Image Concerns and Behavioral Implications

Three Essays in Experimental Economics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2013

vorgelegt von

Carmen Thoma

Referent:	Prof. Dr. Klaus M. Schmidt
Korreferent:	Prof. Dr. Martin G. Kocher
Promotionsabschlussberatung:	6. November 2013

Datum der mündlichen Prüfung: 31.10.2013

Namen der Berichterstatter: Klaus M. Schmidt, Martin G. Kocher, Fabian Herweg

Acknowledgements

First and foremost, I would like to thank Klaus Schmidt for being an excellent supervisor. He always provided me with guiding advice, inspiring encouragement and constant support, reaching beyond the supervision of my research. I would also like to thank Martin Kocher for agreeing to serve as the second supervisor of my thesis and for sharing his large experimental knowledge with me. I am also thankful to Fabian Herweg for agreeing to serve as the third supervisor and for providing me with helpful and inspiring comments.

I also want to thank my colleagues at the Seminar for Economic Theory and the Munich Graduate School of Economics for generating such an excellent atmosphere and making my doctoral studies so enjoyable. Special thanks goes to my co-authors Sandra Ludwig and Miriam Schütte. Our collaboration was of huge value with benefits going far beyond the quality of our projects. Moreover, I would like to thank colleagues I had the chance to meet at seminars, workshops, and conferences for their excellent comments and their interest in my work.

I am also grateful to Ernesto Reuben for inviting me to Columbia Business School - a stay that I have enjoyed and profited from a lot - with an invaluable souvenir.

I would also like to thank the staff at MELESSA for providing assistance with conducting my experiments. Financial support from the DFG via SFB TR 15, LMUexcellent and the Förderverein Kurt Fordan is gratefully acknowledged.

More than anything I thank Julian and my family for their love and care - the best support I can imagine.

Carmen Thoma

Contents

List of Tables

List of Figures

Preface

Contrary to standard economic theory, non-payoff maximizing human behavior can be observed in many situations in the everyday and working life. For example, people give to charity, hire someone out of sympathy, and pay for the gym although they never end up going there. While humans are surely driven by monetary incentives to a certain extent, they are also influenced by psychological factors, one of which is their social image. For instance, recent economic studies find evidence that social image concerns can partly explain seemingly altruistic behavior. There is evidence that individuals act less selfishly in dictator games (Dana et al., 2007), and are more willing to volunteer (Carpenter and Myers, 2010), or to donate (Ariely et al., 2009) if their behavior is more likely to be observed. The effect of observability of actions on behavior suggests that people care about what others think about them, i.e. that they have concerns for their social image. Thus, people want to behave in line with highly valued characteristics, such as altruism, generosity, or trustworthiness when being observed by others.

In this dissertation we present three experimental studies that analyze further effects of social image concerns on human behavior. The first two chapters investigate how image concerns influence self-assessment, whereas the third chapter explores whether social image concerns are a motivator for promise keeping.

Beliefs about one's abilities influence many decisions, such as career choices and risk taking. Biased beliefs about one's abilities can lead to suboptimal decision making. For example, it has been shown that women seldom choose competitive payment schemes, whereas men tend to choose them very frequently (Niederle and Vesterlund, 2007). Chapter 1 analyzes whether social image concerns drive the less self-confident behavior of women in comparison to men. We investigate whether women have shame to overestimate themselves in public, and thus lower their self-assessment if its accuracy is publicly observable. As a consequence they might appear less self-confident, and even shy away from challenging situations in which they have to prove themselves. This may present one potential explanation for the small share of women in top level jobs and the large gender wage gap. Chapter 2 explores whether women's shame of overestimation might be caused by a societal antipathy towards overconfidence. We analyze whether people favor an individual who is underconfident or overconfident about his relative performance, and whether an underconfident or an overconfident individual is expected to exert more effort in a real effort task. Moreover, we explore whether individuals anticipate the potentially unequal signaling value of under- and overconfidence, and therefore strategically bias their self-assessment if its accuracy is publicly observable.

While standard economic theory predicts that communication does not influence human behavior, experimental studies show that communication indeed enhances cooperation, e.g. in trust games and social dilemmas (Charness and Dufwenberg, 2006). As cooperation usually increases overall welfare, an expanding literature investigates why and under which circumstances communication leads to more cooperation. Chapter 3 contributes to this literature by exploring whether social image concerns make individuals stick to a promise to cooperate, as they do not want to be perceived as a promise breaker. If this holds true, transparency of actions might further increase the positive effect of communication on cooperation in non-contractible settings.

All three chapters use a common method to tackle the respective research questions. As there is no field data available regarding the topics addressed by this dissertation, we generate our own data in an anonymous laboratory setting. We are aware that the laboratory setting, in contrast to settings in the field, might lack reality. However, there are some considerable advantages to collecting data in the laboratory setting. For example, it enables us to analyze gender differences in shame of overestimation, while controlling for other gender differences, e.g. in risk preferences. Moreover, it allows us to limit the exogenous variation to only one factor, that is whether image concerns can occur or not. This enables us to separate the effect of image concerns from other influencing factors. Additionally, in the laboratory it is possible to give subjects only one characteristic, which in the first two chapters, is the accuracy of their self-assessment, and in the third chapter, whether subjects keep their promise or not. Thus, if we observe a treatment difference in behavior due to image concerns, it is caused by subjects' aim to influence exactly this characteristic. This allows us to determine what kind of image the subjects are concerned about, e.g. appearing modest or self-confident or not being perceived as a promise breaker. Moreover, in Chapter 2, if subjects prefer the under- or overconfident agent, it is caused by a preference for under- or overconfidence and not by any other correlated characteristic, such as appearance, which is difficult to control for in the field.

The first chapter - a joint study with Sandra Ludwig - pays attention to the large and problematic gender differences in the labor market: the absence of women in top level jobs and the gender wage gap. Recent economic studies show that women shy away from competitive payment schemes (e.g. Niederle and Vesterlund, 2007), challenging tasks (Niederle and Yestrumskas, 2008), and wage negotiations (Babcock and Laschever, 2003), while men do not. Women's lack of beliefs in their abilities, higher risk-aversion, and lower enthusiasm for competition seem to partly explain this gender difference. In our study we suggest a new effect, shame of overestimation, which may cause gender differences in making career influencing decisions. We define shame as a negative emotion that an individual may have when being not as able or successful as publicly announced. Choosing a competitive or challenging work environment implies that an individual is sufficiently convinced that he will succeed. He might feel ashamed if he does not. To avoid this shame women might not choose these situations at all. Similarly, subsequent to a wage negotiation an employee has to show that he is really worth his money. Thus, a woman's fear of being not as successful as she announced might make her claim to be less successful than she actually thinks she is, and in consequence demand a lower wage than a man. It might even make her shy away from wage negotiations completely.

In a between subjects design we analyze whether women's and men's stated self-assessment changes if we exogenously vary the observability of the accuracy of agent's self-assessment across two treatments. In both treatments the self-assessment is based on subjects' relative performance in a real effort task, and subjects have a monetary incentive to state the rank, which they think they most likely have achieved.

The results show that women rank themselves significantly lower if another participant can infer whether they over- or underestimated themselves than when he cannot, despite the anonymous setting. We do not find such a treatment difference for men. When comparing women's and men's self-assessment, we find that women rank themselves significantly lower than men if shame is possible, but rank themselves similarly if shame is not possible. We conclude that women have shame to overestimate themselves in public, whereas men have not. To avoid this kind of shame, women downgrade their stated belief leading to a gender difference in self-assessment.

The observed gender difference in shame of overestimation might explain why women shy away from situations, in which they have to prove themselves, and men do not. Not entering these situations deters women from reaching top level positions and wage increases. Moreover, women's less self-confident appearance caused by shame might also be obstructive for their career and wage. Firstly, if the true performance is unknown, stated self-assessment might serve as a performance signal, e.g. during job interviews, making the less self-confident woman less likely to be selected than a man. Second, even if the difference in self-assessment is anticipated, self-confidence might be a highly valued characteristic in leading positions giving advantage to equally qualified but more self-confident men. Third, employees usually do not get paid more than they ask for, leaving modest women with a lower wage than men.

Motivated by the findings in the first chapter, the second chapter explores how self-assessment regarding one's own relative performance is perceived by others, and whether individuals strategically adapt their self-assessment to others' preferences. As a biased self-assessment can lead to systematic biases in individuals' decision making, such as overinvestment (Malmendier and Tate, 2005), an expanding number of economic studies addresses the topic of self-assessment. While earlier mainly psychological studies claim that people are overconfident, recent economic studies contradict these findings (Clark and Friesen, 2009; Moore and Cain, 2007). Therefore, an expanding literature tries to identify factors influencing self-assessment, e.g. the difficulty of the task (Hoelzl and Rustichini, 2005). We contribute to this literature by analyzing whether individuals' self-assessment is influenced by image concerns. We explore whether individuals bias their self-assessment to appeal to others, and whether this behavior is justified, i.e. whether overconfident and underconfident people are perceived differently. We thereby focus on two aspects: whether overconfident subjects are considered as more or less likable than underconfident subjects, and whether they are expected to exert more or less effort in a real effort task.

In two experimental treatments participants have to choose one out of two subjects, thereby only observing whether each of the two subjects over- or underestimated himself. In the first treatment participants select the subject that then receives a monetary reward, without any monetary incentives for themselves. In the second treatment participants have a monetary incentive to select the subject that they expect to perform better in a subsequent task. The only information participants receive about the two subjects is the accuracy of their self-assessment. Subjects' self-assessment is based on their relative performance in a real effort task.

The results show that underconfidence beats overconfidence in both respects. Underconfident subjects are rewarded significantly more often than overconfident subjects, and are significantly more often expected to perform better in the real-effort task. It seems as if subjects being less convinced of their performance are considered as more sympathetic, and are expected to be more ambitious to increase their performance, whereas overconfident subjects are rather expected to rely on their high self-confidence instead of trying to improve. Elicited beliefs reveal that the stronger performance signal of underconfidence is not anticipated, whereas its higher sympathy value is anticipated. Moreover, the comparison to a control treatment, in which the incentives to be selected are eliminated, reveals that men strategically downgrade their self-assessment in order to be rewarded by others. Yet, we do not find a difference in self-assessment for women. Women either do not deflate their self-assessment in order to be rewarded by other, or they even do so in non-strategic settings, in which the accuracy of their self-assessment is observed by others. This behavior might be driven by non-monetary image concerns of women and is in line with the findings in Chapter 1.

The results of the second chapter contribute to the strand of literature exploring advantages and disadvantages of overconfidence in comparison to underconfidence. At the same time the findings add to the questions why and in which situations individuals might (rationally) exhibit a bias in their self-assessment. The results show that being underconfident can be advantageous regarding the perception of others. Thus, subjects might state a lower self-assessment than they actually have if others can observe its accuracy. Combining the results of the first and the second chapters, we suggest that women might have internalized the antipathy towards overconfidence, thus downgrading their actual self-assessment even in non-strategic settings.

It is well known that breaking a promise is deemed negative in society. In the third chapter - a joint study with Miriam Schütte - we analyze whether subjects stick to their promise in order to avoid the image of being a promise breaker. While rationality predicts that communication does not influence individuals' behavior, experimental studies have shown that communication can be an effective tool to enhance cooperation. As cooperation among interacting parties usually increases economic welfare, we analyze whether the cooperation enhancing effect of communication can be further increased in settings with social image concerns.

In a between subject design participants play a one-shot sequential trust game in which the second mover can promise the first mover that he will cooperate. We exogenously vary whether the second-mover's action is revealed to the first mover, and analyze whether revelation leads to more cooperation of the second-mover. If second movers want to avoid being perceived as a promise breaker, they should more likely cooperate in case their action is revealed than when it is not. Due to our design we are able to attribute a difference in cooperation to social image concerns and to exclude other explanations provided by the literature so far: promise keeping per se (Vanberg, 2008) and guilt aversion (Charness and Dufwenberg, 2006). The preference for promise keeping per se does not depend on the observability of the second mover's action. First order beliefs and consequently second order beliefs cannot be influenced by our treatment variation either since first-movers do not know whether the second-mover's action is revealed to them or not. As subjects might also care about being perceived as selfish (Tadelis, 2011), revelation might increase cooperation independent of communication. Therefore, we compare the results to a control treatment without communication.

However, our results do not support the relevance of social image concerns on promise keeping. In the treatment with communication we do not find a significant effect of revealing the second-mover's action on his likelihood to cooperate. We do not observe an effect of revelation in the treatment without communication either. Thus, in our setting there is neither a social image concern of being perceived as a promise breaker nor concern of being perceived as being selfish.

Without revelation the large share of 81% of promises is kept, limiting the scope for a further increase with revelation and highlighting the high preference for promise keeping per se due to self image concerns (Vanberg 2008). However, the between subject variation of the pre-defined messages available to the second mover reveals that individuals indeed have a preference for keeping their promise per se, but tend to break statements of intent ("I will cooperate" in comparison to "I promise to cooperate"). Nevertheless, first movers seem to trust both messages similarly, which seems to be anticipated by second movers. Thus, second movers' feelings of guilt should not vary between breaking a promise and breaking a statement of intent. Therefore, if guilt aversion was the reason for the high promise keeping rate, second-movers should also stick to their statements of intent. As they do not, we exclude guilt aversion (as found by Charness and Dufwenberg, 2006) as an explanation for the high promise keeping rate in our experiment. The variation in pre-defined messages also shows that the set of available messages highly influences the positive effect of communication on cooperation. Cooperation of second movers highly increases in comparison to the treatment without communication if they can either send a promise or an empty message, which does not announce their action. In contrast, cooperation with communication is not higher than without if second movers can also choose to send a statement of intent, as this seems easier to break than a promise.

The following three chapters are all self-contained and have their own introduction and appendix. Each chapter can thus be read independently of the other two.

Chapter 1

Do Women Have More Shame than Men? An Experiment on Self-Assessment and the Shame of Overestimating Oneself*

1.1 Introduction

Frequent and much discussed observations in labor markets are the absence of women in top level jobs and the gender wage gap.[1] Recent studies suggest that this may be due to the fact that women in comparison to men shy away from competition, demanding work environments, and negotiations about their wage.[2] This behavior seems to be partly driven by women's lower self-assessment of their ability, higher risk-aversion and lower

*This chapter is based on joint work with Sandra Ludwig from the University of Ulm.

[1]See e.g. Bertrand and Hallock (2001).

[2]For example, Balafoutas and Sutter (2012), Charness et al. (2012), Datta Gupta et al. (2013), Dohmen and Falk (2011), and Niederle and Vesterlund (2007) show that women are less competitive than men; Niederle and Yestrumskas (2008) show that women choose challenging tasks less often than men; Babcock and Laschever (2003), Bowles et al. (2005), and Gerhart and Rynes (1991) show that women negotiate their wage less than men.

competitiveness.[3]

In this paper we analyze another mechanism, the effect of shame, that may imply gender differences in occupational decisions. Shame may also shed light on why women exhibit a lower self-assessment. We define shame in our context as the negative moral emotion that an individual may have when she is not as able or successful as she *publicly* announced.[4] For example, an agent might have shame if subsequent to a wage (or promotion) negotiation or a job interview the employer observes that the agent is not as able as she claimed to be. Similarly, choosing a competitive or demanding work environment can be seen as a public statement of being sufficiently confident to succeed. An agent might feel ashamed if someone else (the employer or competitor) observes her suffer defeat. Agents may want to avoid shame and thus make less confident statements about their abilities, or even shy away from situations, in which they might end up feeling ashamed. Specifically, we investigate whether women make less confident statements about their ability when their true ability is observable than when it is not – because they want to avoid the shame of overestimating their ability – and whether men's statements are less sensitive to the observability of their ability.

We conduct a controlled laboratory experiment, in which the subjects first perform an incentivized task. Afterwards, they are randomly assigned to be principals and agents. One principal is matched with two agents. Both agents estimate the relative rank of their performance in the task compared to other participants and receive a payment if their guessed rank is correct. According to their monetary incentives, agents should state the rank they think is most likely correct.

To isolate the effect of shame, we vary the (potential) exposure to shame across two

[3]There exist other explanations for gender differences in the labor market such as discrimination against women and preference differences regarding e.g. child rearing (see e.g. Altonji and Blank, 1999; Goldin and Rouse, 2000).

[4]In questionnaire studies, psychologists analyze which emotions individuals classify as shame. A consistent definition of shame, however, does not exist, and the distinction between shame and related emotions such as guilt and embarrassment is difficult. A long-standing notion is that shame is related to situations with public exposure and disapproval of one's failing, while guilt does not depend on public exposure (for a discussion, see Tangney, 2002). The distinction between shame and embarrassment is even less clear and strongly debated. For an overview see Sabini et al. (2001).

treatments. In both treatments, the principal observes the agents' self-assessments. The only difference between the two treatments is that in one treatment (*Info*) the principal additionally observes the agents' true ranks, i.e. the principal can infer whether the agents over- (or under-) estimated themselves. In the other treatment (*NoInfo*), the principal does not observe the agents' true ranks and thus cannot infer the accuracy of the agents' self-assessments. Note that if agents state a lower rank in *Info* than in *NoInfo* (given equal performance in both treatments), then the only obvious explanation is that agents try to avoid shame. Social preferences, overconfidence per se, risk-aversion, or preferences for competition cannot explain a treatment difference in guessed ranks, as we only vary the observability of the accuracy of the agents' self-assessment.

In our experiment, we find neither a gender difference in performance nor a performance difference between treatments. Yet, we observe that women in *Info* rank themselves significantly lower than women in *NoInfo*. For men, we observe no significant treatment effect; if anything, the effect is in the opposite direction. Thus, shame-aversion might explain the different behavior of women and men in settings in which others observe or learn over time the accuracy of their self-assessment. In addition, shame-aversion may strengthen the frequently observed gender difference in self-assessment.[5] While we also find that women rank themselves significantly lower than men in treatment *Info*, the gender difference in guessed ranks disappears in *NoInfo*. In addition, it seems that to some extent women downgrade their guessed ranks consciously: In *Info* significantly more women (compared to *NoInfo*) think that if their guessed rank is not correct, then their actual rank will rather be superior than they guessed previously.

What causes the shame to overestimate (and not to underestimate) oneself? In a post-experimental questionnaire almost 90% of the subjects indicate that overestimating oneself is deemed negative in society. In contrast, only about 20% state that underestimating

[5] For gender differences in self-assessment see e.g. Balafoutas et al. (2012), Beyer (1990), Beyer and Bowden (1997), Möbius et al. (2012), Niederle et al. (2013), Reuben et al. (2012). It is, however, difficult to compare the size of the gender difference in self-assessment between studies, and to explore whether shame or the absence of shame drives differences between studies, since experimental conditions vary, in particular, how the self-assessment is elicited.

oneself is deemed negative in society.[6] Yet, we observe no gender difference in these statements. Yet, we find that subjects expect men, but not women, to overestimate their performance in the real task. Given these expectations, women in comparison to men may (believe they) worsen their social standing to a greater extent when overestimating themselves and others observe it. This may imply that only women downgrade their self-assessment in *Info* due to a stronger (anticipated) social disapproval of their overconfidence.[7]

Regardless of the root cause for the shame to overestimate oneself, we find that women react stronger to it than men. The stronger reaction could be explained by the general psychological finding that women experience self-conscious emotions (SCE) such as shame, embarrassment, and guilt more than men.[8] Women's shame-aversion may lead to more cautious behavior and self-promotion when their performance becomes, at least with some probability, verifiable afterwards. For example, in job interviews, or when employees have to assess the extent to which they reached their targets in the context of management by objectives, women may state lower beliefs about their abilities/achievements. Similarly, in wage negotiations women may ask for lower wages than men, or may not ask for a wage increase at all, because their claim reflects their belief about their ability. Moreover, women may not enter competitive environments, because entering expresses a high confidence in their ability, but the outcome may disprove that they are of high ability.

Besides the literature on gender differences in the labor market and in self-assessment cited before, the recent theoretical and experimental study by Ewers and Zimmermann (2012) is related to our study. Their focus is on signaling effects in self-assessment – motivated by image concerns. Confirming their theoretical predictions, they provide evidence that subjects try to signal high ability in front of an audience. They also analyze whether subjects signal modesty, a characteristic similar to shame, and only find weak evidence.

[6] The second chapter provides evidence that people prefer underconfident over overconfident subjects.

[7] Evidence from the psychological literature suggests that the society evaluates the same kind of behavior differently for men and women. Bowles et al. (2007) find that women are penalized when trying to negotiate a higher wage, while men are not. Eagly (1987) and Rudman (1998) show that self-promoting women are evaluated worse than modest women, while there is no such difference for men.

