we show the *W subjects* the correct answers to these questions. We provide them with the answers after they answered the questions themselves, estimated their own number of correct answers, and stated who is more likely correct, but before they classify the *AT subjects'* bias and estimate the *A subjects'* ability. We want the information about the correct answers to be a strong signal. Therefore, we deliberately choose tricky multiple-choice questions that induce overconfidence. They look very simple, but are in fact not: the answer that seems to be the correct one, is actually wrong.

Unsurprisingly, the tricky questions lead to more pronounced overestimation. 90% of the subjects overestimate their ability. Table 3 summarizes our findings on the overestimation of own abilities. It shows the proportions of subjects who estimate their ability correctly, over-, or underestimate it in the various treatments and for the different sets of questions. Testing for differences in the subjects' self-assessment between treatments (for the same type of questions) indicates no significant differences.[9]

2.4.1 The impact of information

We observe that the information provided through the different wording in the instructions increases the awareness about the bias of others. For the difficult questions, only 42% of the subjects believe that the *A subjects* are unbiased. 32% (26%) think that they overestimate (underestimate) their number of correct answers.[10] Thus, the information leads to the insight that the *A subjects* are biased, but not that they are actually overconfident. Moreover, the *W subjects'* average belief about the ability of the *A subjects* is only 0.05 units smaller than $\bar{q}$. This is much larger than the average ability of the *A subjects*. Hence, people are also biased when evaluating the ability of others. For the tricky questions, 91% of the *W subjects* believe that the *AT subjects* are overconfident.

---

[9] Difficult questions: testing for differences between Treatments A and B/A and W/B and W, respectively, using a Chi-square test yields $p = 0.289/0.437/0.319$. Tricky questions: testing for differences between Treatments AT and W using a Fisher exact test yields $p = 1.00$.

[10] According to a Chi-square test, more subjects in Treatment W than in Treatment C think that others are biased, $p = 0.049$. There is no significant difference between Treatments W and B.

Yet, they do not recognize the extent of the overconfidence bias—although they get to know the correct answers. They think that the average ability of the *AT subjects* is about 2.9. Indeed, it is only 1.2 and the *AT subjects* believe it is 4.6. Median tests indicate that the *W subjects'* guesses are significantly smaller than the *AT subjects'* average belief of 4.6 ($p = 0.0002$) and larger than the true average of 1.2 ($p = 0.0001$).[11]

2.4.2 Relation between beliefs about oneself and others

In this subsection, we ask how subjects' beliefs about others relate to the belief about their own ability or their own bias. This analysis does not only help to better understand the belief people hold about others, but also some driving forces behind people's overconfidence.

Self-correction mechanism Asking how the belief about others is related to a *W subject's* own bias helps to shed some light on the overconfidence phenomenon. Do those subjects, who believe that others are biased, correct the belief about themselves accordingly (self-correction mechanism)? For the difficult questions,[12] we observe that those *W subjects*, who think that the *A subjects* are overconfident, overestimate their own ability by 1.2 on average. They adjust their belief too much upwards. Thus, they end up being heavily overconfident. Those subjects, who say that the *A subjects* are roughly correct, overestimate their own ability by 0.75 on average. They do not recognize the bias of others, do not adjust the own belief, and therefore, end up being overconfident. They, however, are less overconfident than those subjects who think that others are underconfident. And finally those, who say that the *A subjects* are overconfident, are in fact underconfident with a bias of 0.5. Being aware of the overconfidence phenomenon, they try to correct their belief so as not to end up overconfident. But they over-correct and end up underconfident. Specifically, we observe that the estimate q of those subjects, who say that others are overconfident (underconfident), is significantly larger (smaller) than the estimate q of those, who say that others are unbiased (Mann–Whitney U tests $p = 0.004$ and $p = 0.011$, respectively).[13]

Figure 2 illustrates this relationship graphically. The figure plots three cumulative distribution functions of *W subjects'* biases—one function for *W subjects* who think that *A subjects* are unbiased, overconfident, and underconfident, respectively. We observe that the cumulative distribution function of those subjects who think that the *A subjects* are overconfident lies below the distribution functions for those who think that others are unbiased, or underconfident.

---

[11] Psychologist show that outcome feedback does not significantly reduce the bias of people (see, e.g., Pulford and Colman 1997). Our results show in addition that such feedback does not reduce the bias people have when evaluating others.