[8] Else-Quest et al. (2012) provide an overview and a meta-analysis of studies on SCE.

However, they do not analyze gender differences in modesty signaling.

The rest of this chapter is structured as follows. In Section 1.2 we describe our experimental design. In Section 1.3 we report our results. In Section 1.4 we provide a closer look at the effect of shame, analyzing whether individuals are aware of the reactions to shame and of the gender difference in self-assessment. We also discuss why only women seem to be shame-averse. We conclude in Section 1.5.

1.2 Experimental Design and Hypotheses

In our experiment subjects are randomly assigned to the roles of principals and agents.[9] One third of the participants are principals, the remaining two thirds are agents. Participants are randomly matched into groups of three, where each group consists of one principal and two agents. The setting is completely anonymous. Participants do not learn the identity of the other subjects in their group, neither during nor after the experiment. Before learning about the two different roles and the allocation to groups, all participants perform a real effort task. Afterwards the agents assess their relative performance in the task (explained below). In order not to distort effort, subjects do not know about the self-assessment when completing the task. Participants only know that another part of the experiment will follow. Participants receive separate instructions for the task and for the self-assessment part and complete each part only once. The instructions are handed out to participants and read aloud at the beginning of the experiment and after the task respectively.[10]

In the real effort task subjects add up sets of five two-digit numbers. Subjects are not allowed to use a calculator, but to use the provided scratch paper. After a subject has entered and confirmed her result for a set of numbers, a new set of numbers appears on the screen. Once a result has been confirmed, subjects cannot go back and revise their

[9]The experiment was framed neutrally. While we refer to “principals” and “agents” in the following, we used the neutral terms participant A and B to describe the roles in the experiment.

[10]Translated instructions are provided in the appendix.

result. On the screen the task looks as follows:

54	27	63	10	89		*Result:*

Each set of numbers is randomly generated. Subjects perform this task for 7.5 minutes. They may solve as many problems as they can. On the screen they see the remaining time as well as their number of correctly and wrongly solved problems by then. Before the 7.5 minutes start, there is a practice phase of two minutes during which subjects can get acquainted with the software interface, while no money is earned. We chose the number-adding task as performance depends not only on effort, but also on ability.[11] Furthermore, the task is easy to understand and the performance is easy to measure. In addition, several other studies use this task and predominantly do not observe a gender difference in performance.[12]

Each agent receives two tokens for every arithmetic problem she solves correctly.[13] Principals receive no payment for the task. When adding the sets of numbers, however, subjects do neither know their role, nor whether they will be paid for the task or not. They only know that two thirds of all subjects, which are randomly determined at the beginning of the experiment, receive two tokens per correct answer, while the remaining third receives no payment for the task. As soon as the task is finished, subjects learn their role.

After having completed the real effort task, each subject is assigned (but not told) a rank between 1 and 22. To determine a subject's rank we compare her performance to the performance of 21 participants of another session ("baseline treatment"). Rank 1 refers to the best performance in this group of 22 subjects and rank 22 to the worst performance. Each subject is ranked to the same 21 participants of the baseline treatment.

[11] Thus, subjects may base their self-assessment not only on their performance in this task, but also on their performance in school, study, etc. This is confirmed by our results as the estimated relative performance is correlated with the grade in the final secondary school examinations (abitur), in the following called final school-grade (see Section 1.3).

[12] See e.g. Balafoutas et al. (2012), Eriksson et al. (2009), and Niederle and Vesterlund (2007). An exception is Niederle et al. (2013). Note however, that a gender difference in performance would not be a problem in our experiment, as we mainly analyze treatment effects for each gender separately (as we explain in more detail below), rather than gender differences within one treatment.

[13] During the experiment subjects earned tokens. At the end of the experiment, tokens were converted into Euros where 1 token=25 Eurocent.

The baseline treatment was the first session that we conducted, and we use it to determine the performance ranking only. The participants completed the identical task and had exactly the same instructions and incentives for this task, i.e. subjects knew that they would be paid by piece-rate with a probability of two thirds. As a subject is compared to the 21 participants of the baseline treatment, her rank is independent of the performance of the other subjects in her session. More precisely the ranking is determined as follows. A subject is assigned rank $r \in \{1, ..., 22\}$ if she performed better or as good as $22 - r$ participants of the baseline treatment. A subject performed better than a participant of the baseline treatment if she solved more problems correctly. In case she solved the same number of problems correctly, the subject is better if she made less mistakes. The subject performed as good as a participant of the baseline treatment if both solved the same number of problems correctly and made the same number of mistakes.

Each agent is asked to estimate her rank between 1 and 22. In the following, we refer to an agent's estimate as her "guessed rank".[14] If an agent's guessed rank is correct, i.e. equals her actual rank, she receives 50 tokens (12.50 Euros). She receives no payment if her guessed rank differs from her actual rank.[15]

In each group the agent whose guessed rank is better is automatically "chosen" – irrespective of the accuracy of her guessed rank.[16] Note that a better guessed rank means that the stated number is smaller, and not that the guessed rank is closer to the actual rank. The actual performance of the chosen agent affects the expected payment of the

[14]To be precise, the principals also guess their rank, yet, their guessed rank has no payoff consequences. Therefore, their guesses are not comparable to the agents' guesses and we do not analyze them. We let them guess their rank to keep them busy and to avoid that participants can infer who is a principal due to their inactivity.

[15]This "all-or-nothing" payment rule has the advantage that it is easy to understand and it ensures that each subject has the incentive to state the rank which she thinks is most likely her actual rank (i.e. to state the mode of the ranks on which she places a positive probability) – irrespective of her risk-preferences. A quadratic-scoring rule in contrast is much more difficult to understand. In addition, the use of a quadratic-scoring rule is problematic if subjects are not risk-neutral (see e.g. Holt, 1986; and Savage, 1971). Since women tend to be more risk-averse than men (see Eckel and Grossman (2008) or Croson and Gneezy (2009) for an overview), a quadratic-scoring rule may induce gender differences in guessed ranks, which we want to avoid.

[16]If both agents in a group have the same guessed rank, one of them is randomly chosen. The selection occurs automatically to avoid confounding effects.

principal.[17] The principal receives a payment of $(22 - rank) \cdot 3$ tokens, whereupon the rank is either the actual rank of the chosen agent or a random rank between 1 and 22, both with equal probability. Thus, the better the actual rank (i.e. the smaller the number) of the chosen agent, the higher the expected payoff of the principal. The principal does not learn whether the actual rank of the chosen agent or a random rank determines her payoff. We introduced the random rank to avoid that the principal can deduce the agent's actual rank from her payoff. This is crucial as in one treatment (see *NoInfo* below) the principal is not supposed to learn the agents' actual ranks. An agent receives no payment for being chosen. We only set incentives to correctly guess the rank. We deliberately abstract from monetary incentives for being chosen to isolate the effect of shame. Such monetary incentives might induce agents to overstate their ability for strategic reasons, i.e. to lie. There might not only be gender differences in the willingness to lie in general, but also depending on whether lying can be detected.[18]

At the end of the experiment each agent learns her actual rank and whether she is chosen or not. The principal learns the guessed ranks of both agents in her group and – depending on the treatment – she additionally learns the agents' actual ranks. This means we vary the principal's information about the agents' actual ranks across treatments. In the first treatment ("*NoInfo*") the principal only learns the guessed ranks of both agents, but not their actual ranks. In the second treatment ("*Info*") the principal learns the guessed and the actual ranks of her agents. *Before* guessing their ranks, the agents are informed that the principal learns their guessed ranks and whether or not she additionally learns their actual ranks. Note that the setting is completely anonymous, i.e. the principal is shown the agents' guessed and/or actual ranks on the screen, but the agents' identities are not revealed.

[17]We form groups of two agents and one principal to generate more observations of agents' self-assessments compared to a matching of one agent and one principal. The principal's earnings depend on the agents' self-assessments and the chosen agent's performance (i) to intensify the principal-agent relationship in the anonymous laboratory situation and (ii) to make the setting more reasonable and realistic (as in an application or wage negotiation setting).

[18]For example, Charness et al. (2012) find that men tend to increase their self-assessment for strategic reasons, while women do not. The studies by Houser et al. (2012) and Lundquist et al. (2009) indicate that women are less likely to lie.

How can a treatment difference in the agents' guessed ranks be explained? The only difference between treatments is whether or not the principal learns the agents' actual ranks; the agents' monetary incentives for their guessed ranks are identical in both treatments. Therefore, risk preferences, social preferences, preferences for competition or overconfidence per se cannot explain a treatment difference in guessed ranks. What can explain the treatment difference is shame. An agent may have shame if the principal can infer the accuracy of the agent's guessed rank (as in *Info*). In particular, an agent may feel ashamed if she stated to be better than she turns out to be, i.e. if she overestimated herself. To avoid this shame, the agent may guess a worse rank when the principal learns the accuracy of her guessed rank. In our setting an agent's shame might be intensified by the fact that the guessed ranks also determine which agent is chosen and thereby whose actual performance affects the principal's expected payoff.[19]

Apart from the shame subjects may have if the principal observes that they overestimated themselves, subjects may feel ashamed just because they themselves learn that they overestimated themselves or because the experimenter learns it. Both kinds of shame towards oneself and the experimenter, however, cannot explain a treatment difference: Agents learn their actual ranks and thus the accuracy of their guess in both treatments, and also the experimenter always observes the accuracy of guessed ranks. When we talk about shame in the following, we refer to the shame agents feel towards the principal.

Evidence from psychology suggests that women experience emotions more intensely than men – according to self-reports (see e.g. Brody, 1997; Grossman and Wood, 1993; and Else-Quest et al. (2012) for a meta-analysis). Women's stronger emotional experience can increase their disutility from overestimating themselves. Due to the gender difference in emotional experience and to anecdotal evidence from personnel managers that women shy away from promoting their abilities, we expect that particularly women have shame and try to avoid it by guessing a worse rank in *Info*. Thus, we propose the first hypothesis:

Hypothesis 1. *The guessed ranks of women in Info are worse than in NoInfo.*

[19]Note however, that overestimating as well as underestimating one's performance can negatively affect the principal's payoff since not the best performer might be chosen.

Since we expect men to be less prone to shame, we state the second hypothesis:

Hypothesis 2. *Men's guessed ranks in Info and NoInfo do not differ.*

Moreover, we expect that shame largely explains the gender gap in self-assessment – given that men and women have the same ability – and state as third hypothesis:

Hypothesis 3. *Women assess themselves lower than men in Info. The gender gap in self-assessment in NoInfo is strictly smaller than in Info.*

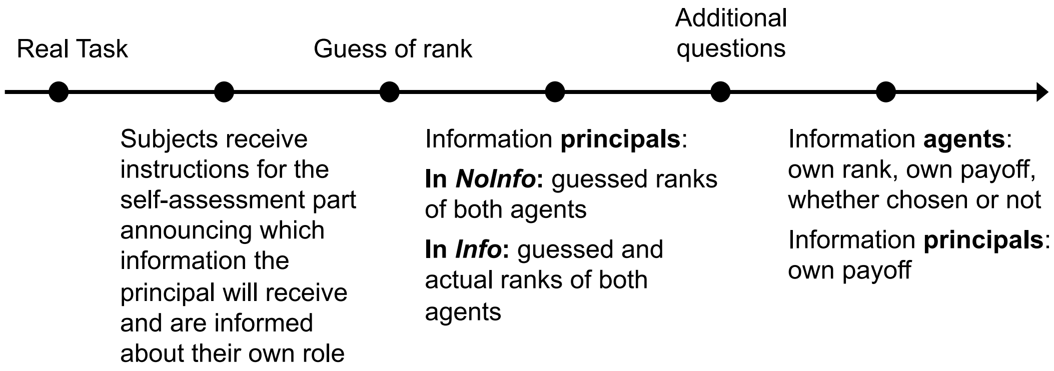
In both treatments we elicit additional information from all subjects before informing them about the accuracy of their guessed rank. First, we ask each subject whether she thinks that her actual rank would rather be better or worse than she previously estimated in case her guessed rank would turn out to be wrong. Subjects, whose guessed rank indeed turns out to be wrong, receive two tokens if their answer is correct. This question may shed light on whether subjects are aware of their potentially different behavior in the two treatments. Second, we elicit the subjects' beliefs about the average actual rank and the average guessed rank of all female as well as of all male agents in their session. These beliefs provide an insight into whether people expect a general tendency to over- or underestimation and gender differences in self-assessment. One of these four estimates is randomly chosen for payment and subjects receive 16 tokens if their corresponding answer is correct.[20]

To summarize the course of the experiment Figure 1.1 illustrates the order in which subjects take their actions and receive information. Finally, the subjects complete a questionnaire, which asks for their gender, age, subject of study, final school-grade, and elicits their risk preferences and degree of self-esteem. In addition, we ask subjects whether they think that over- (under-) estimating oneself is deemed negative in society.

We conducted the computerized experiment in the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) at the University of Munich during April

[20] To avoid any kind of hedging strategies, all five additional, incentivized questions were not announced in the instructions.

Figure 1.1: The Course of the Experiment



2011. Participants were recruited via ORSEE (Greiner, 2004). Including the baseline treatment, 171 subjects participated in our experiment (mainly students from the universities in Munich). The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007). Participants were assigned individual computer terminals and could not see other participants' decisions. We ran four sessions per treatment. Subjects were randomly assigned to sessions and could take part in one session only. The gender composition of each session was roughly half women and half men.[21] Each session lasted about one hour and subjects earned 14.93 Euros on average (including a show-up fee of 4 Euros).

1.3 Main Experimental Results

First of all we check whether there are differences in the subjects' performance between treatments as such differences could influence our later treatment comparisons in guessed ranks. Neither women's nor men's performance differs between treatments: Female agents solve on average 16.2 problems correctly in *NoInfo* and 15.9 in *Info*, whereas male agents solve on average 17.2 problems correctly in *NoInfo* and 16.8 in *Info* (two-sided Mann-Whitney-U-tests (MWU) yield $p = 0.779$ and $p = 0.799$ for women and men respectively).

[21] We invited the same number of men and women to each session. Due to different show-ups, however, not exactly 50% men and women participated.

The number of wrongly solved problems does not differ between treatments either ($p = 0.230$ for women and $p = 0.703$ for men, MWU, two-sided).

Since the subjects' self-assessment is measured in relative ranks, we also analyze whether the ranks differ between treatments. Here, we consider a subject's optimal guessed rank, i.e. the guessed rank that conditional on a subject's performance maximizes her expected earnings. Thus, we determine the rank that is most likely assigned to her given her performance and the observed performance distribution of all participants in all sessions.[22] Comparing the optimal guessed ranks between treatments, we find no treatment differences, neither for women nor for men: The average optimal guessed rank of women in *NoInfo* is 12.1 and in *Info* it is 12.5 ($p = 0.833$, MWU, two-sided). For men the average optimal guessed rank in *NoInfo* is 11.0 and in *Info* it is 11.5. ($p = 0.800$, MWU, two-sided). Thus, to maximize earnings, neither women's nor men's guessed ranks should differ between treatments.

Table 1.1 summarizes for each gender and treatment the number of observations, the average number of correctly and wrongly solved problems, the average optimal rank, the average guessed rank, and the corresponding standard deviations.

Table 1.1: Summary Statistics

		# observations	# correct	# wrong	optimal guessed rank	actual guessed rank
Women	*NoInfo*	25	16.2 (5.3)	2.9 (1.9)	12.1 (6.2)	7.8 (3.6)
	Info	24	15.9 (5.0)	2.3 (1.8)	12.5 (5.8)	10.3 (4.6)
Men	*NoInfo*	25	17.2 (5.2)	3.0 (2.2)	11.0 (5.8)	6.8 (3.3)
	Info	26	16.8 (6.8)	2.5 (1.2)	11.5 (6.7)	6.4 (4.5)

The sample consists of all agents. Standard deviations are given in parentheses.

Although women's performance is the same in both treatments, women rank themselves significantly worse in *Info* by as much as 2.45 ranks on average ($p = 0.035$, MWU, two-

[22] In order to calculate the optimal guessed ranks, we ran Monte-Carlo simulations, in which we randomly drew 500,000 groups consisting of 21 participants out of the performance distribution of all participants (with replacement). We then calculated for any given performance level the rank within each simulated group. The optimal guessed rank equals the mode of all 500,000 simulated ranks.

sided): In *NoInfo*, women's average guessed rank is 7.80, while it is 10.25 in *Info*. This observation confirms our first hypothesis.

Result 1. *Women state significantly lower beliefs about their performance if the accuracy of their self-assessment is observable.*

The guessed ranks of men, in contrast, do not differ significantly between treatments ($p = 0.394$, MWU, two-sided), which confirms our second hypothesis. If anything, the effect even seems to go in the opposite direction: In *NoInfo*, their average guessed rank is 6.80, while it is 6.38 in *Info*.

Result 2. *There is no significant difference between men's stated beliefs about their performance when the accuracy of their self-assessment is observable or not.*

Women's lower self-assessment in *Info* in comparison to *NoInfo* also becomes apparent in Figure 1.2, which illustrates the distribution of women's guessed ranks. Figure 1.3 shows the distribution of men's guessed rank, illustrating that men rather tend to rank themselves higher instead of lower in *Info* than in *NoInfo*.

Figure 1.2: Distribution of Guessed Ranks for Women

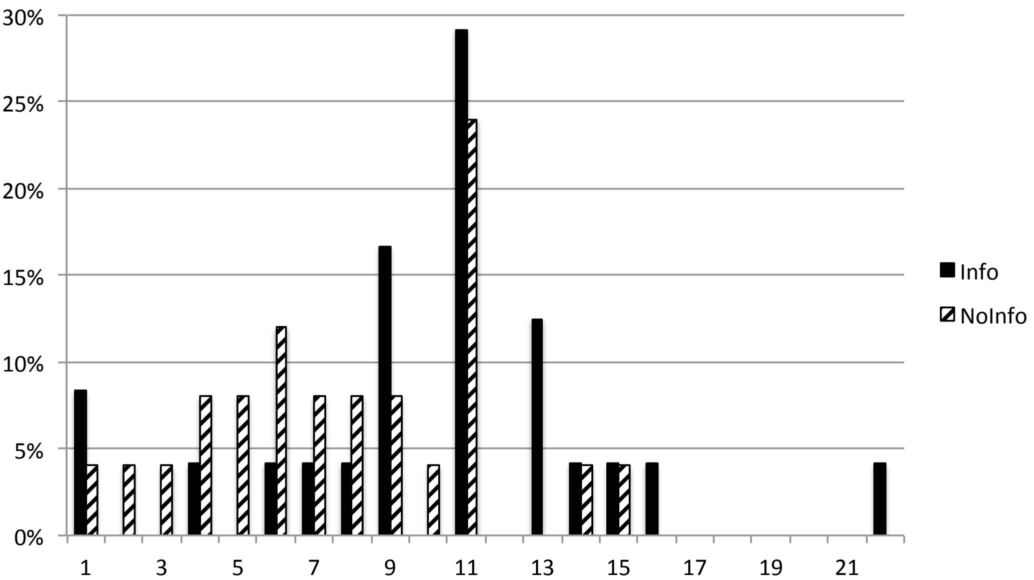
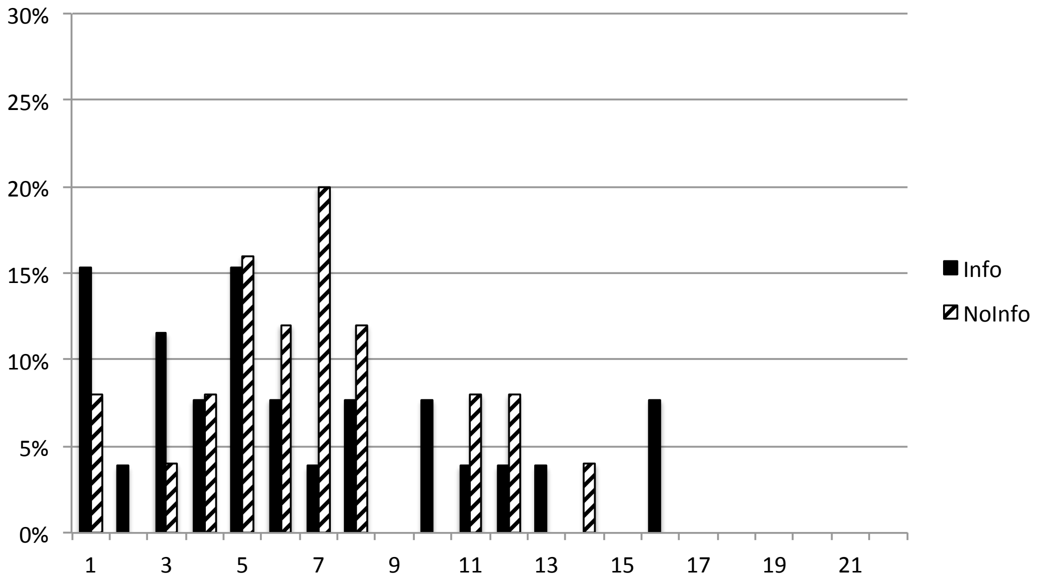