[12] As shown before, with the tricky questions 91% of the *W subjects* recognize that the *AT subjects* are overconfident. As this is induced by the exogenous information (seeing the correct answers), a belief in others' overconfidence does not relate to the own bias.

[13] The pattern in Treatment B is similar (those who think that others are underconfident are most heavily overconfident). However, differences between those who think others are overconfident (underconfident) and those who think they are unbiased are not significant. As most subjects think others are unbiased, there are only few observations of people thinking that others are biased.

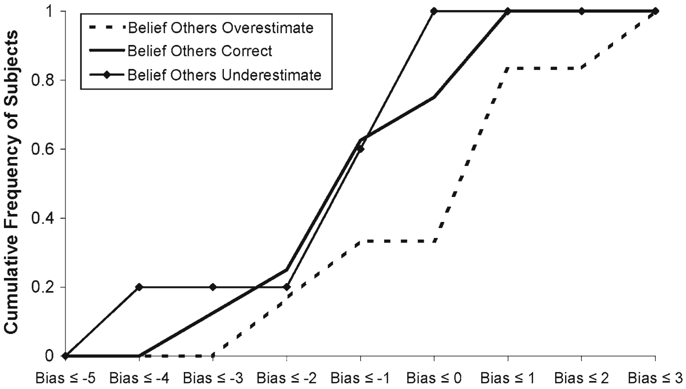

Fig. 2 *W subjects*' own bias given their belief about the bias in the *A subjects'* self-assessment. A negative bias refers to overconfidence and a positive bias to underconfidence

Social projection How do subjects form their (biased) belief about others? Relating the own belief to the belief a subject holds about others helps to shed some light on the formation of this belief. We observe that subjects seem to use the information they have about themselves ("social projection"[14]). For the difficult questions, the belief about the own ability and the estimate about the others' average ability are positively correlated (Spearman rank-order correlation test, Spearman's rho: 0.737, $p = 0.0002$). Figure 3 illustrates this observation. It shows that *W subjects* with lower beliefs q (up to 3) guess on average that the *A subjects'* ability is lower than $\bar{q} = 3.4$. Subjects with higher beliefs (from 4 on) estimate that the *A subjects'* ability is larger than 3.4. Hence, overall, social projection can explain the bias in the evaluation of others as subjects are biased themselves.

Is this also true for the tricky questions? Recall that we show the *W subjects* the correct answers to the questions before they evaluate the *AT subjects* who answered the tricky questions (but after they evaluate themselves). How do the *W subjects* use the information the observation of the correct answers provides? We observe that the belief about the own ability is no longer correlated with the belief about the others' ability. The own ability, however, is (Spearman's rho: 0.488, $p = 0.034$). This indicates that *W subjects* use the information to roughly calculate their own ability. Then they take this number as the basis for their belief about the *AT subjects*. Again, subjects conclude from their own results on the results of others. Yet, it is surprising that they do not adjust their belief about others more. Their own average ability is just 1.5, which is significantly smaller than the belief of 2.9 they state for the others (Wilcoxon signed rank test, $p = 0.001$, two-tailed). From the nature of the questions, and their own results they should have noticed that it is impossible for the others to answer on average 2.9 questions correctly. Self-impression motives may be a driving force

---

[14] Similar to social projection, is the (false) consensus effect (compare Mullen et al. 1985, or Engelmann and Strobel 2000). People who are prone to a false consensus effect tend to overestimate the degree to which their own behavior or beliefs are shared by others. In our experiment, the false consensus effect would imply that subjects overestimate the frequency with which their own estimate q is present in the population. While our results hint in this direction, we cannot test whether the subjects indeed *over*estimate the frequency.