Figure 1.3: Distribution of Guessed Ranks for Men



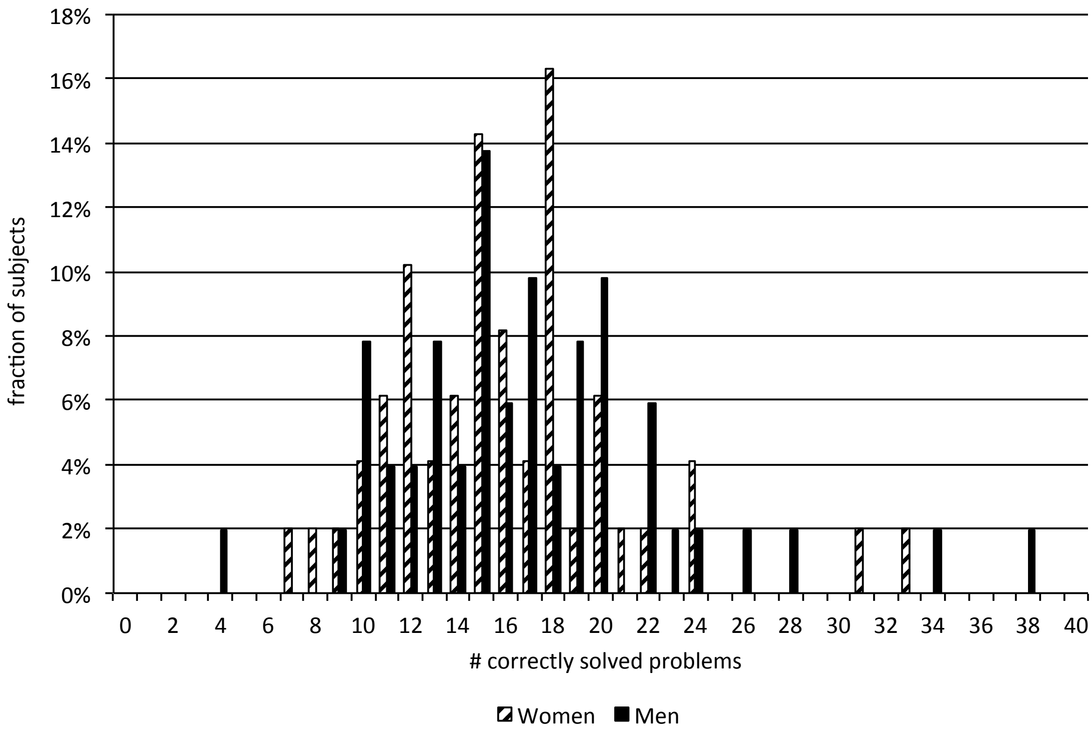
From the first two results we conclude that women feel ashamed if they state a better rank than they actually have, in case their actual rank is observable. To avoid this shame they downgrade their self-assessment. Men do not have this kind of shame, or are at least less shame-averse than women, such that their (incentivized) self-assessment does not change.

Since only women, but not men, downgrade their beliefs in *Info*, shame-aversion may partly explain the frequently observed gender gap in self-assessment. Therefore, we next compare the gender gap in self-assessment across treatments. In *Info*, women rank themselves significantly and substantially worse than men by almost 4 ranks on average ($p = 0.004$, MWU, two-sided): The average guessed rank of women is 10.25, while it is 6.38 for men. This gender gap in guessed ranks is not driven by a different performance of women and men as their number of correctly solved problems (see Table 1.1) does not differ significantly (two-sided MWU-tests yield $p = 0.613/0.520/0.520$ for *Info*/*NoInfo*/both treatments pooled).[23] Accordingly, optimal guessed ranks do not differ across gender either ($p = 0.606/0.534/0.401$ for *Info*/*NoInfo*/both treatments pooled, MWU, two-sided).

[23]Men solve on average one problem more than women. Yet, this result is rather driven by some high-performing men (i.e. outliers) and not representative for the whole group (see Figure 1.4).

Figure 1.4 illustrates the distribution of the correctly solved questions for women and men for the pooled sample of both treatments.[24]

Figure 1.4: Distribution of Correctly Solved Problems



Thus, in treatment *Info* we confirm evidence found by earlier studies on women stating lower beliefs about their performance than men (see footnote 6). Yet, we cannot confirm the finding that women state lower beliefs in treatment *NoInfo* (see Table 1.1). In *NoInfo* the gender difference in guessed ranks is not significant ($p = 0.349$, MWU, two-sided). Hence, our observations confirm Hypothesis 3 and suggest that women are not less self-confident than men per se, but are rather more modest and reserved when others learn whether they over- or underestimated themselves.

Result 3. *Women state significantly lower beliefs about their performance than men if the accuracy of their self-assessment is observable, otherwise, the gender difference in self-assessment disappears.*

[24]The figure looks very similar for each treatment.

The previous results are confirmed by ordered probit regressions. The results of the regressions are reported in Table 1.2. We regress the guessed rank on performance (number of correctly solved problems), a female dummy, a dummy for treatment *Info*, and risk attitude.[25] Thus, the reference category is man in *NoInfo*.

Table 1.2: Ordered Probit of Guessed Rank

	Coefficient (Robust Std. Error)			
	(1)	(1a)	(2)	(2a)
Performance	-0.113***	-0.126***	-0.115***	-0.125***
	(0.020)	(0.020)	(0.019)	(0.022)
Female	0.457***	0.414**	0.035	-0.112
	(0.178)	(0.171)	(0.099)	(0.172)
Info	0.168	0.097	-0.265	-0.414
	(0.183)	(0.200)	(0.162)	(0.154)
Female**Info*			0.890***	1.071***
			(0.200)	(0.297)
Risk attitude	-0.123**	-0.133*	-0.127**	-0.133*
	(0.055)	(0.073)	(0.051)	(0.070)
Additional controls	No	Yes	No	Yes
# of observations	100	100	100	100
# of sessions	8	8	8	8
Pseudo R-squared	0.085	0.125	0.095	0.137
Log Pseudo Likelihood	-215.6	-206.2	-213.4	-203.3

The sample consists of all agents and the regression clusters on sessions. The additional controls are final school-grade, subject of study, age, number of siblings, degree of self-esteem, dummy for mother working or not.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

To elicit risk preferences, individuals indicated on a scale ranging from 0 to 10 whether they are willing to take risks (or try to avoid risks). 0 represented a very weak willingness to take risks, while 10 represented a strong willingness to take risks. Dohmen et al. (2011) show that this general risk question is a good predictor of actual risk-taking behavior. In specifications (1a) and (2a) we include further controls: age, final school-grade, a dummy for a quantitative orientation in the subject of study (economics, mathematics,

[25] Additionally, we ran OLS regressions (see Table A1.1 in the appendix). The results remain qualitatively unchanged.

natural sciences), self-esteem, number of siblings, and a dummy that is one if an agent's mother is not working. The additional information stems from self-reported questionnaire responses at the end of the experiment. Self-esteem is measured by Rosenberg's (1965) self-esteem scale, where a higher score indicates a higher self-esteem. In specifications (2) we additionally include an interaction term for female and *Info*. In all specifications we consider the guessed ranks between 12 and 22 as one category, as only 16 of all 100 participants rank themselves worse than rank 11.

The number of correctly solved problems (performance) is a strong predictor for the guessed rank. The result is robust over both specifications.[26] In specifications (1) the coefficient of the female dummy is positive and significant, meaning that a woman states a worse rank than a man. The coefficient of the treatment dummy, however, is not significant. Thus, when looking at the sample of all agents, men and women, guessed ranks do not differ between treatments. Yet, in specifications (2) the interaction effect is positive and significant. Moreover, the coefficient of the female dummy becomes insignificant.[27] This confirms Results 1 - 3.

The Accuracy of Agents' Self-Assessment

Given that our focus is on people's (women's) shame to overestimate themselves, it is interesting to see whether there is indeed a tendency to overestimation in the number-adding task. Empirical evidence often suggests that people tend to overestimate their abilities, yet, overconfidence seems to depend on the task, the incentives, and the techniques to elicit self-confidence.[28] To address the issue of overestimation, we next consider the average "accuracy" of the agents' guessed ranks. A meaningful measure of accuracy should take account of the fact that the group of 21 participants in the baseline treatment

[26]Note that the performance in the number adding task is neither correlated with the final school-grade nor with the quantitative subject of study dummy (Spearman rank order correlation, Spearman's $\rho = 0.021/0.127$, $p = 0.834/0.207$, $N = 100$). Yet, the coefficient of the final school-grade is positive and significant, meaning that subjects who performed better at school guess a better rank.

[27]Note that men are significantly less risk-averse than women ($p = 0.066$, MWU, two-sided). If we do not control for risk preferences in the regressions, qualitative results remain unchanged with the exception that in (2) and (2a) the coefficient of *Info* becomes significant.

[28]See e.g. Benoit and Dubra (2011), Burks et al. (2013), Hoelzl and Rustichini (2005), Klayman et al. (1999), and Pulford and Colman (1997).

– which determines a subject's 'actual rank' in the experiment – is relatively small and might contain outliers. Therefore, we calculate the accuracy of an agent's guessed rank as the difference between her optimal guessed rank and her actual guessed rank. If an agent's guessed rank coincides with her optimal guessed rank, her accuracy is 0, whereas a positive (negative) accuracy means that she over- (under-) estimated herself. According to our measure of accuracy, the majority of subjects overestimates the relative performance. This holds true for women and men in each treatment (63-80 % overestimate themselves). Women and men similarly overestimate themselves in *NoInfo*. The average accuracy for women is 4.32 and for men it is 4.16 ($p = 0.922$, MWU, two-sided). Yet, women tend to overestimate themselves less than men in *Info* (the average accuracy of women is 2.25, while it is 5.15 for men). The difference is marginally significant ($p = 0.077$, MWU, two-sided).[29]

Given that many women overestimate their relative performance, but guess lower ranks in *Info*, one might wonder whether women's guessed ranks become more accurate in *Info* due to shame-aversion. This would presume that rather those women who show a tendency to overestimate themselves downgrade their beliefs due to shame-aversion (e.g. because they are to some extent aware of their tendency to overestimate themselves). Indeed, when we consider the *absolute accuracy* of women's beliefs, on average, women's guessed ranks are closer to the optimal guessed rank in *Info*. Yet, the difference is not significant ($p = 0.292$, MWU, two-sided), which suggests that shame-aversion also affects those women, who tend to underestimate their performance. For men, we observe no significant treatment effect in the absolute accuracy either.

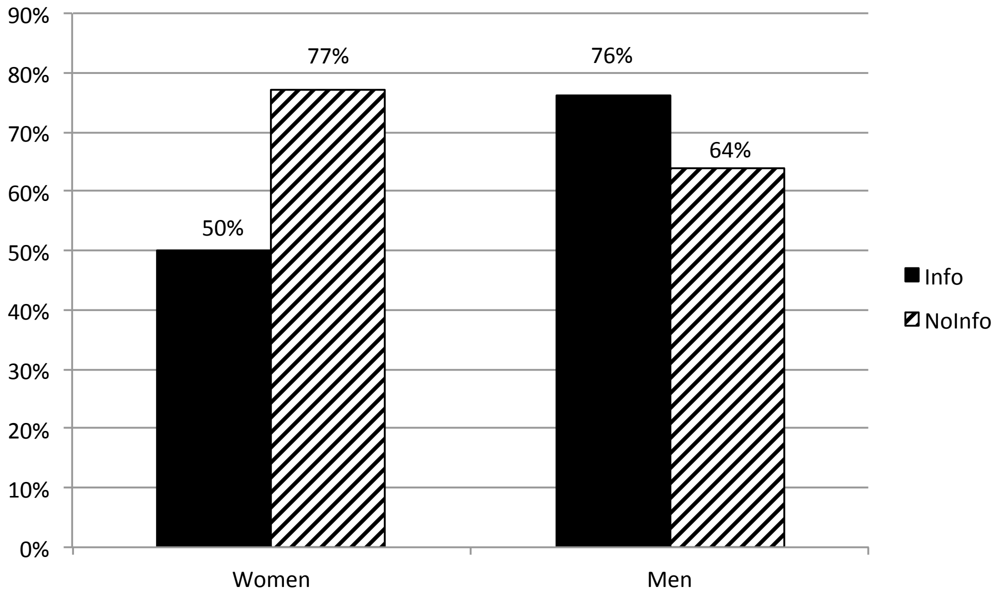
1.4 Causations of the Gender Difference in Shame

We observe that women state a worse rank in *Info* than in *NoInfo* while men (though insignificantly) state a weakly better rank in *Info* than in *NoInfo*. Before we try to explain

[29] If we use the agents' actual instead of optimal guessed ranks, gender differences across treatments are not affected, only the level of overestimation is lower as subjects in the baseline treatment performed slightly worse than participants in the experiment on average.

why men and women react differently to the potential exposure to shame, we analyze whether subjects adjust their guessed rank consciously or out of habit when someone else learns their guessed rank as well as their actual rank. To get a first indication, we ask the agents after they submitted their guessed rank, but before they know if it was correct, whether they think that their actual rank would rather be better or worse than their guessed rank, in case the latter turned out to be wrong. Subjects receive two tokens (0.5 Euro) if their guessed rank indeed turns out to be wrong and their answer is correct.

Figure 1.5: Percentage of Subjects Responding that their Actual Rank is Worse than Previously Guessed



These answers are illustrated in Figure 1.5. For women the answers differ across treatments: In *NoInfo* 77% of women respond that they would have a worse rank than they previously guessed, while in *Info* only 50% say so ($p = 0.059$, Chi^2, two-sided). This finding suggests that women are at least partly aware of their reaction to shame. They seem to consciously avoid shame even in the anonymous laboratory setting. Similarly, men seem to anticipate that they state better ranks in *Info*: In *Info* 76% indicate they would rather have a worse rank, while in *NoInfo* only 64% do so. However, as for men's guessed ranks, the treatment difference is not significant ($p = 0.311$, Chi^2, two-sided). When comparing the answers across gender within treatments, we obtain another indication that women in contrast to men seem to lower their rank (to some degree) consciously: In *Info* significantly more women than men think that their actual rank is better than stated ($p = 0.048$, Chi^2, two-sided), while there is no gender difference in *NoInfo* ($p = 0.355$, Chi^2, two-sided).

But why do only women consciously avoid shame? Our observations provide some evidence that women's shame-aversion may be attributed to social conventions. We ask the subjects in the questionnaire whether they think that overestimating oneself is deemed negative in society. 85% of all subjects say "yes". In comparison, only 21% say that underestimating oneself is deemed negative in society.[30] These observations can explain why people have shame when they overestimate themselves, but seem to have no shame when they underestimate themselves. People expect their social or self image to suffer, when overestimating themselves. Thus, people who care about their image may try to avoid to overestimate themselves (by stating worse ranks) when others can observe the accuracy of their self-assessment. Yet, we do not observe any gender difference in the answers to both questions. This implies that the mere willingness to keep one's image up cannot explain the gender difference in guessed ranks that we observe when shame is possible.

So why are only women shame-averse? First, as aforementioned, psychological studies suggest that women experience specific emotions stronger than men. In particular, these studies show that women experience more shame, embarrassment, and guilt, but less pride than men.[31] Hence, women might experience shame as defined in our context and the negative attitude towards overconfidence more intensely and thus react to it much stronger. Second, there might be a more negative societal attitude towards self-promoting women than towards self-promoting men (see e.g. Eagly, 1987; Rudman, 1998).

Third, we find that women, in contrast to men, are not expected to be overconfident.

[30] Possible answers to both questions are "yes" and "no". We varied the order of the two questions and restrict to the first question here (even more subjects say that overestimation is deemed negative when it is the second question). We pool both treatments as the results are very similar.

[31] See e.g Else-Quest et al. (2012) for a meta-analysis.

We ask subjects to estimate the average actual rank of women and men and the average guessed rank of women and men in their session (estimates can be given accurate to one-tenth of a rank). For each subject one question is randomly selected and the subject receives 16 tokens (4 Euros) if her answer does not differ more than +/- 1 rank from the true value.[32] For each subject, we calculate her *estimated accuracy of women* ("EAW") and her *estimated accuracy of men* ("EAM") as the difference between her estimated average rank of women/men minus her estimated average guessed rank of women/men. When asked about women, the majority of subjects (52%/64% in *Info*/*NoInfo*) expects them to be underconfident, the median of EAW being negative (-0.6 and -1 rank in *Info* and *NoInfo*, resp.); whereas when asked about men, the majority (50%/76% in *Info*/*NoInfo*) expects them to be overconfident, the median of EAM being positive (0.25 and 2 ranks in *Info* and *NoInfo*, resp.).[33] The expectations about men and women differ significantly in the sense that women are rather expected to be underconfident than men.[34] Figure 1.6 illustrates for each treatment the fractions of subjects expecting that women and men respectively, overestimate, underestimate or correctly estimate their rank in the experiment.

Given that women are not expected to be overconfident, a woman's reputation might suffer more when she is overconfident. In other words, a woman who is overconfident may be perceived more negatively than an overconfident man – or at least the woman may expect a stronger negative attitude towards her overconfidence. In reaction to this (perceived) unequal treatment, women may downgrade their beliefs about their performance if others can infer the accuracy of the beliefs. Hence, society rather observes that women appear less overconfident than men, which reinforces the expectations.

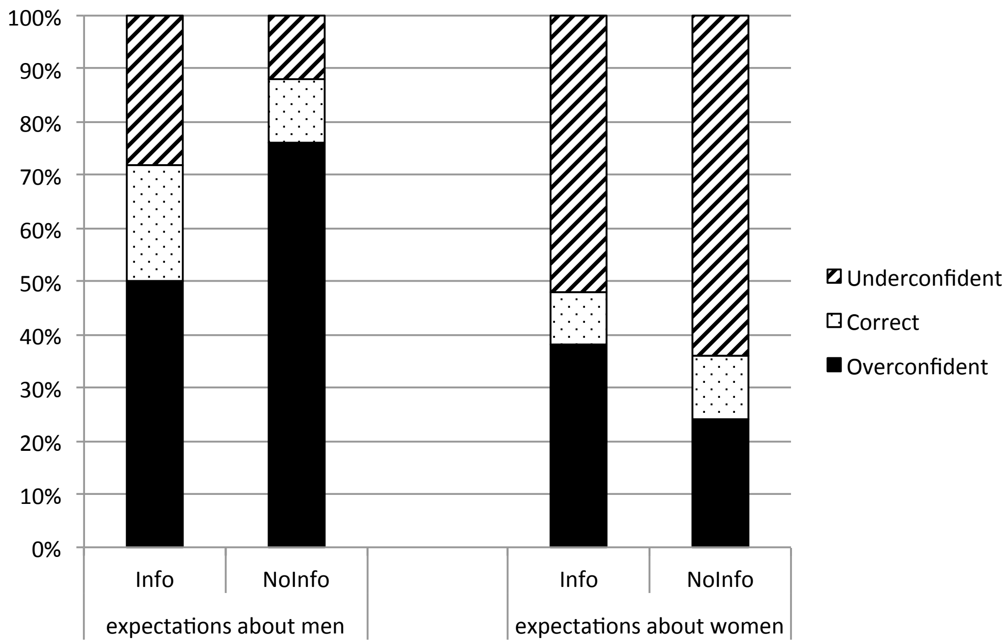
Fourth, educational differences may foster women's shame: Girls may rather be taught to

[32] We ask these four questions in four different orders to control for order effects. We varied the order of the gender as well as the order of guessed and actual rank. According to MWU-tests, there are no significant order effects.

[33] Recall that a better rank equals a lower number such that a positive (negative) EAW or EAM corresponds to expected overestimation (underestimation).

[34] Wilcoxon signed rank tests indicate that in both treatments EAW is higher than EAM ($p = 0.000/0.023$ for *NoInfo*/*Info*, two-sided) and more subjects expect that women are underconfident compared to men ($p = 0.000/0.053$ for *NoInfo*/*Info*, two-sided).

Figure 1.6: Expectations about the Accuracy of Subjects' Self-Assessment



be modest and reticent, while boys are taught to be self-confident and tough – which in turn may shape society's expectations. Women might have internalized these principles and may feel shame in case they do not behave accordingly and others observe it.

1.5 Conclusion

We analyze in a laboratory experiment whether individuals' (incentivized) self-assessment of their performance depends on whether its accuracy is observable to others. We find that women state a lower belief about their performance in case another person learns the accuracy of their self-assessment than if the accuracy is not observable. This behavior can be attributed to shame-aversion: women try to avoid feeling ashamed if another person observes that they overestimated themselves. Men, however, seem to be less prone to shame-aversion: They do not downgrade their beliefs in case the accuracy of their beliefs is observable.

Women's behavior as well as the observed gender difference in behavior cannot be explained by risk preferences, social preferences, preferences for competition, or overconfidence per se: Our results are based on treatment comparisons and the only variation across treatments is whether another person observes an individual's actual performance in addition to her estimated performance.

We also find some indications why only women are shame-averse. Women may expect or actually face a stronger social disapproval if they overestimate themselves: We elicit subjects' beliefs suggesting that men but not women are expected to overestimate themselves.

Note that the interaction of subjects in the experiment is completely anonymous and the only treatment variation is that one other subject observes an agent's self-assessment accuracy. Thus, the effect of shame is presumably even stronger in reality when the agent's actions are observable by more than one person and the agent and the observer(s) know each other.

Our observations contribute to the discussion why women are underrepresented in leading positions and why the gender wage gap is huge although women are equally educated and equally able (according to their grades) than men. Women might present themselves worse than men when applying for jobs, and might not negotiate their wages, because they want to avoid the shame they have if they turn out to perform worse than they claimed to. Similarly, they might not enter competitive or demanding work environments as this could be interpreted as a statement of being sufficiently confident to succeed and they are afraid that others might observe them to fail.

We consider our experiment as a first step to provide evidence of women's shame-aversion. Further research is needed to identify conditions in which shame-aversion occurs and to pin down the impact of shame-aversion in more complex settings, such as wage negotiations and competition. Moreover, it would be interesting to test whether women's shame aversion is culture-specific as society's expectations may differ in respect to the attitude towards women and their education.

1.6 Appendix A1

A1.1 Tables

Table A1.1: OLS Regression of Guessed Rank

	Coefficient (Robust Std. Error)			
	(1)	(1a)	(2)	(2a)
Performance	-0.334***	-0.328***	-0.335***	-0.322***
	(0.048)	(0.351)	(0.040)	(0.035)
Female	1.811**	1.507*	0.035	-0.238
	(0.749)	(0.710)	(0.392)	(0.665)
Info	0.928	0.737	-0.499	-0.884
	(0.563)	(0.635)	(0.564)	(0.578)
Female* *Info*			2.915**	3.39**
			(0.918)	(1.059)
Risk attitude	-0.370	-0.333	-0.375*	-0.332
	(0.202)	(0.239)	(0.188)	(0.221)
Constant	13.90***	11.16**	14.66**	10.197**
	(1.281)	(3.878)	(0.978)	(3.845)
Additional controls	No	Yes	No	Yes
# of observations	100	100	100	100
# of sessions	8	8	8	8
R-squared	0.312	0.412	0.341	0.450

The sample consists of all agents and the regression clusters on sessions. The additional controls are final school-grade, subject of study, age, number of siblings, degree of self-esteem, dummy for mother working or not.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

A1.2 Instructions (translated from German)

Welcome to this experiment. Please read these instructions carefully and follow the instructions on your screen when the experiment has started. At the end of the experiment you will be paid in cash according to your decisions and the decisions of other participants as described in the following. In addition, you receive a fixed payment of 4 Euros for showing-up. During the experiment you are not allowed to speak to other participants, to use cell phones or to start any other programs on the computer. If you break this rule, we have to exclude you from the experiment and its pay-out. If you have any questions, please raise your hand. An experimenter will then come to your seat to answer your questions.

During the experiment we do not speak of Euros but of points. Your earnings will be calculated in points first. At the end of the experiment your points will be converted into Euros whereupon applies:

1 point = 25 Eurocents.

The experiment consists of two parts and a questionnaire. Part 1 is explained in more detail in the following. As soon as all participants have finished part 1, you receive the instructions for part 2. Subsequent to part 2, there is a questionnaire.

Instructions Part 1

In Part 1 of the experiment you will be asked to add five two-digit numbers at a time. Please enter your result in the corresponding box and click "Confirm". Once you have confirmed your result, five new numbers appear, irrespective of whether your result was correct or wrong. On the screen, you will see whether your last result was correct or wrong and how many problems you have solved correctly and accordingly falsely so far. You are not allowed to use a calculator, but the provided scratch paper, to calculate the results. Overall, you have a time period of 7.5 minutes. During this time you can work on as many problems as you can. The remaining time will be shown top right on the screen. After the 7.5 minutes have passed, part 1 is completed and you will receive the instructions for part 2.

Your payment for part 1:

Whether part 1 or only part 2 of the experiment will be relevant for your payment, has been randomly determined at the beginning of the experiment. Two thirds of all participants will be paid for their performance in part 1. They get 2 points for each correctly solved problem. For one third of the participants, the payment will be based on part 2 of the experiment only. At the beginning of part 2, you will be informed on your screen whether you will be paid for part 1.

Procedure of part 1:

As soon as all participants have read these instructions, there will be a test phase of 2 minutes. During this time you can get used to the screen, the handling and the type of problems. You will receive no payment for the test phase. Subsequently, the 7.5 minutes – as described above – will start.

Instructions Part 2

At the beginning of the experiment, each participant was randomly assigned to the role A or B. Two thirds of the participants were assigned to role A (“participant A”) and one third to role B (“participant B”). Which role has been assigned to you, will be shown to you on your screen at the beginning of part 2. If you were assigned to role A, at the end of the experiment you will receive 2 points for every problem you solved correctly during part 1. If you were assigned to role B, you will not receive any payment for part 1.

Moreover, each participant B was randomly assigned to two participants A, i.e. one participant B and two participants A form a group of three. This assignment is random and anonymous. No participant learns the identity of the participants assigned to him, neither during nor after the experiment.

Based on the number of correctly solved problems in part 1, your **rank** within a ranking from 1 to 22 will be determined. For this ranking your number of correctly solved problems will be compared to the number of correctly solved problems of 21 other participants. These 21 participants have already completed part 1 of this experiment in this laboratory at an earlier point of time. In the following, we refer to the "former experiment". The former participants had exactly the same instructions as you in part 1.

The **ranking** is generated as follows:
The participant (either you or a participant of the former experiment) who solved the most problems correctly obtains rank 1. He who solved the second most problems correctly obtains rank 2, and so on. The participant with the lowest number of correctly solved problems obtains rank 22. If two participants (two participants of the former experiment or you and a participant of the former experiment) solved the same number of problems correctly, the one who solved less exercises falsely obtains the higher rank. If this number coincides as well, both participants obtain the same rank and the following (lower) rank is not assigned. Note that you will solely be compared to the participants of the former experiment. Your rank is independent of the other attending participants. Note that a higher rank equals a smaller number (e.g. rank 6 is a higher rank than rank 12. Rank 12 is a lower rank than rank 6).

Assessment:
Each participant A estimates his rank in this ranking from 1 to 22.

In each group, one of the two participants A will be selected based on the assessments of both participants A. The actual rank of the selected participant A will be relevant for the payment of participant B in his group, as described below. In each group, the participant A **who assessed himself on a higher rank** (i.e. he who stated the **smaller** number at his assessment) will be selected. If both participants assess themselves on the same rank, it is randomly decided who is selected.

For the time being, participant B learns the rank-assessments of both participants A, and at the end of the experiment, he also learns their actual ranks.

[The preceding sentence is replaced as follows in Treatment *NoInfo*: *Participant B learns* *the rank-assessments of both participants A, but not their actual ranks.*]

Participant B also estimates his rank, his assessment does not affect other participants and is not paid.

Example "selection": *In a group, participant A1 estimates that his rank is 12. The other participant A2 estimates that his rank is 6. The participant A2 who estimated that his rank is 6 is chosen for the payment of participant B.*

Payment Part 2:

Participant A receives 50 points if his assessment exactly matches his actual rank. If that is not the case, he receives 0 points.

Participant B receives a payment that might depend on the **actual rank** of the selected participant A in his group.

Participant B receives the following number of points: $(22 - rank) \cdot 3$.

The payment of participant B is higher, the higher the rank. With a probability of 1/2, the relevant rank for the payment of participant B is the actual rank of the selected participant A. With a probability of 1/2, it is a rank between 1 and 22, which is determined randomly by the computer, where each rank between 1 and 22 is equally likely. At the end of the experiment participant B learns the actual ranks of both participants A, who have been assigned to him, and whether his payment was determined by the actual rank of the selected participant A or by the randomly drawn rank.

[The preceding sentence reads as follows in Treatment *NoInfo*: *Participant B does neither learn the actual ranks of the participants A who have been assigned to him nor whether his payment was determined by the actual rank of the selected participant A or by the randomly determined rank.*]

Example "payment":

In one group, a participant A1 estimates that his rank is 12. His actual rank is 12. The other participant A2 estimates that his rank is 6. His actual rank is 9. Participant A1 guessed his rank correctly and receives 50 points for his assessment, participant A2's assessment is wrong and he receives 0 points for his assessment. Participant A2 guessed a higher rank for himself than did participant A1. Therefore, participant A2 is selected for the payment of participant B. With a probability of 1/2 participant B receives a payment of $(22 - 9) \cdot 3$ points = 39 points, i.e. the actual rank of the selected participant A is relevant. With a probability of 1/2 a rank r between 1 and 22 is chosen randomly and participant B receives $(22 - r) \cdot 3$ points.

Information of the participants:

Participant A learns:

- after submitting his assessment: whether he was selected or not
- at the end of the experiment: his actual rank

Participant B learns:

- after the submissions of the evaluations: the estimated rank of the selected as well as of the not selected participant A
- at the end of the experiment: the actual rank of the selected as well as of the not selected participant A

[In treatment *NoInfo* the information of participant B reads as follows:

Participant B learns:

- *after the submissions of the evaluations: the estimated rank of the selected, as well as of the not selected participant A*

Participant B does not learn:

- *the actual ranks of both participants A*
- *whether his payment was determined by the actual rank of the selected participant A or by the randomly determined rank*]

Chapter 2

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

2.1 Introduction

While earlier, mainly psychological and social psychological studies claim that people are overconfident, recent economic studies show that individuals' self-assessment is rather precise or underconfident.[1] As a biased self-assessment can lead to systematic biases in individuals' decision making the topic is of high interest to economists.[2] Yet, barely any

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

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

research has been done to analyze how overconfidence in comparison to underconfidence is perceived by others, and whether individuals adapt their self-assessment to others' perception.

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

The experiment consists of two parts. In part 1 all subjects perform an incentivized real effort task which serves as the basis for their self-assessment. In part 2 two thirds of the subjects (agents) are assigned a rank based on their relative performance, whereas each rank is assigned to two subjects. The two agents having the same rank are assigned to one of the remaining participants (principals). Both agents estimate their relative rank and the principal learns by how many ranks each of them over- or underestimated himself.[3] In treatment *SYMP* the principal chooses to whom of the two agents he wants to give 5 Euros. In treatment *PERF* the principal has a monetary incentive to choose the agent who performs better in a repetition of the real effort task. The only information the principal gets is the deviation of the agents' self-assessments and the information that both agents have the same actual rank. This element of the design is essential as subjects on higher ranks are more likely to be underconfident, while subjects on lower ranks are

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

more likely to be overconfident, due to the limited scale for self-assessment. If subjects' actual ranks differed, principals might choose the underconfident agent not because they prefer underconfidence, but because underconfidence might signal a higher actual rank.

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

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

[4]This is in line with women's shame of overestimation observed by the experimental study reported in the first chapter.

[5]Note that the accuracy of agents' self-assessment is still observable in the control treatments, but

somehow imprinted the social norm of modesty and even downgrade their self-assessment in environments in which the (monetary) need for modesty is absent.[6]

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

Only very few recent papers consider the impact of one's self-assessment on others, and whether individuals account for others' perception when stating their self-assessment. Ewers and Zimmermann (2012) theoretically and experimentally analyze whether individuals bias their self-assessment due to image concerns. They find that individuals state a higher self-assessment if reports are observed by an audience (anonymity is lifted) than in private.[7] Yet, they find that self-assessments do not differ if true performance is also

the agents do not have a monetary incentive to be chosen.

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

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

publicly revealed, thus subjects do neither try to signal high ability nor modesty if the accuracy of one's self-assessment is observable. However, in contrast to our study, in Ewers and Zimmermann (2012) individuals have no monetary incentives for strategically biasing their self-assessment. Moreover, their study does not analyze how one's self-assessment is perceived by others, i.e. if one's social image or expected ability is actually increased by stating a higher self-assessment.

These issues are addressed in an experimental study by Charness et al. (2012). They investigate whether individuals bias their stated confidence about their performance strategically to deter or motivate others to enter a two-player tournament, and whether others react to it. They find that males inflate their stated confidence when deterrence is strategically optimal, and that men and women deflate their confidence if encouraging entry is strategically optimal. Moreover, they observe that individuals are less likely to enter the competition, the higher the confidence of the other person is. In line with these results Reuben et al. (2012) observe that men inflate their self-assessment to be voted as the group leader, which turns out to be a successful strategy.[8] However, in both studies individuals' actual performance is unknown and might strongly differ. Expecting all individuals to exhibit the same bias in self-assessment, the ranking of subjects' self-assessment most likely corresponds to the ranking of subjects' true performance. Thus, it is not the bias in self-assessment, which reveals information about the actual performance, but the self-assessment per se. This is different in our experiment, in which both agents have the same actual (relative) performance, enabling us to investigate whether the bias in self-assessment serves as a performance signal or influences one's image. In many real-life situations, in which individuals have to assess their performance, e.g. in promotion interviews or wage negotiations, the true performance is somehow appraised or at least partly known. Thus, the accuracy of the self-assessment might also be evaluated. On the one

"overconfidence is social signaling bias".

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

hand it might influence whether an individual is liked or not, on the other hand it might serve as a performance signal.

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

This chapter might also contribute to the literature observing gender differences in self-assessment.[9] We suggest that the observed antipathy towards overconfidence adds to the explanation of the gender difference in self-assessment as women seem to experience emotions, i.e. the negative attitude towards overconfidence, stronger than men (see e.g. Brody, 1997; Grossman and Wood, 1993). In addition, they might even be punished more harshly than men when being self-confident (Eagly, 1987; Rudman, 1998). Our results could moreover provide an explanation for the findings of the first chapter, which observes that women have shame to overestimate themselves in public, while men have not.

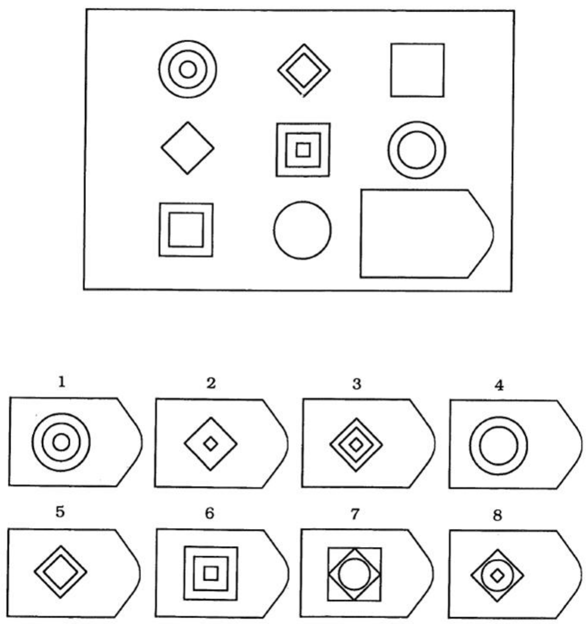
The rest of this chapter is structured as follows. In the next section, we describe the experimental design and the two different treatments. In Section 2.3 we present the main experimental results, i.e. the selection behavior of the principals. In Section 2.4 we analyze whether agents anticipate principals' preferences and whether they strategically bias their self-assessment. We conclude in Section 2.5.

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

2.2 Experimental Design

The experiment consists of two parts and a questionnaire, with separate instructions for each part. In part 1 all participants conduct a real effort task (task 1), which is solving Raven's Advanced Progressive Matrices (APM), a measure of cognitive ability (Raven, 2000). For each matrix participants have to select one out of 8 symbols fitting the visual pattern of the matrix. An example of a matrix is given in Figure 2.1.

Figure 2.1: Example of a Raven Advanced Progressive Matrix



In this task ability and effort are needed to succeed in the task. The participants have five minutes to solve as many matrices as possible. After choosing a symbol, they receive feedback whether or not their chosen symbol is correct, and thereafter, the next matrix appears. Once having chosen a symbol, they cannot go back and correct it, neither is it possible to skip a matrix without making a choice. On subjects' screens the remaining time as well as the number of correctly and wrongly solved matrices is displayed. The maximum number of matrices is 22, none of the subjects managed to get to the last matrix. Subjects are informed about their absolute performance, but neither about their relative performance nor the performance of others. They receive 5 tokens for each matrix they solve correctly. For each wrong answer 5 tokens are deducted from their earnings. Yet, they receive at least 0 tokens for part 1. During the whole experiment participants earn tokens, which are converted into Euros at the end of the experiment, at an exchange rate of 1 Euro for 10 tokens. Before the five minutes start, participants solve two matrices as a trial without payment. After the five minutes part 1 is finished and the instructions for part 2 are distributed. The instructions for both parts are read aloud.

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

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

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

As the scale for subject's relative self-assessment is limited, it naturally occurs that subjects on a higher rank are more likely to underestimate their rank and subjects on a lower

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

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

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

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

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

If one agent over- and the other agent underestimates himself each by 1 rank, the decision the principal takes for this case becomes relevant. The principal takes the decision for 12 different cases, which are listed in Table 2.1, where a negative deviation means that the agent underestimates himself and a positive deviation means that the agent overestimates himself. As accuracy might matter for the principal's choices, using the SVM decreases the number of sessions needed and allows for a cleaner data analysis, as each principal takes the decisions for the same cases. We picked these 12 cases in order to include the most realistic outcomes and to check the robustness of principals' preferences.[12]

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

Table 2.1: The Cases for which Principals Take a Decision

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

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

In cases 1a and 1b both agents have the same absolute deviation. In cases 2a - 6b agents' accuracies differ and the less accurate agent is listed on top. In the a-cases the more self-confident agent has a larger absolute deviation, in the b-cases he has a smaller absolute deviation. For every case the principal selects one of the two agents, i.e. the one who he wants to receive the 50 tokens. His choice becomes relevant for the case, which actually applies to the two agents assigned to him. If none of the cases 1-12 applies to the two agents, principals take a 13th decision: Principals choose whether the agent who estimates the higher (better) rank or the agent who estimates the lower (worse) rank shall receive the 50 tokens. The accuracy of the agents' estimated ranks is not considered in this decision.[13] If the two agents estimate the same rank, and therefore have the same deviation, a case not covered by the decisions of the principal, chance determines which agent receives the 50 tokens.[14]

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

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

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

Treatment Performance (*PERF*):

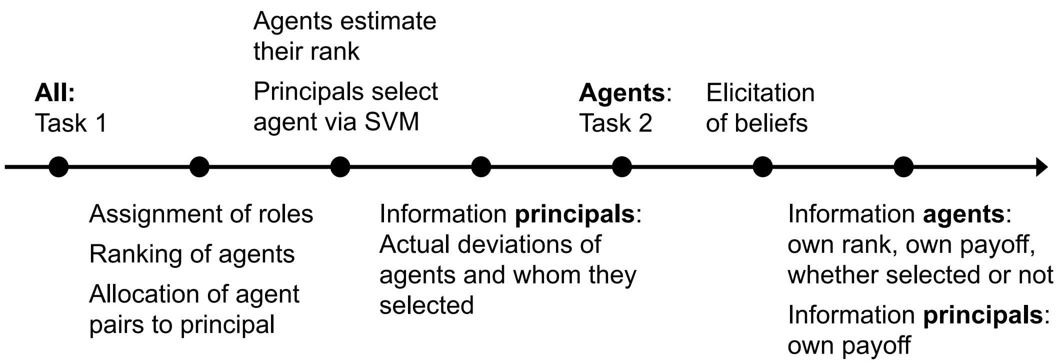
In *PERF* the principal takes the same 13 decisions as in *SYMP*, i.e. using the strategy vector method, the principal chooses one out of two agents, based on the deviations of their self-assessments. However, the motivation for the choice of the principal is different. While in *SYMP* the principal's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

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

The principals do not participate in task 2. They bet on the agent chosen in cases 1-13. If the agent, who the principal bets on, wins the competition, the principal receives 50 tokens. If both agents estimate the same rank, chance determines on whom the principal bets. After his decisions the principal learns about the actual deviations of both agents and which case is relevant. As in *SYMP*, the principal does neither learn the actual nor the estimated ranks of the agents. An agent receives 50 tokens if the principal bets on him. This monetary incentive is introduced to analyze whether agents bias their self-assessment in order to be selected. The analysis is shown in chapter 2.4. The payment is only announced on the agents' screens, thus principals do not know that an agent receives money when they bet on him, in order to avoid that feelings of sympathy and inequity aversion influence principals' choices. An agent learns whether the principal bet on him, but only after task 2. This is common knowledge. An agent does not learn the estimated rank of the other agent, only when the two estimated the same rank, they learn that a chance move decides on whom the principal bets.