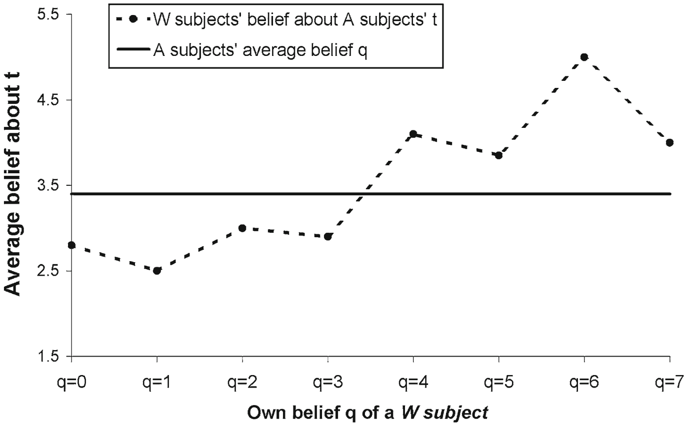

Fig. 3 Average belief about the *A subjects'* average ability $\bar{t}$ given the own belief q of a *W subject*

behind this result. Subjects do not want to know or admit that they had such mistaken beliefs about themselves. Hence, they do not really calculate their own number of correct answers, but make some "favorable guess". They then use this favorable guess for the evaluation of the others. Hence, subjects do not fully process the available information.

Self-impression motives The previous analysis showed that people overestimate their own ability and that this causes, by social projection, also a bias in the evaluation of other people. We now analyze the relationship between these biases in more detail. Are people more or less objective when evaluating others? When estimating the own and the others' ability for the difficult questions in Treatment W, subjects have similar information about themselves than about the others (their own belief q and the average belief $\bar{q}$ of the others). However, people know themselves better than they know others. This knowledge should make the evaluation of others more difficult. Hence, one expects that subjects are more biased when evaluating others than when evaluating themselves. The evidence on this is rather mixed. Our results show that those subjects who are overconfident themselves are on average less overconfident when evaluating the others. Similarly, those subjects who are underconfident are less underconfident when evaluating the others (actually they are slightly overconfident).[15] This indicates that biased subjects do not fully process the information they have about themselves. One possible reason for this behavior are self-impression motives. People often want to impress themselves and thus ignore information. For example, a subject may overstate his estimate to impress himself how "smart" he is, and thus ends up being overconfident (cf. e.g., Bodner and Prelec 2003). When evaluating others no such concerns arise and thus people are more objective. In contrast, the few subjects who are unbiased overestimate the others' ability on average. These subjects seem to process the information about themselves correctly (as they are unbiased). Since they have worse information on the others than on themselves, they are more biased

---

[15] For those subjects who are under-/overconfident themselves the own bias and the bias about others differ significantly: $p = 0.014/0.111$, Wilcoxon signed-rank tests, one-tailed.

about the others' ability.[16] Overall, the *W subjects'* own absolute bias is by 0.26 larger (though not significantly, Wilcoxon signed-rank test, $p = 0.199$, one-tailed) than the bias in assessing the others' average ability. Thus, subjects are not generally worse in evaluating others' abilities.[17]

Self-serving bias Finally, we analyze a *W subject's* belief about who he thinks is more likely correct or biased – he himself or the others (i.e., the *A subjects* when the difficult questions are considered, or the *AT subjects* for the tricky questions). The largest group of the *W subjects* thinks that it is more likely that they do not make a mistake themselves while the others are biased (37% for the difficult questions; 32% for the tricky questions). Only about a sixth of the subjects thinks that the others are unbiased, while they make a mistake themselves (16% for both the difficult as well as the tricky questions); 16% (32 %) think both are biased, and 31% (21%) think both are unbiased for the difficult (tricky) questions. Overall, the majority of the subjects thinks that they themselves are correct (69% for the difficult questions; 53% for the tricky ones).

These observations provide evidence for a very strong better-than-average effect. If one believes that people are unbiased, one should rather state that the others are on average correct and oneself makes a mistake. If mistakes are random, i.e., if people are unbiased, mistakes should cancel out for the population, but not for a single subject. Hence, individuals do not only think that they have better abilities relative to others (as psychologist observe), but also that they are better in judging their own ability than are others in this task.

Finally, the relation between a subject's own bias and the comparative evaluation of biases provides some further evidence that subjects try to correct their biases. Those who say that they are more likely correct than are the *A subjects* are on average more overconfident than those subjects who do not state this (the biases for the two groups are -1 versus -0.17 for the difficult questions and -2 versus -1.62 for the tricky questions).