In both treatments all subjects are asked additional incentivized questions. In particular they estimate the mean deviation of agents and the choice behavior of principals. After the announcement of the payoffs, subjects complete a questionnaire asking for their age, gender, subject of study, choice motivation, and a self-assessment of risk preferences (Dohmen et al., 2011). Figure 2.2 illustrates the course of the treatment *PERF*. *SYMP* has the same course except that there is no task 2.

Figure 2.2: The Timeline of *PERF*



Experimental Procedure

We conducted the computerized experiment in the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) at the University of Munich during spring 2012. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007) and participants were recruited via ORSEE (Greiner, 2004). In total 216 subjects participated in the experiment (mainly students from the universities in Munich). We ran 5 sessions of *SYMP* and 4 sessions of *PERF*, whereof in each treatment 2 sessions were control sessions for agents, on which we will comment in Section 2.4.

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

2.3 Experimental Results

2.3.1 Treatment *SYMP*

Table 2.2 reports the shares of principals choosing either the one or the other agent in each of the 13 different cases in *SYMP*. There are 40 principals in total. In 2 sessions (16 principals), which were control treatments for the agents, the principals' choices were only hypothetical and had no consequence on the decision, which agent received the 50 tokens. This decision was taken by chance to analyze whether agents' self-assessment is influenced by the principals' choices. The analysis is provided in the next section. The principals' decisions in these 2 sessions are not significantly different from the principals' choices in the other 3 sessions with actual choices, thus we pool the data.[15]

Table 2.2: Principals' Selection Behavior in *SYMP*.

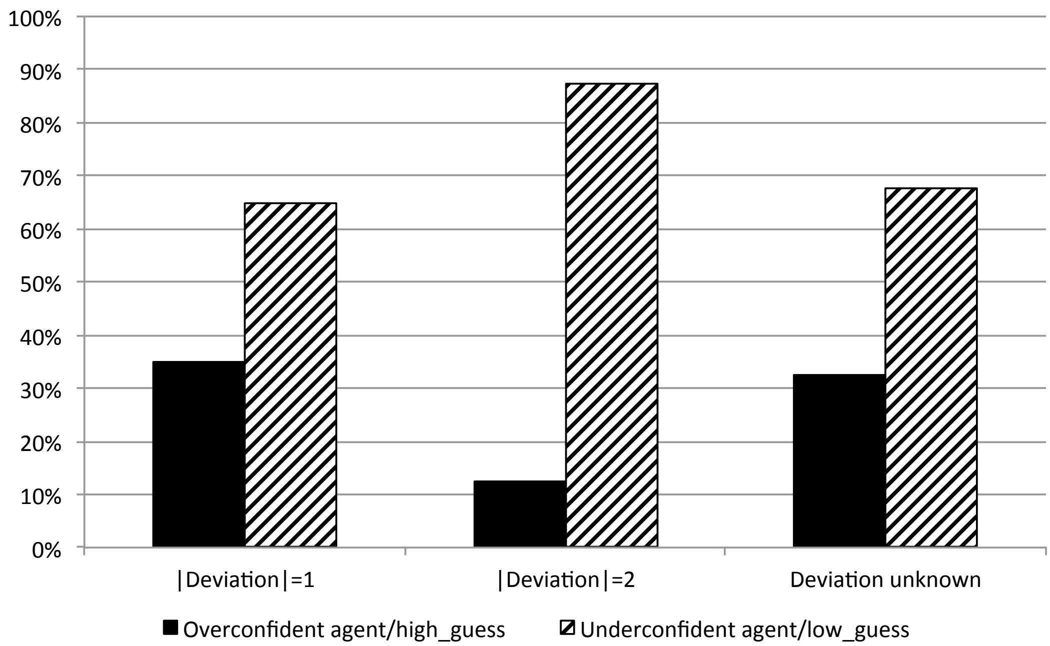
Case	1a	1b	2a	2b	3a	3b	
Deviation of agent 1	1	2	2	-2	3	-3	
Deviation of agent 2	-1	-2	-1	1	-1	1	
Share selecting agent 1	35%	12.5%	5%	32.5%	2.5%	27.5%	
Share selecting agent 2	65%	87.5%	95%	67.5%	97.5%	72.5%	
Binomial test (p-value)	0.040	0.000	0.000	0.019	0.000	0.003	
McNemar test (p-value)	0.004	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

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

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

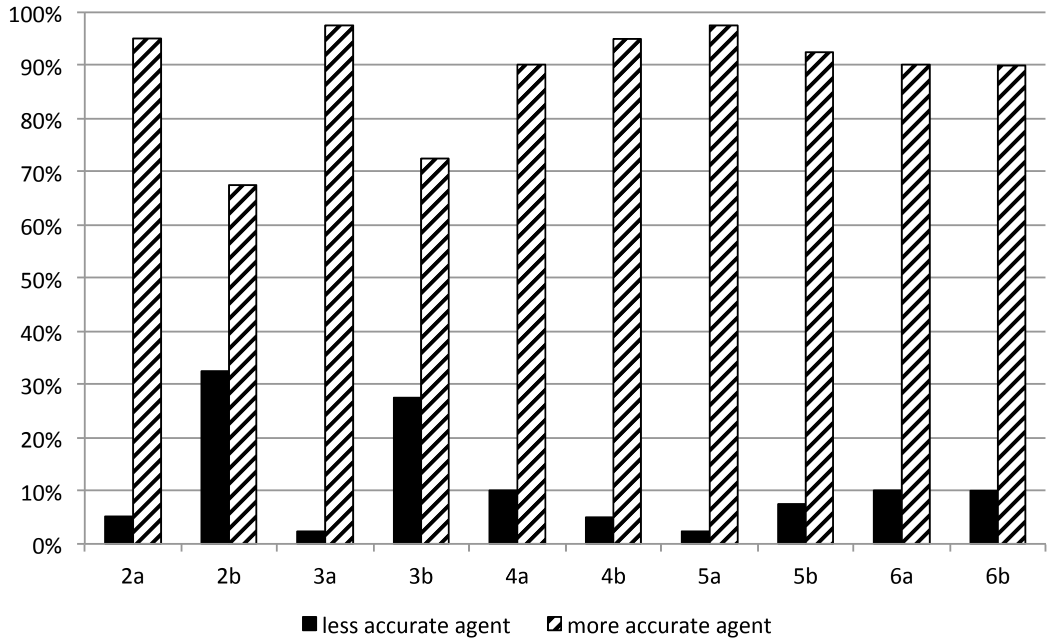
The first important observation is that principals favor agents who are underconfident rather than overconfident. In case 1a 65% of principals (26 out of 40) reward the agent who underestimates his rank by 1 rank, whereas only 35% of principals (14 out of 40) reward the agent who overestimates his rank by 1 rank. This distribution is significantly different from a 50:50 split (one-sided binomial test, p=0.040). The result is even stronger in case 1b, in which 87.5% of principals (35 out of 40) reward the agent who underestimates his rank by 2 ranks and only 12.5% of principals (5 out of 40) reward the agent who overestimates his rank by 2 ranks (one-sided binomial test, p=0.000). In case 7 in which the absolute deviation is unknown, 67.5% of principals reward the agent estimating the lower rank (*low_guess*) and only 32.5% reward the agent estimating the higher rank (*high_guess*) (one-sided binomialtest, p=0.019). These results are illustrated in Figure 2.3.

Figure 2.3: Principals' Choices for Cases 1a, 1b, and 7 in *SYMP*



If agents' deviations differ (cases 2a-6b), the agent with the smaller absolute deviation is selected significantly more often - in every case - as shown by the results in Table 2.2. Figure 2.4 illustrates principals' choice behavior, whereas for each case the right bar represents the share of principals selecting the more accurate agent.

Figure 2.4: Principals' Choices for Cases 2 - 6 in *SYMP*



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

These results are confirmed by probit regressions, which are reported in Table 2.3.[16]

[16]The results are robust to OLS regressions.

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

Table 2.3: Probit of Selection of the Less Accurate Agent in *SYMP*

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

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

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

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

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

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

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

2.3.2 Treatment *PERF*

The principals' choice behavior for all 13 decisions in *PERF* is reported in Table 2.4.

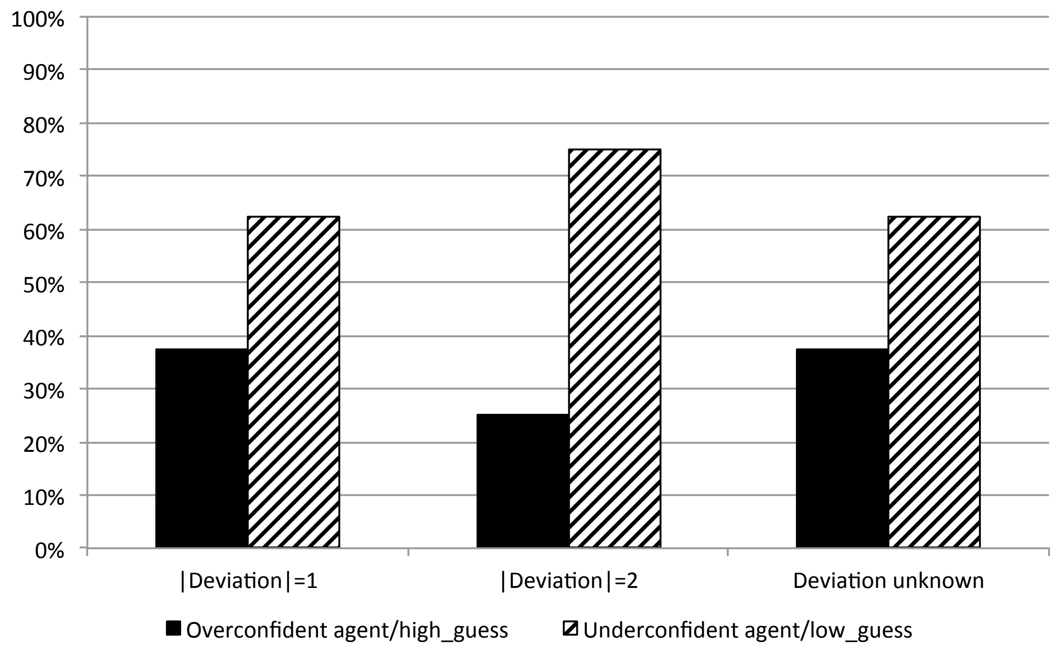
Table 2.4: Principals' Selection Behavior in *PERF*

Case	1a	1b	2a	2b	3a	3b	
Deviation agent 1	1	2	2	-2	3	-3	
Deviation agent 2	-1	-2	-1	1	-1	1	
Share selecting agent 1	37.5%	25%	12.5%	53.1%	15.6%	46.9%	
Share selecting agent 2	62.5%	75%	87.5%	46.9%	84.4%	53.1%	
Binomial test (p-value)	0.108	0.004	0.000	0.430	0.000	0.430	
McNemar test (p-value)	0.424		0.004		0.002		
Case	4a	4b	5a	5b	6a	6b	7
Deviation agent 1	1	-1	2	-2	2	-2	high
Deviation agent 2	0	0	0	0	1	-1	low
Share selecting agent 1	28.1%	53.1%	15.6%	40.6%	21.9%	46.9%	37.5%
Share selecting agent 2	71.9%	46.9%	84.4%	59.4%	78.1%	53.1%	62.5%
Binomial test (p-value)	0.010	0.430	0.000	0.189	0.001	0.430	0.108
McNemar test (p-value)	0.057		0.057		0.021		

Every principal took all 13 decisions. # of observations is 32. A positive deviation represents "overconfidence", whereas a negative deviation represents "underconfidence". The Binomial test tests for a difference between selection rates and a 50:50 split for each case. The McNemar test tests for a difference between cases a and b.

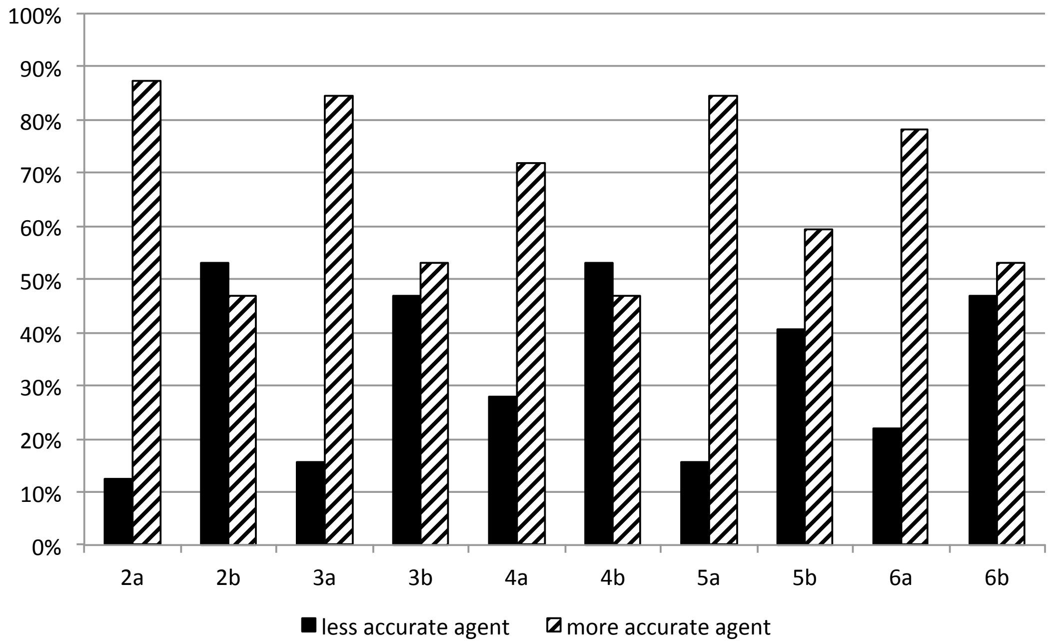
We observe that the underconfident agent is selected more often than the overconfident agent when the agents' absolute deviations are equal or unknown. In case 1a 62.5% of principals (20 out of 32) bet on the agent who underestimates himself by 1 rank, while only 37.5% (12 out of 32) bet on the agent who overestimates himself by 1 rank. This result is not significant, but close to marginally significant (one-sided binomial-test, $p = 0.108$). Yet, in case 1b 75% of principals (24 out of 32) bet on the agent underestimating himself by 2 ranks, while only 25% (8 out of 32) bet on the agent overestimating himself by 2 ranks. This is significantly different from a 50:50 split (one-sided binomial-test, $p = 0.004$). In case 7 62.5% select the agent estimating the lower rank and 37.5% select the agent estimating the higher rank ($p = 0.108$). The selection rates for cases 1a, 1b, and 7 are illustrated in Figure 2.5.

Figure 2.5: Principals' Choices for Cases 1a, 1b, and 7 in *PERF*



The principals' choice rates in *PERF* for the cases 2-6 are illustrated in Figure 2.6. Almost all principals select the agent stating the *lower* rank if he has the smaller absolute deviation (cases 2a, 3a, 4a, 5a, 6a). Yet, there is not a single case, in which the agent stating the *higher* rank is selected significantly more often than the agent estimating the lower rank, even if he has the smaller absolute deviation (cases 3b, 4b, 5b, 6b). In

Figure 2.6: Principals' Choices for Cases 2 - 6 in *PERF*



these cases selection rates are not significantly different from a 50:50 split as confirmed by binomial tests reported in Table 2.4. Note that even for cases 4b and 5b, in which one agent correctly assesses himself and the other agent underestimates himself, roughly 50% of principals select the underconfident agent. This shows that principals, betting on the underconfident agent in cases 2b and 3b, do not only choose the underconfident agent to avoid betting on the overconfident agent (as it happens in *SYMP*). Instead, they seem to be convinced that the underconfident agent will perform better in task 2. This also becomes apparent when conducting the same two probit regressions as in *SYMP*. The results are reported in Table 2.5.[18]

As above, the dependent variable is 1 if a principal selects the less accurate agent. The results in column (1) show that the less accurate agent is more likely to be selected if he is underconfident and the more accurate agent is overconfident than if he is overconfident and the more accurate agent is underconfident. However, in contrast to *SYMP* we observe that the coefficient of *inacc_low* in column (2) is also highly significant. This means that

[18]The results are robust to OLS regressions.

Table 2.5: Probit of Selection of the Less Accurate Agent in *PERF*

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

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

even if the more accurate agent is assessing himself correctly (or the deviation of his self-assessment goes in the same direction), the less accurate agent is more likely to be selected if he is underconfident than if he is overconfident.

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

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

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

2.3.3 Treatment Comparison

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

These treatment differences are also confirmed by two probit regressions, which are reported in Table 2.6.[20] As in the earlier probit regressions, the dependent variable is 1 if the principal selects the less accurate agent, and it is 0 if he selects the more accurate agent. We conduct two separate regressions for the cases 2-3 and the cases 4-6. *inacc_low* is a dummy, being 1 for the cases, in which the less accurate agent is the agent stating

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

[20]The results are robust to OLS regressions.

the lower rank, *PERF* is a treatment dummy, and *PERF*inacc_low* the interaction of the two, in order to check whether the treatment difference is even larger for the agent stating the lower rank. In both regressions we cluster standard errors on principals.

Table 2.6: Probit of Selection of the Less Accurate Agent

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

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

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

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

2.4 Further Analysis

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

2.4.1 Anticipation of Principals' Preferences

After the principals' selection, agents estimated principals' selection behavior for the cases listed in Table 2.7. For each case they estimated how many of the 8 principals in their session selected the agent with the deviation listed first. One question was randomly chosen for payment and participants received 1 Euro if their answer did not differ more than +/-0.5 from the correct answer. The agents' average estimations are reported in Table 2.7.

Table 2.7: Agents' Average Estimation of Principals' Choice Behavior

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

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

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

2.4.2 Strategic Adaptation of Self-Assessment

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

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

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

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

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

optimal guessed rank instead of the actual rank to calculate his deviation. The optimal guessed rank is the rank that is most likely assigned to an agent, given his performance and the performance distribution of all agents in all treatments.[22] We calculate the deviation of an agent as his optimal guessed rank minus his actual rank. Thus, a negative deviation represents underconfidence, a positive deviation represents overconfidence. Table 2.8 gives an overview of agents' average guessed and optimal guessed ranks as well as their average deviation in all treatments.

Table 2.8: Guessed and Optimal Guessed Ranks of Agents

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

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

While agents are slightly underconfident in *SYMP* (average deviation -0.13), they are slightly overconfident in the other three treatments.[23] The deviation across *SYMP* and *PERF* is significantly different (MWU, two-sided, p=0.021). We cannot differentiate whether agents state a higher rank in *PERF* or a lower rank in *SYMP*.[24] Moreover, the differences between treatments and control treatments are not significant (MWU, two-sided, SYMP: p=0.562; *PERF*: p=0.236). As participants do not seem to anticipate the preference for underconfidence in *PERF* the absence of a difference in deviations between *PERF* and *PERF-CON* is not surprising. However, we expected a lower self-

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

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

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

assessment in *SYMP* than in *SYMP-CON*. A possible explanation for why this is not the case is that in *SYMP-CON* as in *SYMP*, principals get to know whether an agent over- or underestimated himself. The first chapter shows that women have shame to overestimate themselves, while men have not. Women's shame might bias their self-assessment in the same direction as the ambition of being selected by the principal, leading to a low self-assessment in both treatments. As men seem to be less prone to this kind of shame, we conduct ordered probit regressions controlling for gender, in order to explore whether gender differences in stated ranks exist.

Table 2.9 reports the results of ordered probit regressions for each gender separately.[25] The dependent variable is the guessed rank. The independent variables are the optimal guessed rank (as performance measure), a dummy for *CON*, being one for *PERF-CON* and *SYMP-CON*, a dummy for *PERF*, being one for PERF and PERF-CON, and an interaction dummy for *PERF-CON*.[26] We also control for risk aversion.