3 Discussion

3.1 Selection of questions

There is a lively debate whether overconfidence is a stable trait or induced by certain questions. Lichtenstein and Fischhoff (1977) show that people are overconfident for difficult questions, but underconfident for easy ones (the so called "hard-easy effect"). Gigerenzer (1993) argues that only the biased selection of questions matters for overconfidence, while the bias vanishes under random sampling irrespective of the

---

[16] There are, however, too few observation to test whether the bias is significantly larger.

[17] For the tricky questions the *W subjects'* observe the correct answers after assessing the own ability but before assessing the *AT subjects'* $\bar{t}$. Given this additional information, we cannot hypothesize whether it is more or less difficult to evaluate the others. However, our results again indicate that subjects are not better in evaluating themselves. All but two subjects are overconfident when evaluating themselves. And these overconfident subjects are on average less overconfident when evaluating the others (Wilcoxon signed-rank test, $p = 0.118$, one-tailed).

questions' level of difficulty. Brenner et al. (1996) provide an overview of empirical studies confuting this view. They conclude that overconfidence persists even under random sampling. Yet, whether or not a specific set of questions induces overconfidence is not our question. Instead, we want to test whether *B subjects* realize for the *given* questions (for which people turn out to be overconfident) that others are overconfident. We also do not want to claim that subjects are "more overconfident" for the tricky questions. We just use them—together with the information about the correct answers—as a strong signal about the existence of biases.

Nonetheless, subjects' beliefs about *how* we selected the questions—in a special way or randomly—might drive results. Thinking that we randomly selected the questions could cause the belief that others are unbiased (if subjects think along the lines of Gigerenzer 1993 that overconfidence vanishes for a random selection). Yet, disproving this explanation, we observe a significant difference between the number of people who state that others are biased/unbiased in Treatments C and W (cf. Sect. 2.4.1). If the belief about how we selected the questions drove the belief that others are unbiased, we should see no difference.

3.2 Incentive compatibility

Similar to the procedures employed by Hoelzl and Rustichini (2005), Blavatskyy (2009), or Urbig et al. (2009), we reward subjects for correct decisions and implement a neutral decision frame. Our payoff structure for the various decision tasks ensures that even risk averse subjects select the alternative on which they place the highest probability (a similar payoff structure is, e.g., used by Wilcox and Feltovich 2000; Bhatt and Camerer 2005).[18] Moreover, by using neutrally framed payment tables, like Table 1 or 2, we do not have to ask a subject directly "how many questions do you think you have correct", or "do you think that others overestimate their ability" (see Treatments A, B, and C).

However, as in many experiments on belief elicitation (for a recent overview see, e.g., Blanco et al. 2008), potential hedging problems may arise in our experiment. First, in all treatments (except Treatment C), a subject might answer no question correctly to make for sure the correct estimate about her ability. As this hedging strategy seems obvious, we avoided it by implementing in all treatments the aforementioned two-step procedure (see Sect. 2).[19]

Second, in Treatments B and W, a subject might employ a hedging strategy between the decisions "estimate your own ability", "what is the others' bias", "estimate the others' ability", and the statement on the relative bias. Here a caveat applies as we did not

---

[18] In comparison, a quadratic scoring rule has the advantage that larger deviations are punished more, but it requires risk neutrality (for more details on scoring rules see, e.g., Savage 1971). The latter is likely to be violated even for small stakes (see, e.g., Holt and Laury 2002; Rabin 2000).

[19] An alternative to this two-step procedure, in which subjects do not get to know the payoff table when answering the questions, would have been the "hidden chaining" procedure used by Blavatskyy (2009). Before subjects answer the questions, he shows them a payoff table for a decision task that is conducted at the second stage. However, he does not tell them that this decision task relates to the number of correctly answered questions.

eliminate all possible hedging strategies, but only the obvious ones. As mentioned, we explicitly prevented hedging possibilities for the decisions “what is the others’ bias” and “estimate the others’ ability” (the estimate of the others’ ability has to be consistent with the statement regarding the others’ bias), and for the two decision parts in Treatment W (one part is randomly selected). Yet, subjects could hedge their stated beliefs for the tasks “estimate your own ability”, “what is the others’ bias”, and the statement on the relative bias. Here we implemented a more indirect procedure that is in line with many experimental studies on belief elicitation. We kept the payoffs for the statement of the relative bias and the estimation of one’s own ability low compared to the payoffs for estimating the direction of the others’ bias.[20]