Table 2.9: Ordered Probit of Guessed Rank

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

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

Column 1 reports the results for men, column 2 reports the results for women. We observe that for men the coefficient of *CON* is marginal significant (p=0.072) and negative, i.e.

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

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

men rank themselves lower in *SYMP* than in *CON*. It seems as if men strategically lower their self-assessment in *SYMP* to increase their selection chances.

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

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

2.5 Conclusion

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

Elicited beliefs of agents show that they do not expect the principals to select the underconfident agent more often when performance is the critical selection criterion. However, they anticipate that underconfidence is deemed more likeable than overconfidence.

The comparison of stated self-assessments to a treatment, in which the principal cannot make a selection choice (non-strategic setting) shows that men slightly deflate their self-assessment strategically to be rewarded by the principal. Yet, women do not. An explanation might be given by women's shame of overestimation (as reported in the first chapter) suggesting that women even deflate their self-assessment in the non-strategic setting, as principals still learn the deviation of agents' self-assessment. Thus, apart from monetary consequences women might expect their social image to suffer. Moreover, besides the conscious adaptation of self-assessment, individuals might have internalized the negative attitude towards overconfidence and might be modest in situations, in which no strategic concerns are at place.

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

2.6 Appendix A2

A2.1 Tables

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

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

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

A2.2 Instructions (translated from German)

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

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

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

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

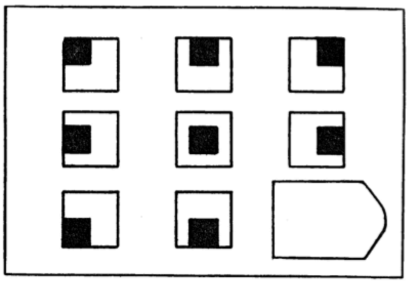
1 Token = 10 Eurocent.

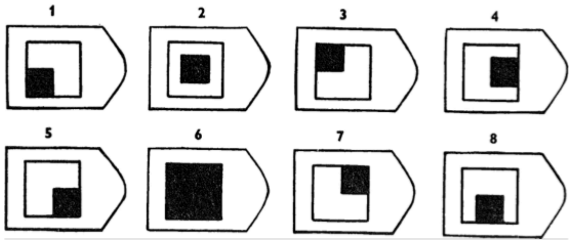
The experiment consists of two parts and a questionnaire. Part 1 will be explained below. Once all participants have finished part 1, you will receive the instructions for part 2. After part 2 a questionnaire will follow.

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

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

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





In this example the correct solution is symbol number 5.

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

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

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

Your payment for Part 1

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

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

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

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

The Course of Part 1

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

Instructions Part II[27]

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

Ranking

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

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

Remember:

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

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

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

Self-assessment of the B-persons

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

Matching of B-pairs

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

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

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

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

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

Assignment

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

[**Payment Part 2**

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

[[**A-persons' bet**

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

A-persons' selection decisions

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

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

Case 1:

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

Case 2:

- One person is overestimating himself by 2 ranks

- One person is underestimating himself by 2 ranks

...

Case 12:

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

- One person is underestimating himself by 2 ranks

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

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

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

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

A-person will also take a more general selection decision, called case 13.

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

In Case 13 the A-person has the following choice:
[[To choose]] [[To bet on]]
- the person who estimated the higher (better) rank
- the person who estimated the lower (worse) rank

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

Equal self-assessment of the B-persons:

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

A-persons' information

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

Summary and course of action of part 2

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

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

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

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

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

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

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

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

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

Chapter 3

Promises and Image Concerns*

3.1 Introduction

Cooperation among interacting partners is essential for economic success in many situations, as joint value creation often exceeds individual achievements. These situations become challenging as soon as cooperation cannot be contractually enforced, but relies on mutual trust by the interacting partners. Among a large literature focusing on how to improve cooperation, various experimental studies show that communication can be an effective tool to enhance it (see, e.g. Bochet and Putterman, 2009; Cooper et al., 1992; Ellingsen and Johannesson, 2004). While several articles analyze whether cheap talk can be effective and how this depends on the communication protocol and the game structure (see for instance Blume and Ortmann, 2007; Camera et. al., 2011; Ellingsen and Östling, 2010; Kriss et al., 2011; Mohlin and Johanneson, 2008), we contribute to the literature focusing on *why* individuals stick to a commitment, given that rationality predicts a deviating behavior. In particular, we analyze whether and to what extent social image concerns motivate people to stick to a given promise. More precisely, as breaking a promise is deemed negative in society, avoiding the image of being a promise breaker might induce individuals to keep their word. Consequently, we study whether an individ-

*This chapter is based on joint work with Miriam Schütte from the University of Munich.

ual is more likely to act in line with a given promise if its violation is more obvious to its receiver.

In order to test whether social image concerns influence promise keeping behavior, we conduct a controlled laboratory experiment. Here, subjects are randomly matched in pairs of two and play a one-shot sequential trust game similar to the one used in Charness and Dufwenberg (2006). A first mover (`A`) decides whether to enter the game or to opt out, the latter choice inducing a low outside option for both players. If `A` enters the game, a second mover (`B`) chooses between a selfish option, yielding a payoff of zero for `A`, and cooperation, in which case a chance move determines whether `A` gets a positive payoff or 0.[1] Prior to the strategic decisions, the second mover sends one out of three pre-defined messages to the first mover, one of which is a promise to cooperate. In order to test for social image concerns, we vary the ex-post observability of the second mover's action. While in condition *Rev* `A` learns `B`'s action choice, in condition *NoRev* she cannot infer whether a payoff of zero is due to `B` behaving selfish or just to bad luck.[2] We hypothesize that a higher share of `Bs` cooperate if `B`'s action is revealed to `A` (*Rev*) than if it is concealed (*NoRev*), assuming that a fraction of `Bs` has a preference for avoiding the image of being a promise breaker.

By the choice of our experimental design, we attempt to differentiate social image concerns from other possible reasons for promise keeping by second movers. Up to now, the literature mainly provides two motivations why individuals might stick to their promises. Firstly, Charness and Dufwenberg (2006) explain promise keeping by simple guilt, i.e. the aversion to disappoint other people's expectations, as introduced by Batigalli and Dufwenberg (2007). If the second mover promises cooperation, the first mover expects a higher payoff, which increases the second mover's guilt in case he refuses to cooperate. However, in our experiment only the game structure and the payoffs are common knowledge, but `Bs` are privately informed about the revelation condition. `As` are not even aware that different

[1]While rational behavior predicts the second mover to behave selfish, and therefore the first mover not to enter the game, mutual cooperation is the unique Pareto-optimal outcome, which generates the highest joint payoff.

[2]Conditions are assigned randomly to pairs.

conditions exist. Thus, As' first order beliefs, and consequently Bs' second-order beliefs should not vary across conditions, inducing the same amount of guilt for non-cooperation in both conditions.[3] Secondly, Vanberg (2008) claims that subjects have a preference for keeping their promises per se, independent of others' expectations. This assumption cannot explain a difference in Bs' behavior across conditions either, as the preference for keeping a promise should be independent of A's ex-post information.

Yet, as the revelation of B's action choice might also induce a concern of being perceived as selfish (Tadelis, 2011), we conduct a control treatment without communication (*NoCom*). We claim that the effect of revelation on behavior in treatment *Com* is larger than the respective effect in *NoCom*, indicating that the differential effect is due to the mere aversion of being perceived as a promise breaker, additional to the aversion to an egoistic image.

With pre-play communication, we observe marginally significantly more cooperation in *Rev* than in *NoRev*. This effect does not seem to be driven by shame to be selfish alone, as without communication revelation even marginally decreases cooperation rates in *Rev* compared to *NoRev*. However, although conditions are identical at the pre-play communication stage, the number of promises sent is significantly higher in *Rev* than in *NoRev*. Thus, the higher *Roll* rate in *Rev* might only be driven by a higher number of promises and not by image concerns of being perceived as a promise breaker. When comparing the share of promises kept, we do observe a slightly higher rate in *Rev* (85%) than in *NoRev* (81%), however the difference is not significant. Thus, we fail to prove our hypothesis that avoiding the image of being perceived as a promise breaker plays a significant role in the individual decision to keep a given promise.

It is worth noting that the high promise keeping rate without revelation (81%) limits the scope for further increase. In treatment *Com1*, where Bs can choose between a promise to cooperate, a statement of intent, and an empty message, this high promise keeping rate might be partly due to the fact that Bs, who attempt to influence their interaction

[3] Otherwise A might expect B to choose *Roll* with a higher probability if his choice is revealed, inducing higher simple guilt in *Rev* than in *NoRev* (if we assume consistent beliefs).

partner without planning to cooperate, have the possibility to send a statement of intent. In order to reduce the promise keeping rate without revelation by forcing this type of subjects to either break a promise or refrain from influencing the interaction partner, we exclude the opportunity of stating an intention in a further treatment, *Com2*. However, we do not observe a significant effect of revelation in *Com2* either.

Still, this design variation provides another interesting finding. The menu of messages available to `B` seems to play a significant role for the effectiveness of communication, as `B`s are significantly more likely to keep a promise than to stick to a statement of intent. Hence, intentions seem to be less costly to break than promises. In contrast, `A`s, who are unaware of the available messages, seem to trust intentions to the same amount as promises.

Literature

This chapter is mainly related to two strands of the economic literature. First, there is an expanding literature analyzing the effect of non-binding communication on behavior. Experimental studies show that communication can increase coordination (Blume and Ortmann, 2007; Ellingsen and Östling, 2010; Kriss et al., 2011), generosity in a dictator game (Andreoni and Rao, 2011; Mohlin and Johannesson, 2008), and most relevant for our study, cooperation (Bochet and Putterman, 2009; Charness and Dufwenberg, 2006, 2010; Cooper et al., 1992; Ellingsen and Johannesson, 2004; Vanberg, 2008). So far, mainly two reasons for the effectiveness of communication have been identified. On the one hand, guilt aversion in the sense of Batigalli and Dufwenberg (2007) has been found to induce promise keeping (see for instance, Charness and Dufwenberg, 2006, or Beck et al., 2013).[4] On the other hand, individuals can exhibit a preference for promise keeping per se, that is, promises have a commitment value (Ismayilov and Potters, 2012; Vanberg, 2008). Likewise, individuals might face costs of lying (Fischbacher and Föllmi-Heusi, 2013; Hurkens and Kartik, 2009; Lundquist et al., 2009; Mazar et al., 2008). However, none of these papers consider social image concerns as a reason why people stick to their

[4] However, the effect of guilt aversion has been found to be relatively small (Ellingsen et al., 2010).

promises.

A second related strand in the behavioral economics literature studies the effect of social image concerns on behavior. Following Ariely et al. (2009), "image motivation [. . .] refers to an individual's tendency to be motivated partly by others' perceptions." That is, individuals dislike to publicly violate a social norm, such as altruism or modesty. Correspondingly, evidence for individuals behaving more selfishly or greedily if their action is less likely to be observed, has for example been found in experimental dictator games (Andreoni and Bernheim, 2009; Broberg et al 2007; Dana et al., 2006; Dana et al., 2007; Grossman, 2010a, 2010b; Koch and Normann, 2008; Larson and Capra, 2009) and in the context of volunteering (e.g. Carpenter and Myers, 2010; Linardi and McConnell, 2008) or donations (Ariely et al., 2009; DellaVigna et al., 2012; Lacetera and Macis, 2010).

Similar to our experimental design, Tadelis (2011) builds on the framework of CD (2006) and varies the ex post information of the first mover. He shows that image concerns to appear selfish (the "shame" effect) exist and increase cooperation, especially if anonymity is lifted, by announcing the second mover's action choice to all participants in the room. However, subjects in his setting are not able to communicate.

In our study, we combine these two strands of literature and investigate whether social image concerns are even more pronounced with communication, due to the aversion of being perceived as a promise breaker. Bracht and Regner (2013) also analyze social image concerns in a similar trust game with communication, however, they focus on the correlation of behavior to proneness to shame and guilt, which they elicit via a psychological test.[5] While Bracht and Regner (2013) analyze the effect of transparency and communication separately, we focus on how communication interacts with the effect of revelation on behavior. To the best of our knowledge, social image concerns have rarely been analyzed in the context of communication.

The remainder of this chapter is structured as follows. Section 3.2 introduces the experimental design and the leading hypotheses. In Section 3.3, we analyze and discuss the

[5] Bracht and Regner (2013) find that disposition to guilt predicts behavior, but not disposition to shame.

experimental results. We compare our results to previous research in Section 3.4, and Section 3.5 concludes.

3.2 Experimental Design and Hypotheses

3.2.1 Experimental Design

At the beginning of the experiment, role `A` is assigned to half of the subjects while the other half is assigned role `B`. One subject with role `A` and one subject with role `B` are randomly matched to form a pair.[6] Each pair subsequently plays the one-shot trust game depicted in Figure 3.1, which is akin to the one used by Charness and Dufwenberg (2006), henceforth CD (2006). The upper number refers to `A`'s payoff, the lower one to `B`'s.[7]

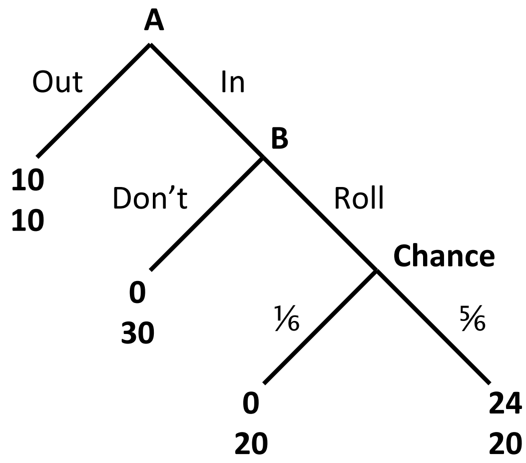
`A` ("she") decides whether to enter the game (*In*) or not (*Out*). Without learning `A`'s decision, `B` ("he") decides whether to keep a payoff of 30 tokens for himself while `A` receives nothing (*Don't Roll*), or to let a die decide over `A`'s payoff (*Roll*). In this case, `A` receives a payoff of 24 tokens with probability $\frac{5}{6}$ and a payoff of 0 with probability $\frac{1}{6}$, while `B` earns a payoff of 20 tokens in any case. In order to elicit `B`'s action choice we use the strategy method, i.e. `B` decides on his action independent of whether `A` enters the game or not. At the end of the experiment, one token is converted into 0.25 Euros.

We conduct three treatments, called *Com1*, *Com2* and *NoCom*. In *Com1* and *Com2* `B` sends one out of three predefined messages to `A`, prior to playing the trust game. *Com1* and *Com2* differ only in the type of messages that can be sent. In *Com1* `B` can choose between a promise ("I promise to choose *Roll*."), an intention ("I will choose *Roll*.") or an empty message ("Hello, how are you? I'm fine."). In *Com2* `B` can choose between the same promise and two empty messages ("Hello!" and "How are you?"), i.e. `B` cannot send an intention in *Com2*. As the design of *Com1* and *Com2* is the same except for the

[6] In the following, we refer to the player with role `A` (`B`) as `A` (`B`).

[7] In comparison to CD (2006), stakes are lower in our set-up, as one session consists of two separate experiments, which are both paid out (see Section 3.2.3). However, the proportions of the payoffs resulting from different strategies are similar.

Figure 3.1: The Trust Game

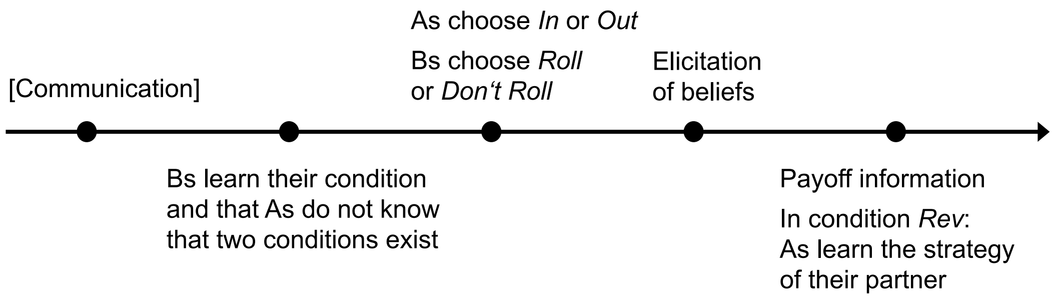


message choices, we sometimes refer to the pooling of both communication treatments as *Com*. *NoCom* is a control treatment, which is identical to the other two treatments, but without pre-play communication.

Without any further information, A cannot infer whether B has chosen *Roll* or *Don't Roll* whenever she experiences a payoff of 0. However, we are interested in the influence of social image concerns on B's cooperative behavior (see Section 3.2.2), that is, whether B cares about how he is perceived by A. Consequently, we vary whether A can observe B's action choice at the end of the experiment or not, which yields two conditions within each treatment. Before playing the trust game, half of the pairs is randomly assigned to condition "Revelation" *(Rev)*, the other half plays condition "No Revelation" *(NoRev)*. In condition *Rev* B's choice will be revealed to A at the end of the experiment, whereas A does not learn B's behavior in condition *NoRev*.

B is informed about the condition he plays before choosing between *Roll* and *Don't Roll*, but after having sent a message to A. Thereby, we ensure that only the action choice, and not the type of message sent, is affected by the condition. In other words, when B chooses the message to be sent, both conditions are exactly equal and Bs' communication behavior should not differ across conditions. Hence, any difference in *Roll* rates across conditions is then due to the variation of the observability of B's action choice and not to a difference in messages across conditions. Figure 3.2 provides an overview of the course of the experiment.

Figure 3.2: The Sequence of the Experiment



A neither learns the condition she is playing in nor is she aware that two different conditions exist until the end of the experiment. The instructions are the same for A and B and inform the participants only about the course and the payoffs of the game, without commenting on information structures.[8] B receives private information about the condition he plays via his screen during the experiment. By not informing A, we ensure that A's first-order belief about B's behavior is constant across conditions. Furthermore, B is explicitly informed about A's unawareness that two conditions exist, thus his second-order belief about A's expectations should not vary across conditions. Therefore, guilt aversion, i.e. the aversion to disappoint A's expectations cannot cause a difference in B's behavior across conditions. We explain the concept of guilt aversion in more detail in Section 3.2.2.

A's first-order and B's second-order beliefs are elicited after the trust game, but before subjects learn their payoffs. As were asked: *"What do you think, how many of the x Bs in the room have chosen Roll?"*, where x was substituted by the number of Bs in the session. For Bs, eliciting beliefs is a bit more involved. In a sequential game like the one we consider, B's choice only becomes relevant for those As who choose *In*, thus only the first-order beliefs of those As should matter for B's behavior and his second-order belief. Hence, we asked all Bs: *"We asked all As: "What do you think, how many of the x Bs in*

[8]However, the instructions emphasize that all Bs throw a die such that Bs' decisions can not be inferred, which is likely to induce a prior of getting no information among As.

the room have chosen Roll?" Consider only the `A`*s who chose In. What do you think is the average guess of those* `A`*s?"*[9] Subjects earn a supplement of 6 tokens for a guess deviating by at most +/-1 from the correct answer. This way, we elicit an interim second-order belief conditional on the event of `A` choosing *In*.[10]

3.2.2 Hypotheses

In the following, we derive our hypotheses from a notion of social image concerns and subsequently exclude other possible behavioral explanations for our hypotheses.

Assuming selfish and risk-neutral players, the unique subgame-perfect equilibrium in the trust game illustrated in Figure 3.1 is (*Out, Don't Roll*). However, while the classical game theory claims that non-binding communication cannot influence the players' strategies if information is symmetric, it has been observed in the laboratory that communication indeed enhances cooperation in trust games – promises are made, taken as credible and frequently kept. While CD (2006) argue that subjects keep their promises due to guilt aversion, that is, to not disappoint the increased expectations of the truster, Vanberg (2008) claims that people have a tendency to keep their promise per se, independent of the truster's expectations. Still, in their experiments a considerable share of trustees break a given promise.[11] We analyze whether a change in the set-up, i.e. introducing transparency about the trustee's action induces more trustees to be true on their word. More precisely, we investigate whether social image concerns of being perceived as a promise breaker exist and induce individuals to stick to their word. This yields our main hypothesis, which we break down to testable hypotheses in the following.

Main Hypothesis. *The aversion of being perceived as a promise breaker exists and is one reason for why people keep a given promise.*

[9]This procedure is analogous to the one in CD (2006).

[10]One could argue that observing the actual choice of the `A`-player is far more influential for beliefs than a hypothetical choice. However, we think that this effect is negligible given that the results show a high correlation between second-order beliefs and actual strategy choices.