This procedure alleviates the hedging problem, but might not totally prevent it. However, it is hard to come up with a hedging strategy for the decision tasks “estimate your own ability” and “what is the others’ bias”. The results of Treatments B and C confirm this. In Treatment C, subjects did not estimate their own ability, but only the bias of the others, while in Treatment B they did both. As mentioned, there is no significant difference in the number of subjects who state that others are unbiased/biased. Hedging between the estimation of the others’ bias and the statement on the relative bias in Treatment W might be more of an issue. For example, a subject could partially insure himself against a low payoff by stating “the others are unbiased” and “I think my estimate, as well as the estimate of the others is wrong”. However, we observe for the difficult questions in Treatment W that only one subject made such an “inconsistent statement”[21]—indicating that subjects are unlikely to recognize or exploit (complicated) hedging possibilities.[22] This supposition is also supported by a recent study by Blanco et al. (2008). They examine explicitly whether subjects understand and exploit hedging possibilities between different interrelated tasks. The study suggests that subjects are unlikely to do so.

4 Conclusion

Overconfidence is an everyday life phenomenon. People overestimate their driving abilities, students their scores in exams, couples the likelihood of not getting divorced, employees their chances of a promotion, managers the success of their investment and merger strategies. Most people, however, are not aware of this bias. We observe that people do not only overestimate their own abilities, but also abilities of others, and do not detect others’ overconfidence bias. We examine the formation of this belief about

---

[20] For example, Rutström and Wilcox (1996) [p. 11] state that “The maximum payoff for each stated belief [. . .] is deliberately low to make belief statements less interesting in expected payoff terms than strategy choice itself, which is typical of designs like this using a scoring rule.”

[21] Concerning the decision tasks that refer to the tricky questions, subjects received information on the correct answers to the questions after they estimated their own ability and made the statement on the relative bias, but before they stated their belief regrading the others’ bias. Thus, “inconsistent statements” arise more often—they, however, are induced by the exogenous information.

[22] For the tasks “estimate your own ability” and the statement on the relative bias, we observe that the ability estimates do not differ between treatments in which subjects complete both tasks, and the ones where they only estimate their ability. Hence, hedging for these two tasks seems not to be an issue.

others: how does it depend on the available information? How do beliefs about other people's characteristics (i.e., their abilities and bias) depend on own characteristics? We observe a strong relation between the belief people hold about themselves or their own bias and the belief they hold about others. We discuss how these results are driven by a self-correction mechanism, social projection, and self-impression motives.

Our results have implications for the theoretical modeling of interactions between overconfident players. In recent years, a literature emerged that includes overconfidence biases in theoretical models asking about, for example, the optimal incentive contract for overconfident agents (Santos-Pinto 2008; de la Rosa 2007), the behavior of overconfident agents in tournaments (Santos-Pinto 2007), team production (Gervais and Goldstein 2007), or in oligopolistic markets (Eichberger et al. 2009; Englmaier 2004). A typical assumption in such models is that people know others' biases. Our results, however, give quite pessimistic predictions regarding the knowledge about other people's (like an opponent's) bias.

Based on our results it seems important to examine the un-/awareness of others' biases in real organizations and markets. For example, the awareness of biases in individual judgments that lead to costly mistakes (like overly optimistic estimates of planning horizons, construction costs, or material properties) may induce firms to implement safety factors or buffer times. The employment of such repairs indicates that firms are to some degree aware of individual's shortcomings like overconfidence.[23] However, frequent delays in production (like for the Airbus A380, or the 787 Dreamliner of Boeing[24]) show that such repairs are often insufficient. That is, firms are not aware of the full extent of the bias. Furthermore, while empirical evidence shows that overconfident managers have a negative impact on the firm's profits (Malmendier and Tate 2005; Malmendier et al. 2007; Malmendier and Tate 2008), theoretical models suggest that there are also positive aspects of employing them (compare, e.g., Englmaier 2004; Goel and Thakor 2008). Hence, the question arises whether managers are selected systematically with respect to their overconfidence bias, or whether shareholders are simply not aware of the bias and its possible positive or negative implications for the firm.

Acknowledgments We would like to thank Johannes Abeler, Ethan Cohen-Cole, Leonidas Enrique de la Rosa, Simon Gächter, Paul Heidhues, Denis Hilton, Erik Hölzl, Hannah Hörisch, Steffen Huck, Philipp Kircher, Georg Kirchsteiger, Alexander Koch, Martin Kocher, David Laibson, Ulrike Malmendier, Felix Marklein, Luis Santos-Pinto, Drazen Prelec, Burkhard Schipper, Karl Schlag, Matthias Sutter, and an anonymous referee for helpful comments and discussions. Financial support from the Bonn Graduate School of Economics is gratefully acknowledged.

---

[23] Heath et al. (1998) provide examples of organizational practices that may repair individual shortcomings like, for example, overconfidence (so called "cognitive repairs") and discuss their effects.

[24] The first A380 was delivered 20 months later than Airbus intended (see, e.g., http://news.bbc.co.uk/2/hi/busi\discretionary-ness/5405524.stm). In June 2009, Boeing announced the fifth delay of the maiden flight of its Dreamliner 787 (see, e.g., http://news.bbc.co.uk/2/hi/busi\discretionary-ness/8115147.stm).

Appendix: Instructions

(translated; original instructions are in German).

General information in all treatments

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

Instructions Treatment A and Treatment AT

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 receive the instructions for stage 2 after having answered the 7 questions.

Stage 1:

- You have to answer 7 multiple-choice questions. For each question there are 4 possible answers to choose from. Exactly **one** of these possible answers is correct. You select your answer 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 a question. During these 45 seconds you can give your answer at any time. The time that is left is displayed. When the time has run out, the next question is shown automatically.
- **Please note:** If you do not select an answer or do **not** click OK before the time has run out, this is equivalent to a wrong answer.
- Once you have answered all questions, the computer calculates how many questions you answered correctly. **You receive the information how many correct answers you have after the experiment, i.e., after stage 2.**

Payoff stage 1:
For each correct answer you receive 190 tokens and for each wrong one you receive 10 tokens.

Stage 2: [*Instructions for stage 2 are distributed after stage 1 is finished*]
In stage 2 you choose one out of eight possible **actions 0, 1, 2, 3, 4, 5, 6, 7** by entering one of these numbers in the corresponding cell on the screen and you confirm your choice by clicking "OK".

Payoff stage 2:

	Number of correct answers							
	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 3	30	30	525	30	30	30	30	30
Action 4	30	30	30	525	30	30	30	30
Action 5	30	30	30	30	525	30	30	30
Action 6	30	30	30	30	30	30	525	30
Action 7	30	30	30	30	30	30	30	525

The table shows the payoffs that you receive depending on your choice and on your number of correct answers in stage 1. **You are not told how many questions you answered correctly before the experiment is finished.**

Your total payoff:
Your total payoff in 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 the action you have chosen (your payoff in stage 2). In addition you receive 525 tokens.

Instructions Treatment B

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 this choice 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 depends on Experiment I. You receive the instructions for stages 2, 3 and 4 after having answered the 7 questions.

Stage 1:
[*Exactly as in Treatment A.*]

Stage 2: [*Instructions for stages 2, 3, and 4 are distributed after stage 1*]
[*Exactly as in Treatment A.*]

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 the choice between actions 0, 1, 2, 3, 4, 5, 6, 7 in stage 2. In 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.
Relevant results from Experiment I:
Based on the answers and the decisions of the participants in 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 to 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 to one decimal place. Thus, the average action can also take values from 0 to 7 in steps of 0.1.
   Before you make your decisions in stages 3 and 4, **you are told the value of the average action (A)** chosen by the participants in Experiment I.

[*Stage 3 is not discussed in the current paper. It includes a decision about two alternatives that relate the accuracy of a subject's own estimate in stage 2 to the accuracy of the estimate of participants in Experiment I.*]

Stage 4:
You choose between three alternatives: **left, middle and right** by clicking on the corresponding alternative on the screen. You confirm your choice by clicking "OK".

Payoff stage 4:

	Actions		
	Left	Middle	Right
$R < A - 0.5$	315	315	1680
$A - 0.5 \leq R \leq A + 0.5$	315	1680	315
$R > A - 0.5$	1680	315	315

A = Value of the average action of the participants in Experiment I
R = Value of the average number of correct questions in Experiment I
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 payoffs in stages 2, 3 and 4.
- In addition you receive 725 tokens.