[11]CD (2006) observe that 25% of promisers break their promise without revelation, Vanberg 2008 observes a share of 27% (*no switch* condition).

Indeed, it is frequently observed by Economists, Sociologists and Psychologists that people care for how they are perceived by others (e.g. Apsler, 1975; Grossman, 2010a; Lacetera and Macis, 2010; Lewis, 1995; Scheff, 1988; Smith et al., 2002; Tangney, 1995). Applied to our setting, we hypothesize that the trustee is more likely to cooperate if his action choice is revealed than if it is concealed, in a situation where communication is possible.[12]

Hypothesis 1. *The revelation of* `B`*s' action choices induces more cooperation among* `B`*s. In our setting, the Roll rate in* $[Com|Rev]$ *is higher than in* $[Com|NoRev]$.

However, the presence of social image concerns does not necessarily rely on the possibility to communicate. In fact, even without communication evidence for social image concerns has been found, such as the aversion of being perceived as egoistic or greedy (e.g. Ariely et al., 2009; Dana et al., 2006, 2007; Güth et al., 1996; Koch and Normann, 2008; Tadelis, 2011). From a theoretical point of view, Tadelis (2011) proposes a model of "shame" inducing disutility of being perceived as a non-cooperator, in order to explain the effect he observes.[13] Besides the social disapproval of egoism, we are interested in the existence of another social norm which condemns promise breaking, thereby inducing additional social image concerns. Accordingly, we hypothesize that the effect of revelation on *Roll* rates is larger if subjects can communicate than without communication, indicating that the differential effect has to be due to an aversion to be regarded as a promise breaker. Hence, we compare the results of *Com* to the control treatment *NoCom* and state the following hypothesis.[14]

Hypothesis 2. *The effect of revelation on cooperation is larger if pre-play communication takes place. In our setting, the difference between* $[Com|Rev]$ *and* $[Com|NoRev]$ *is larger than the difference between* $[NoCom|Rev]$ *and* $[NoCom|NoRev]$.

[12] We are aware that the experimenter always observes whether a promise is kept or not and that this can also evoke some social image concerns. However, the presence of the experimenter does not vary across conditions.

[13] "Guilt from blame" (Batigalli and Dufwenberg, 2007) also accounts for more cooperation in the *Rev* condition, based on `B` facing disutility from `A` blaming him for a bad outcome.

[14] We consider the *Roll* rates of all trustees in *Com* rather than focusing on those of the promising trustees only, as this allows for a comparison of *Roll* rates to the behavior in the control treatment, *NoCom*.

Yet, communication might enhance cooperative behavior of Bs independent of the observability of Bs' action choice (CD, 2006; Vanberg, 2008). In order to contribute the hypothesized higher *Roll* rate in $[Com|Rev]$ compared to $[Com|NoRev]$ to revelation only, the share of promises has to be equal in both conditions.

Hypothesis 3. *The share of promises among all messages in* $[Com|Rev]$ *is not statistically different from the share in* $[Com|NoRev]$.

Given that Bs do not know the condition they play at the pre-play communication stage, and Bs are randomly assigned to both conditions, promising behavior should not differ across conditions. Still, if and only if Hypothesis 3 holds, we can conclude our main hypothesis from Hypotheses 1 and 2.

Elimination of Alternative Explanations

In the following, we show that, given Hypothesis 1 holds, it can neither be explained by simple guilt nor by promise-keeping per se.

Simple guilt. If B is subject to *simple guilt*, in the sense of Batigalli and Dufwenberg (2007), he is reluctant to cause a lower payoff for A compared to what he believes she expects to earn. Let thus $\alpha_{\mathtt{A}} := \Pr_{\mathtt{A}}(Roll)$ denote A's belief about the probability that B cooperates. Then A expects to earn a payoff of $\frac{5}{6} \cdot \alpha_{\mathtt{A}} \cdot 24 = 20\alpha_{\mathtt{A}}$ upon entering the game. In turn, B forms a belief about A's belief about his action choice, given that A chooses *In*. This results in B's interim second-order belief $\beta_{\mathtt{B}} := E[\alpha_{\mathtt{A}}|In]$. By choosing *Don't Roll* conditional on A choosing *In*, B experiences simple guilt proportional to $20\beta_{\mathtt{B}}$, his belief about the difference between A's payoff expectation and her experienced payoff. In contrast, if B cooperates, any deception by A cannot be due to B's behavior, thus he doesn't feel guilty. Assuming that B's utility is additively separable in his material payoff and his experienced simple guilt, this yields

$$\begin{aligned} u_{\mathtt{B}}(In, Roll) &= 20 \\ u_{\mathtt{B}}(In, Don'tRoll) &= 30 - \theta^{SG} \cdot 20\beta_{\mathtt{B}}, \end{aligned}$$

where θ^{SG} denotes B's sensitivity to simple guilt.

Simple guilt can explain why communication is able to influence behavior. If B makes a promise, he believes that he influences A's belief about his behavior, i.e. $\beta_{\mathtt{B}}$ increases. Ceteris paribus, this induces a lower payoff for choosing *Don't Roll*, hence a larger share of Bs chooses *Roll* after having sent a promise.

While simple guilt delivers an explanation for why communication fosters cooperation, it cannot explain the effect in Hypothesis 1. As As do not learn the condition they play, their first-order beliefs cannot depend on whether Bs' behavior is revealed or not. Bs know about the unawareness among As and thus their second-order beliefs cannot depend on the condition either. Thus, ceteris paribus, guilt aversion predicts the same *Roll* rates for conditions *Rev* and *NoRev*. As the condition is not known to both players at the time communication takes place, the amount of promises should be the same in both conditions. Given this assumption, guilt aversion predicts the same *Roll* rates whether B's decision is revealed or not, which contradicts Hypothesis 1.

Self-image concerns ("Promise keeping per se"). Vanberg (2008) argues that there exists a preference for promise keeping per se independent of the truster's expectations. He shows that in case an individual faces a different player than the one he made a promise to, his action choice does not depend on whether the new partner has received a promise by another player before or not. In a similar vein, Ellingsen and Johannesson (2004) introduce the notion of "lying cost". They propose a model where inequity averse players suffer from a fixed personal cost of being inconsistent, $l \geq 0$, which in turn leads to a higher commitment power and credibility of promises.

However, whether B's action in the trust game is revealed to A in the end or not does not make a difference to B if he is a "promise-keeper per se". Thus, given an equal number of promises in both conditions, Hypothesis 1 can not be solely induced by promise-keeping per se.

3.2.3 Experimental Procedure

The experiment was conducted in the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA). Subjects were recruited using the online recruitment system ORSEE (Greiner, 2004), and the 406 participants in 17 sessions consisted mainly of students. Upon entering the laboratory, subjects were randomly assigned to 24 visually isolated computer terminals. The instructions were distributed and read out loud by one of the experimenters. Questions were answered individually at the subjects' seats. Before the experiment started, subjects filled out a short questionnaire ensuring the comprehension of the rules.

The experiment was the first of two independent experiments conducted in one session. Before the experiment started, participants were informed that two independent experiments would be conducted without any further information about the second experiment. Both experiments were paid out at the end of the session, where the average earning was 12.6 EUR, including a fixed show-up fee of 4 EUR. In the first experiment, As received 3.5 EUR on average, while the mean among Bs was 5.2 EUR. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007). Each session ended with a detailed questionnaire on demographics and social preferences and lasted about 50 minutes.

3.3 Experimental Results

In this section we first analyze the effect of revelation on Bs' behavior (Section 3.3.1), followed by an investigation of the effects of communication (Section 3.3.2).

3.3.1 The Effect of Revelation on Bs' Behavior

In the following, we pool the data of *Com1* and *Com2* to *Com* in order to analyze the differences between $[Com|Rev]$ and $[Com|NoRev]$. This procedure is justified as the effect of revelation on Bs' behavior does not differ between the two communication treatments. Data considering each treatment separately is gathered in the appendix.

Our first result provides some evidence for Hypothesis 1.

Result 1. *The Roll rate in* $[Com|Rev]$ *is higher than in* $[Com|NoRev]$*, with the difference being marginally significant.*

Indeed, while 63% of Bs choose *Roll* in $[Com|Rev]$, this share amounts to only 51% in $[Com|NoRev]$ (test of proportions, one-tailed, Z=1.450, p=0.074).[15] Thus, Hypothesis 1 is confirmed on a marginally significant level, indicating that subjects in a situation where communication is possible behave more cooperatively when their action is revealed than when it is not.

The next step is to take a closer look at the source of this marginally significant effect. We claim that Bs behave more cooperatively in $[Com|Rev]$ than in $[Com|NoRev]$ as they do not want to be perceived as a promise breaker. In order to confirm this claim, we have to confirm that the higher *Roll* rate in $[Com|Rev]$ is not only caused by an image concern of being perceived as selfish, but rather induced by the combination of communication and revelation (Hypothesis 2). Therefore, we compare the observed effect of revelation in *Com* to the one in *NoCom*. Figure 3.3 illustrates the shares of Bs choosing *Roll* in *Com* and *NoCom* separated by condition.

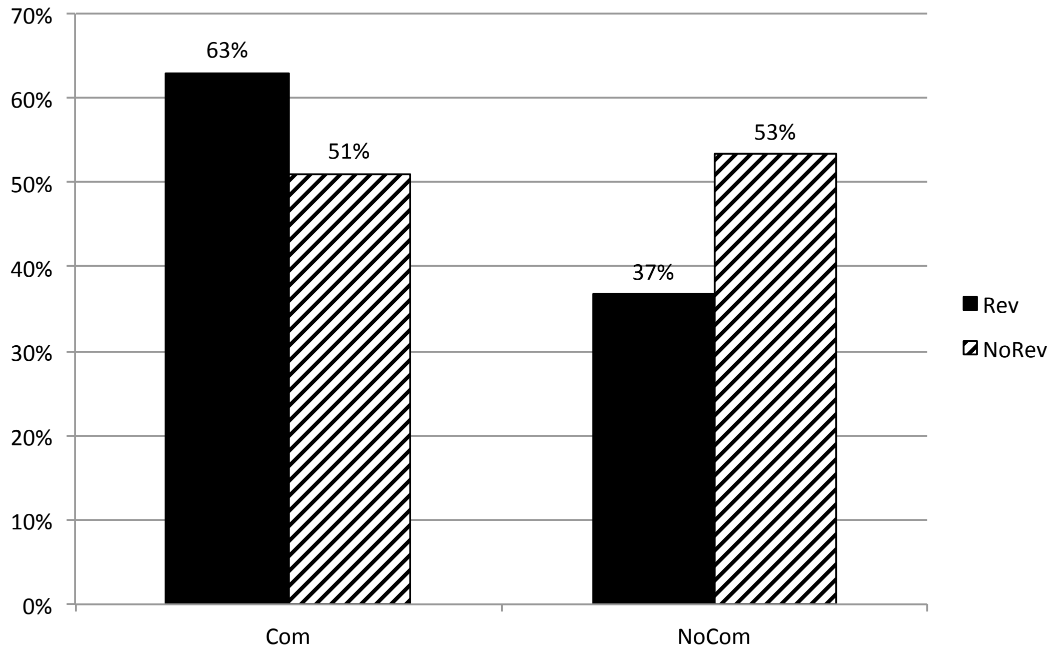
First, we consider the *NoCom* treatment separately and find no evidence for an image concern of being perceived as selfish.

Result 2. *The Roll rate in* $[NoCom|Rev]$ *is marginally significantly lower than the one in* $[NoCom|NoRev]$*. Hence, there is no evidence for the existence of image concerns of being perceived as selfish.*

Indeed, only 37% of Bs choose *Roll* in $[NoCom|Rev]$ whereas 53% cooperate in

[15] If we consider *Com1* and *Com2* separately, the effect goes in the same direction, but is no longer significant (*Com1*: 54% vs. 42%, Z=1.062, p=0.144; *Com2*: 72% vs. 61%, Z=1.000, p=0.159, one-tailed test). Throughout this chapter, the Z-Statistics reflect the test of proportions (see Glasnapp and Poggio, 1985) and p-values are on one-tailed tests, because we use our underlying hypotheses, except when reported otherwise.

Figure 3.3: *Roll* Rates of Bs Separated by Condition



[$NoCom|NoRev$]. This difference is marginally significant (test of proportions, one-tailed, Z=1.292, p=0.098). Thus, if revelation changes Bs' behavior in *NoCom*, it rather decreases cooperative behavior.[16] This rather unexpected result is unlikely to be a demand effect as Bs are only informed about their own condition, i.e. that their behavior is revealed or not revealed to A, but not about the existence of the other condition. The low *Roll* rate in [$NoCom|Rev$] might be a sullen behavior due to the sudden announcement that B's action will be revealed to A, which was not mentioned in the instructions.

While Result 1 and Result 2 already suggest the confirmation of Hypothesis 2, i.e. that the effect of revelation on cooperation is larger in *Com* than in *NoCom*, we conduct a probit regression to compare the differences across conditions in *Com* and *NoCom*, delivering the following result (Hypothesis 2).

Result 3. *The difference in Roll rates between* [$Com|Rev$] *and* [$Com|NoRev$] *is significantly larger than the one between* [$NoCom|Rev$] *and* [$NoCom|NoRev$]. *Hence, Hypothesis 2 is confirmed.*

[16] This result is in contrast to Tadelis (2011).

The results of the probit regression are reported in Table 3.1. The dependent variable is 1 if B chooses *Roll* and 0 otherwise. The independent variables are a dummy for *Com*, a dummy for *Rev*, and an interaction of the two.

Table 3.1: Regression of Choosing *Roll*

	PROBIT Coefficient (p-value)	OLS Coefficient (p-value)
Com	-0.049	-0.019
	(0.770)	(0.775)
Rev	-0.424	-0.167
	(0.105)	(0.108)
*Com**Rev*	0.731	0.287
	(0.030)	(0.040)
Constant	0.084	0.533
	(0.279)	(0.000)

We cluster standard errors on sessions (17 sessions). Number of observations is 203. In the Probit regression Pseudo R-squared is 0.023 and log Pseudo Likelihood is -136.961. In the OLS regression R-squared is 0.031.

We observe that the only significant coefficient is the one of the interaction term *Com**Rev* (p=0.030), which is positive, showing that cooperation among Bs is increased by revelation only in treatment *Com*. The negative and almost marginally significant coefficient of *Rev* indicates the negative effect of revelation on cooperation without communication. The results are robust to an OLS regression, which is also reported in Table 3.1. Thus, Hypothesis 2 is confirmed.

Yet, it remains to show that Bs' communication behavior does not differ between [*Com*|*Rev*] and [*Com*|*NoRev*] (Hypothesis 3). For no apparent reason, we are not able to confirm Hypothesis 3.

Result 4. *The share of promises among messages in* [*Com*|*Rev*] *is significantly higher than the one in* [*Com*|*NoRev*]. *Hence, Hypothesis 3 is violated and we are not able to conclude the main hypothesis about the existence of an image concern of being perceived as a promise breaker.*

Table 3.2 provides an overview of the messages sent from B to A in *Com*.

Table 3.2: Overview of Messages Sent in *Com*

		Promise	Intention	Empty
Com	*Rev*	48/71 68%	6/71 9%	17/71 24%
	NoRev	37/72 51%	13/72 18%	22/72 31%
	Z stat. (p-value)	1.975 (0.024)	-1.692 (0.045)	0.888 (0.187)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

There is neither a difference in the design nor in the instructions of the two conditions. B does not even know that two different conditions exist when sending his message. Still, we observe a significantly higher share of Bs sending a promise in [$Com|Rev$] than in [$Com|NoRev$] (68% vs. 51%, one-tailed test of proportions, Z=1.975, p=0.024).[17] On the other hand, we also observe a significantly smaller share of intentions in [$Com|Rev$] than in [$Com|NoRev$] (Z=1.692, p=0.045), yielding a similar share of intentions and promises (pooled) in both conditions (76% in [$Com|Rev$] vs. 70% in [$Com|NoRev$], p=0.448, two-tailed test). However, as further analyzed in Section 3.3.2 and reported in Table 3.3, subjects sending a promise choose *Roll* significantly more often than subjects sending an intention or an empty message (in *Rev* Z=5.568, p=0.000, in *NoRev* Z=5.183, p=0.000). Thus, we cannot pool intentions and promises, and the communication behavior has to be considered as largely different in both conditions, indicated by a significantly higher share of promises in *Rev* than in *NoRev*.

Therefore, we cannot confirm our main hypothesis via Hypotheses 1 and 2. In order to further investigate what drives the higher *Roll* rate in [$Com|Rev$] in comparison to [$Com|NoRev$], we examine the behavior of subjects sending a promise separately and compare it between conditions, thereby accounting for the different number of promises.

[17] This difference is not driven by one or two sessions, but occurs in all sessions of both communication treatments. It is only marginally significant if we consider *Com1* and *Com2* separately (see the appendix).

If the combination of revelation and communication drives the higher *Roll* rate in [$Com|Rev$], the share of promise keepers should be higher in condition [$Com|Rev$] than in [$Com|NoRev$]. As shown in Result 2, revelation itself does not lead to a higher *Roll* rate in comparison to no revelation, hence image concerns of being perceived as selfish play a negligible role in our setting. This allows us to conduct a separate analysis on the set of Bs sending a promise and attribute a difference in *Roll* rates among promising Bs across conditions to the image concern of being perceived as a promise breaker.[18]

Result 5. *The share of Bs keeping their promise among Bs who give a promise is slightly higher in* [$Com|Rev$] *than in* [$Com|NoRev$]*, however the difference is not significant.*

From Result 5 we conclude that the higher *Roll* rate in [$Com|Rev$] in comparison to [$Com|NoRev$] is mostly driven by the higher number of promises, and not by social image concerns of being perceived as a promise breaker. Table 3.3 reports the *Roll* rates for each type of message sent in both conditions.

Table 3.3: *Roll* Rates by Type of Message Sent in *Com*

		Promise	Intention	Empty
Com	*Rev*	41/48 85%	1/6 17%	3/17 18%
	NoRev	30/37 81%	4/13 31%	3/22 14%
	Z stat. (p-value)	0.534 (0.270)	-0.650 (0.258)	0.344 (0.365)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

In *NoRev* already 81% of promising Bs stick to their word, which leaves little scope for further increase by revelation. Still, in *Rev* the share is even higher with 85%. Although the effect goes in the predicted direction, the difference is not large enough to be significant (Z=0.534, p=0.270).

[18] If there was a higher *Roll* rate in [$NoCom|Rev$] than in [$NoCom|NoRev$], this analysis would not be meaningful since we cannot compare the effect of revelation among Bs sending a promise in *Com* to the overall effect in *NoCom*. Therefore, we started off with considering overall *Roll* rates in *Com*.

Result 5 is further supported by probit regressions of the decision to choose *Roll*, which are reported in Table 3.4. Here, we categorize messages into promises and no promises, where we categorize intentions as "no promise", as Bs' behavior after having sent an intention is not significantly different from the behavior after having sent an empty message (see Table 3.3 and Section 3.3.2). Column 1 of Table 3.4 reports the results of *Com*, Column 2 of *NoCom* and Column 3 (4) reports the results of a regression including both treatments with (without) controls.[19]

Table 3.4: Probit of Bs Decision to Choose *Roll*

	Coefficient (p-value)			
	Com	*NoCom*	All	All
Promise	1.742		0.808	0.715
	(0.000)		(0.000)	(0.000)
Rev	-0.059	-0.322	-0.382	-0.424
	(0.846)	(0.157)	(0.142)	(0.105)
Promise* *Rev*	0.216		0.527	0.623
	(0.561)		(0.158)	(0.085)
NoPromise			-0.639	-0.535
			(0.116)	(0.233)
NoPromise* *Rev*			-0.264	-0.347
			(0.464)	(0.346)
Risk	0.097	0.205	0.132	
	(0.064)	(0.039)	(0.003)	
Female	0.486	-0.235	0.259	
	(0.022)	(0.339)	(0.150)	
# of observations	131	60	191	191
# of sessions	11	5	16	16
Pseudo R-squared	0.314	0.106	0.240	0.212
Log Pseudo Likelihood	-61.519	-36.908	-100.433	-104.052

The regressions cluster on sessions. The reference category is *NoRev*, or [*NoCom*|*NoRev*] respectively. The sample consists of all Bs in all sessions, except of one session of *Com2*, which we exclude due to a lack of controls. Results (in column 4) do not change if we include the session. Results for *Com* and *NoCom* (columns 1 and 2) do not change when excluding the controls.

[19] Due to a lack of controls we excluded one session of *Com2*. Results for *Com* and *NoCom* do not change when excluding controls and/or including the excluded session. Results in Column 4 do not change when including this session either. Moreover, the results are robust to OLS regressions.