Instructions Treatment W—Part I

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) has been answered and decision 1A (1B) has 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:
[*As in Treatment A except that subjects answer two different blocks of 7 multiple-choice questions (the hard and the tricky questions). Subjects are paid as in Treatments A and B for one of the two blocks of questions that is randomly selected in the end.*]

Instructions Treatment W—Part II [*distributed after Part I is finished*]

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 "OK".

Payoff decision 1A:
If your statement coincides with the actual number of correctly answered questions in block A ("Your estimate is correct"), you receive 525 tokens, if it does not coincide ("Your estimate is not correct"), you receive 30 tokens.

Relevant information on Experiments I and II:
[*Description equivalent to Treatment B just for Experiments I and II and instead of "average action" of the participants we say "average estimate".*]

Decisions 2A, 3A, and 4A:
Before you make decisions 2A, 3A, and 4A, **you are told the value of the average estimate (E)** of the participants in 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 estimates are correct": your estimate 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 estimate is better”: your estimate is correct and the distance between E and R in Experiment I is larger than 0.5.
- “Average estimate is better”: your estimate is wrong and the distance between E and R in Experiment I is smaller than or equal to 0.5.
- “Both estimates are wrong”: your estimate is wrong 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 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.**

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 questions (R) in Experiment I is by more than **0.5 smaller** than the average estimate (E).
- The participants estimate their actual number of correctly answered questions on average almost **correct**. This means that the average number of correct questions (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 questions (R) in Experiment I is by more than **0.5 larger** than the average estimate (E).
  You choose between the three alternatives **(overestimate, correct, underestimate)** by clicking on the corresponding alternative and confirming with OK.

Payoff decision 3A:
When the alternative you have chosen is true, you receive 1680 tokens, when it is not true, you receive 315 tokens.

Decision 4A:
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.

Payoff decision 4A:
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, you receive 105 tokens, otherwise you receive 20 tokens. [Explanation of distance as before.]

Decisions 1B–4B [*No information on results in decisions 1A-4A is provided at this point.*]

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

Your total payoff in the experiment:

- 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 for 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.

Questions (Treatments A, B, and W):

(+) indicates correct answer

When did the Holy Roman Empire of the German Nation stop existing?
– 1618/1918/1815/1806 (+)
Which frequency has home power in middle Europe?
– 220 volt/110 volt/60 hertz/50 hertz (+)
Who wrote "Iphigenie auf Aulis"?
– Goethe/Euripides (+)/Schiller/Sophokles
How many symphonies did Joseph Haydn write?
– 104 (+)/41/21/9
Which one is no chemical element?
– selenium/calcium/arsenic/americium
How do you call the dark spots of the moon?
– Mare (+)/Mire/Mure/More
Which boxers fought the "Rumble in the Jungle"?
– Joe Frazier and George Foreman/George Foreman and Muhammed Ali (+)/ Evander Holyfield and Mike Tyson/Muhammed Ali and Joe Frazier

Tricky questions (Treatments AT and W):

Which food has the most kilocalories per 100g?
– crispbread (+)/apple/zucchini/cured eel

Which of these countries conveyed the most crude oil in 2001?
– Thailand/Angola/China (+)/United Arab Emirates
Which city (including outskirts) has the most inhabitants?
– Madrid (Spain) (+)/Prague (Czech Republic)/Bangalore (India)/Munich (Germany)
Which of these mountains is the highest?
– Olymp/Drachenfels/Zugspitze/Etna (+)
Which mature animal (male) weighs most on average?
– Australian koala bear/German shepherd/Belgian carthorse (+)/West African giraffe
Who has/had his following "title" for the longest period?
– Helmut Kohl: chancellor/Johannes Paul II: pope/Arnold Schwarzenegger: "Terminator"/Franz Beckenbauer: "emperor‘" (+)
In which country is the tea consumption per capita the highest (data from 1998)?
– Paraguay (+)/Italy/India/Bahamas

References

Allport, F. H. (1924). *Social psychology*. Boston, MA: Houghton Mifflin.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Bacharach, M., Guerra, G., & Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision, 63*(4), 349–388.

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

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior, 52*(2), 424–459.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2008). Belief elicitation in experiments: Is there a hedging problem? IZA discussion papers 3517, Institute for the Study of Labor (IZA).

Blavatskyy, P. R. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty, 38*(1), 39–49.

Bodner, R., & Prelec, D. (2003). Self-signalling and diagnostic utility in everyday decision making. In I. Brocas & J. D. Carrillo, *The psychology of economic decisions, vol. 1: Rationality and well-being* (pp. 105–126). New York: Oxford University Press.

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

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

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

Cesarini, D., Sandewall, O., & Johannesson, M. (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization, 61*, 453–470.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal form games. *Review of Economic Studies, 75*(3), 729–762.

Croson, R. T. A. (2000). Thinking like a game theorist: Factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization, 41*(3), 299–314.

de la Rosa, L. E. (2007). *Overconfidence and moral hazard*. Mimeo: University of Aarhus.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30*(2), 163–182.

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106.

Eichberger, J., Kelsey, D., & Schipper, B. (2009). Ambiguity and social interaction. *Oxford Economic Papers, 61*(2), 355–379.

Engelmann, D., & Strobel, M. (2000). The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics, 3*(3), 241–260.

Englmaier, F. (2004). *A strategic rationale for having overconfident managers*. Mimeo: University of Munich.

Fischbacher, U. (2007). Z-Tree—Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology, 3*(4), 552–564.

Gervais, S., & Goldstein, I. (2007). The positive effects of biased self-perceptions in firms. *Review of Finance, 11*, 453–496.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives*. London: Routledge.

Goel, A. M., & Thakor, A. V. (2008). Overconfidence, CEO selection and corporate governance. *Journal of Finance, 63*, 2737–2784.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63 (pp. 79–93)*. Goettingen: Gesellschaft fuer Wissenschaftliche Datenverarbeitung.

Guerra, G., & Zizzo, D. J. (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior & Organization, 55*(1), 25–30.

Heath, C., Larrick, R., & Klayman, J. (1998). Cognitive repairs: How organizational practices can compensate for individual shortcomings. In B. Staw & L. Cummings, *Research in organizational behavior* (Vol. 20, pp. 1–37). Greenwood, NJ: JAI Press.

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review, 88*(4), 947–969.

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

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

Huck, S., & Weizsäcker, G. (2002). Do players correctly estimate what others do? Evidence of conservatism in beliefs. *Journal of Economic Behavior & Organization, 47*(1), 71–85.

Krueger, J. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 17–42). New York: Psychology Press.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*(2), 311–328.

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

MacDonald, T., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers' predictions differ from those made by observers? *Personality and Social Psychology Bulletin, 25*, 1417–1429.

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

Malmendier, U., & Tate, G. (2008). Who Makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics, 89*(1), 20–43.

Malmendier, U., Tate, G., & Yan, J. (2007). Corporate financial policies with overconfident managers. NBER working paper no. 13570.

Manski, C. F. (2004). Measuring expectations. *Econometrica, 72*(5), 1329–1376.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*(2), 213–225.

Miller, D., & Turnbull, W. (1986). Expectations and interpersonal processes. *Annual Review of Psychology, 37*, 233–256.

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

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica, 70*(3), 971–1005.

Offerman, T., Sonnemans, J., & Schram, A. (1996). Value orientations, expectations and voluntary contributions in public goods. *Economic Journal, 106*(437), 817–845.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences, 23*(1), 125–133.

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

Russo, E., & Schoemaker, P. (1992). Managing overconfidence. *Sloan Management Review, 33*, 7–17.

Rutström, E. E., & Wilcox, N. T. (1996). Stated beliefs versus empirical beliefs: A methodological inquiry and experimental test. Working paper 6-18, University of Central Florida.

Santos-Pinto, L. (2008). Positive self-image and incentives in organizations. *The Economic Journal, 118*, 1315–1332.

Santos-Pinto, L. (2007). Positive self-image in tournaments. *MPRA Paper* 3140.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association, 66*(336), 783–801.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143–148.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193–210.

Urbig, D., Weitzel, U., & Stauf, J. (2009). What is your level of overconfidence? A strictly incentive compatible measurement of absolute and relative overconfidence. Working Papers 09-20, Utrecht School of Economics.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806–820.

Wilcox, N.T., & Feltovich, N. (2000). *Thinking like a game theorist: Comment*. Mimeo: University of Houston.

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