In all 4 regressions the dependent variable is B's decision, represented by a dummy variable which takes the value 1 if B chooses *Roll* and 0 otherwise. Promise (NoPromise) is a dummy variable for sending a (no) promise in *Com*, *Rev* is a dummy for the condition *Rev*, and Promise**Rev* (NoPromise**Rev*) is an interaction dummy of the two. In Column 1-3, we also include two controls, a measurement of risk and a female dummy.[20]

We observe that the probability to choose *Roll* is significantly higher if B sends a promise (p=0.000) than if he sends another message or does not communicate.[21] However, the coefficient of the interaction dummy Promise**Rev* is far away from being significant (p=0.561), indicating that the probability to choose *Roll* when having sent a promise is not further increased by revelation. Moreover, *Rev* does neither have a general significant effect in *Com* (p=0.846) nor in *NoCom* (p=0.157). As shown by the non-parametric test, if anything, revelation without communication even leads to less cooperative behavior as the coefficient of *Rev* in Column 2 is negative and the p-value is not far from being marginally significant.

In column 3 we report the results of the probit regression including both treatments. The reference category is a subject in [*NoCom*|*NoRev*]. In order to account for the different number of promises in the two conditions, we separate the subjects in *Com* into promisers and non-promisers and include a dummy for each group. Thus, NoPromise only takes the value 1 for Bs not promising in *Com* and it is 0 for subjects in *NoCom*. Altogether, we have 6 categories, with [*NoCom*|*NoRev*] being the base case including dummies for all other cases. Similar to *Com*, we observe that Bs sending a promise have a higher probability to choose *Roll* (p=0.000). The coefficients of *Rev* and and NoPromise**Rev* are not significant, showing that revelation does not change behavior when no promise has been sent. The coefficient of Promise**Rev* is positive, but not significant (p=0.158). Still,

[20] We elicited risk preferences through subjects' self-assessment on a scale from 0 to 10, with 0 indicating that a subject has a very weak willingness to take risks, while a score of 10 means that a subject has a strong willingness to take risks. Dohmen et al. (2011) show that this general risk question is a good predictor of actual risk-taking behavior.

[21] However, the causality is not clear. B might send a promise as he knows he will choose *Roll*, or he might choose *Roll* due to the promise sent. We will address this point in Subsection 3.3.2.

it becomes marginally significant (p=0.085) when excluding the two control variables.[22] It seems that *Rev* marginally increases the probability of choosing *Roll* conditional on sending a promise, yet, the effect is very small and not robust. Thus, we are not able to prove our main hypothesis.

3.3.2 The Effect of Communication on Cooperation

One major reason for our effects being only marginally significant might be that in [*NoCom*|*NoRev*] already 81% of all Bs sending a promise stick to it, restricting the scope for further increase in promise keeping with revelation.

We started off with conducting treatment *Com1*, where Bs choose between a promise, a statement of intent, and an empty message, and observe a very high promise keeping rate even without revelation. In order to achieve more promise breaking in the baseline without revelation, we conducted a second communication treatment, *Com2*, allowing Bs to choose only between the same promise and two empty messages. Thereby, Bs attempting to influence As while planning to take the non-cooperative decision are forced to break a promise.[23] However, this change in the set of messages failed to generate a higher rate of promise breaking in the baseline, such that we are not able to confirm our main hypothesis in *Com2* either.[24] Yet, the comparison of *Com1* and *Com2* reveals some interesting findings about the the effect of communication on cooperation and the differences between promises and intentions, which we will address in this section.

Bs' Behavior and the Choice of Messages

Considering Bs' behavior, it turns out that the set of messages available to B highly influences the effectiveness of communication on cooperation. In the following analysis, we pool the data of *Rev* and *NoRev*, as there is no significant difference in Bs' behavior

[22] Results for column 1 and 2 do not change when excluding risk and female.

[23] As are only informed that there are three messages to choose from, but that they are unaware of the type of messages or the wording. This was complete information.

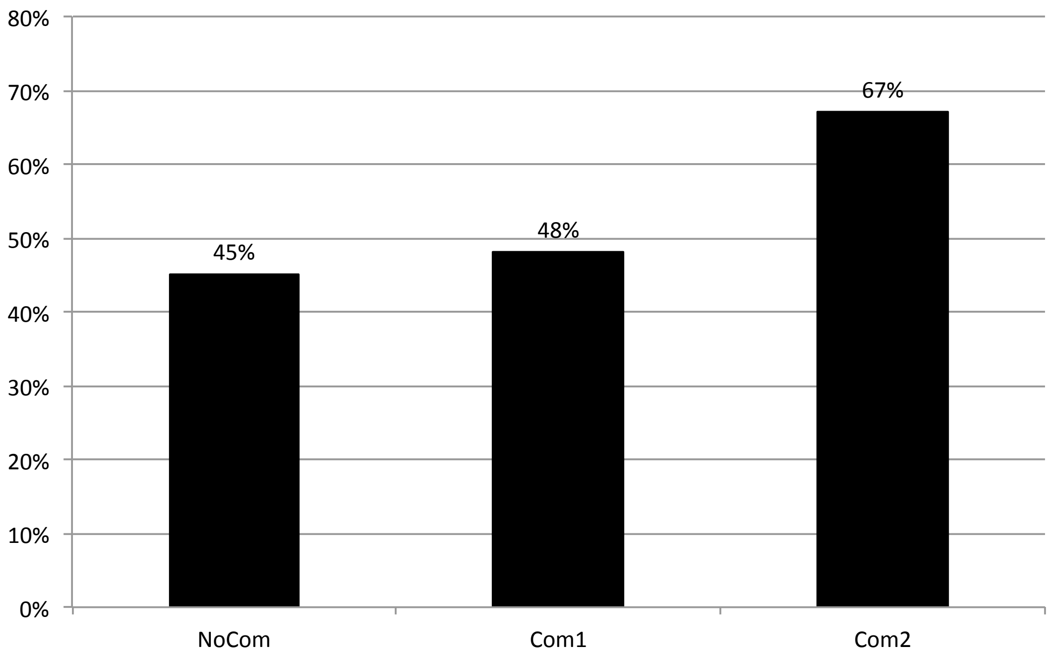
[24] Note that the unchanged communication behavior of Bs allows us to pool *Com1* and *Com2* to *Com* in the analysis of Section 3.3.1.

between both conditions.

Result 6. *While the share of* `Bs` *choosing Roll in Com2 is significantly higher than in NoCom, the share in Com1 is not.*

We conclude that the possibility to send an intention in *Com1* constrains the effectiveness of communication on cooperation. Figure 3.4 illustrates the shares of Bs choosing Roll in all 3 treatments.

Figure 3.4: *Roll* Rates of Bs Separated by Treatment



In *Com2* 67% of Bs choose *Roll*, which is significantly higher than the share of 45% in *NoCom* (Z= 2.502, p=0.006), and than the share of 48% in *Com1* (Z=2.270, p=0.012). In contrast, Bs in *Com1* are as likely to cooperate as Bs in *NoCom* (Z=0.330, p=0.371).

In order to identify the driving forces behind these effects, we analyze the data separated by types of messages. Table 3.5 reports the shares of Bs sending each of the three types of message separately and the corresponding shares of Bs choosing *Roll*.

Table 3.5: Overview of Messages Sent and Subsequent Behavior

	Messages sent			
	Promise	Intention	Empty	Total
Com1	34/71 (48%)	19/71 (27%)	18/71 (25%)	71 (100%)
Com2	51/72 (71%)	–	21/72 (29%)	72 (100%)

	Shares choosing *Roll*			
	Promise	Intention	Empty	Total
Com1	26/34 (77%)	5/19 (26%)	3/18 (17%)	34/71 (48%)
Com2	45/51 (88%)	–	3/21 (14%)	48/72 (67%)

The sample consists of all B-Persons in *Com1* and *Com2*.

Result 7. *The share of* B*s sending a promise is significantly higher in Com2 than in Com1, where roughly one quarter of* B*s choose to send an intention. While the majority of promises is kept, the majority of intentions is broken.*

While 71% of Bs send a promise in *Com2*, only 48% do so in *Com1*, with the difference being highly significant (Z=2.794, p=0.003). In *Com1* 26% of Bs send an intention, which is not possible in *Com2*. In both treatments the majority of promises are kept, in *Com1* 77%, in *Com2* even 88%. In contrast, only 26% of Bs sending an intention stick to it. This share is significantly smaller than the share of Bs keeping their promise (*Com1*: Z=3.554, p=0.000; *Com*: Z=5.083, p=0.000), but not significantly different from the share of cooperating Bs conditional on sending an empty message (*Com1*: Z=0.713, p=0.238; *Com*: Z=0.997, p=0.160).

Note that the share of promises being kept in *Com2* (88%) is even marginally significantly higher than in *Com1* (77%) (Z=1.433, p=0.076), although more Bs send a promise in *Com2* than in *Com1*. This observation, together with the fact that most intentions are broken, yields the following result.

Result 8. *Not sticking to an intention seems to be less costly than breaking a promise.*[25]

[25]This result does not follow from the mere observation that promises are kept and intentions are broken in *Com1* as this might be caused by selection into messages (altruistic subjects send a promise and selfish subjects send an intention). However, the fact that a higher share of promises is sent *and* kept in *Com2* than in *Com1* yields the result.

Result 8 can be caused by the diction of the message itself or by the comparison to message alternatives, indicating that breaking an intention is not the strongest lie. However, such a difference in behavior when sending a promise compared to sending an intention does not seem to occur in CD (2006), who use free-form messages. Therefore, we suggest that the latter reason is more likely to explain the observed phenomenon. Subjects who send an intention might not think "I am indicating to my partner that I will choose Roll", but more likely "I did not promise anything". Thus, unused alternatives seem to play a role, not as a signal to others, but as a self-justification device to behave selfish.[26] To conclude, the set of messages available to subjects in an experimental setting seems to play a crucial role for their behavior.

Finally, the fact that *Roll* rates after sending an intention are not significantly different from *Roll* rates after sending an empty message, but are significantly different from *Roll* rates after sending a promise in *Com1* speaks against the relevance of guilt aversion in our setting. `B` knows that `A` is not aware of the different available messages. Therefore, he should anticipate that both, a promise and an intention message, increase `A`'s first-order belief in comparison to receiving an empty message. Guilt aversion would predict a more cooperative behavior upon sending an intention than upon sending an empty message, and a similar behavior upon sending a promise or an intention.[27] In turn, promise keeping per se, as suggested by Vanberg (2008), is likely to play a role in our experiment, given that a high share of `B`s stick to a given promise even if their action choice is not observable.

`As`' Behavior

As `As` do not know which messages can be sent by `Bs`, we pool *Com1* and *Com2* to *Com* for the analysis of `As`' behavior. We do not differentiate between conditions either, as `As` do neither know about the existence of two conditions nor does the experimental design vary across conditions from `A`'s point of view.

[26] It would be interesting to test whether the share of people sticking to an intention, if the only options are an intention or two empty messages, is similar to the share keeping their promise in Com2.

[27] We cannot directly test for a difference in second-order beliefs as we ask for averages. We can only compare second-order beliefs across *Com1* and *Com2*. These are not significantly different (MWU, 2-sided, p=0.995).

Table 3.6 gives an overview of As' behavior in *Com* and in *NoCom*. For *Com* we report the overall behavior (total) and separated by the message received.

Table 3.6: A's Behavior

	NoCom	*Com*			
		Total	Promise	Intention	Empty
In	21 / 35%	102 / 71%	67 / 79%	14 / 74%	21 / 54%
Out	39 / 65%	41 / 29%	18 / 21%	5 / 26%	18 / 46%
	60 / 100%	143 / 100%	85 / 100%	19 / 100%	39 / 100%

The sample consists of all As.

Result 9. *The share of As choosing In is increased by communication for all kinds of messages. Furthermore, As are equally more likely to cooperate after receiving a promise or an intention than after receiving an empty message.*

We observe a large effect of communication on As' behavior. The share of As choosing *In* increases significantly from only 35% in *NoCom* to 71% in *Com* (Z=4.833, p=0.000). This effect is driven by both promises and intentions. After receiving a promise, 79% of As choose *In*, and after receiving an intention 74% do so (Z=0.488, p=0.313). These two shares are (marginally) significantly higher than the share of As choosing *In* after receiving an empty message, which amounts to 54% (empty vs. intention: Z=1.450, p=0.074; empty vs. promise: Z=2.845 p=0.002).

Interestingly, the share of As choosing *In* after receiving an empty message is significantly different from the respective share in *NoCom* (54% vs. 35%, Z=1.854, p=0.032). It seems as if As receiving an empty message might not have considered the possibility of a promise or an intention, and react to a friendly, though meaningless message.

The difference in As' behavior across treatments is reflected by their first-order beliefs about Bs' behavior. While without communication As believe that on average 45% of Bs choose *Roll*, this belief amounts to 58% with communication (MWU, 2-sided, p=0.001).[28]

[28] In particular, the average first-order belief is 63%, conditional on receiving a promise, 57% conditional on receiving an intention and 47% conditional on receiving an empty message. The first order belief is significantly higher after receiving a promise than after receiving an empty message (MWU, 2-sided, p=0.001), however not significantly higher than after receiving an intention (MWU, 2-sided, p=0.179).

Hence, similar to CD (2006), we observe that communication increases As' first-order beliefs, thus enhances trust among As.

Does Communication Enhance Mutual Cooperation?

We observe a significant increase of mutual cooperation in *Com* compared to *NoCom*, represented by the share of pairs choosing (*In, Roll*) in each treatment (45% vs. 13%, Z=4.270, p=0.000). These shares, reported by type of message sent, are stated in Table 3.7.

Table 3.7: Shares of Pairs Choosing (*In, Roll*)

NoCom	*Com*			
	Total	Promise	Intention	Empty
8/60	64/143	57/85	3/19	4/39
13.3%	44.8%	67.1%	15.8%	10.3%

Result 10. *Communication increases mutual cooperation. However, while promises increase the share of pairs choosing (In, Roll), intentions do not.*

While promises lead to a very high cooperation rate, intentions do not. This difference is mainly driven by the fact that Bs keep their promises, but break their intentions, while As trust both.[29] We conclude that the set of messages available to B plays a crucial role for the effectiveness of communication in experimental settings.

3.4 Comparison to Previous Research

The present experimental design is based on the work by Charness and Dufwenberg (2006), who analyze the effect of free-form communication on cooperation. While their design informs A only about her payoffs, we vary the revelation of B's action choice in order to test for social image concerns. However, if we restrict our data to the *NoRev* condition,

[29] This might have been different, if As had been aware of the messages available to B (compare Charness and Dufwenberg, 2010).

we find largely different results. In this section, we therefore analyze these discrepancies to the work by Charness and Dufwenberg,[30] incorporating their follow-up treatment with predefined messages (Charness and Dufwenberg, 2008 and 2010; henceforth CD, 2010). In CD (2010), Bs could choose between sending a sheet saying “I promise to choose *Roll*.” or an empty sheet, which is closest to our *Com2* treatment without revelation. Apart from the communication protocol, our design differs from CD (2006) and CD (2010) only in B’s relative payoff for choosing *Don’t Roll*, which we slightly increased in order to reduce *Roll* rates without communication (see Section 3.3.2).[31]

Considering Bs’ *Roll* rates, we do not find any difference between *NoCom* (53%) and *Com* (51%) in the *NoRev* condition (Z=0.179, p=0.429). Though slightly more Bs cooperate if we restrict the sample to *Com2* (61%), there is still no significant difference to *NoCom* (Z=0.637, p=0.262).[32] Similarly, communication fails to significantly influence Bs’ behavior in CD (2010) either. While in their experiment the average *Roll* rate increases from 44% without communication to 58% allowing for predefined messages, this difference is only marginally significant on a one-tailed test (Z=1.339, p=0.090).[33]

In contrast, Bs in CD (2006) are significantly more likely to choose *Roll* after free-form communication than without communication (44% vs. 67%, Z=2.083, p=0.019). At first glance, this indicates that Bs feel more committed to a free-form promise than to a predefined one, yielding an increase in *Roll* rates in CD (2006). However, while in CD (2006), 57% of Bs send a promise in the communication treatment, we only observe 51% in *Com* and 39% in *Com1*, the latter difference being almost marginally significant on a two-tailed test (*Com*: 51% vs. 57%, Z=0.594, p=0.552; *Com1*: 39% vs. 57%, Z=1.608, p=0.108, two-tailed tests).[34] Hence, the higher promise rate in CD (2006) might also account for

[30] More precisely, we only use the (5,5) treatment for comparison as it reflects our payoff structure.

[31] Furthermore, Charness and Dufwenberg (2006 and 2010) conduct a pen-and-pencil experiment in the classroom while we use the laboratory and computer screens. However, as we cannot identify any idiosyncratic effect of this design feature, we neglect it in the following analysis.

[32] Note that the difference was significant pooling *Rev* and *NoRev* (Section 3.3.2), but we restrict the sample to *NoRev* here.

[33] Results considering the whole sample in the communication treatment are only reported in CD (2008).

[34] Though in CD (2006), also intentions were classified as promises, we exclude intentions in *Com1* from the comparison. This is reasonable as Bs in our experiment break intentions more often than promises,

part of the increased effect on cooperation.

It is striking that, though CD (2010) find that 85% of all Bs send a promise, which differs statistically from our promise rate in *Com2* (64%, Z=2.293, p=0.022, two-tailed test), their effect of communication on cooperation is only marginally significant. Compared to our result, it seems that promises in CD (2010) induce less commitment among Bs. Indeed, while in *Com2* 87% of all Bs who send a promise keep it, this share is significantly lower in CD (2010) (61%, Z=2.183, p=0.029, two-tailed test).[35] This might be due to the fact that the messages available to Bs are common knowledge in their design, while we leave As unaware of message choices, yielding many Bs to send a promise just in order to avoid the mistrusting signal of an empty sheet.

As to As' behavior, we do not find any evidence that predefined messages in our design dampen cooperation compared to free-form communication. In fact, *In* rates among As in our experiment achieve similar levels as in CD (2006) with free-form messages (71% in *Com* vs. 74% in CD (2006), Z=0.341). Furthermore, as cooperation among As is relatively low without communication in our setting (33%),[36] we observe a highly significant effect of communication on As' behavior (71% in *Com*, Z=3.520, p=0.000), exceeding the effect with free-form messages in CD (2006) (56% without communication vs. 74% with communication, Z=1.777, p=0.038). In contrast, predefined messages in CD (2010) do not induce As to choose *In* more often, if at all, *In* rates decrease (56% without communication vs. 52% with communication, Z=0.336).[37]

While this finding seems to be unintuitive at first sight, it shows that besides differentiating between free-form and predefined messages, subtle design differences can account for huge changes in the credibility of messages. First, while Bs in our experiment choose an empty message if no promise is made (and can not refuse to send a message), the only alternative to a promise in CD (2010) is an empty sheet. It might thus be the case that empty talk in

and behave similarly after sending an intention as after sending an empty message (see Section 3.3.2).

[35] In contrast, the promise keeping rate in *Com* (81%) is similar to the one in CD (2006) (75%, Z=0.567, p=0.571, two-tailed test).

[36] Note that we restrict the sample to *NoRev* only.

[37] There is no effect despite the higher promise rate in CD (2010).

our experiment, though through predefined messages, contains some general pleasantry, thus inducing As to cooperate more often in the present setting compared to CD (2010). Second, the explicit announcement in CD (2010) that promises are not binding might create a social norm reducing both self- and social image concerns for non-cooperation among Bs, which in turn might be anticipated by As. Finally, As in our experiment are not aware of the kind of possible messages, while the exact wording and procedure is common knowledge in CD (2010). As an empty message thus signals uncooperative behavior by B in their setting and might induce As to opt out of the game, it is likely that some Bs in CD (2010) send a promise who would not have done so in other circumstances. If As anticipate this cheap-talk nature of promises, the credibility of a promise is reduced, which is why As seem to trust less in CD (2010) than in our setting. The fact that communication has a larger influence on Bs in our setting than with free-form messages in CD (2006) can only be explained by the strong wording of our predefined promise, as compared to the diverse statements of intent in CD (2006).

To summarize, Bs in our experiment as well as in CD (2010) do not seem to be influenced by communication, while in CD (2006), free-form messages increase *Roll* rates. In contrast, *In* rates in our setting highly increase with communication, with this effect being even stronger than in CD (2006), while messages do not influence As' behavior in CD (2010). Hence, starting from a slightly lower cooperation level without communication than CD (2006), we obtain a similar effect of communication on (*In, Roll*) rates, which is also highly significant (13% in *NoCom* vs. 40% in *Com*, 50% in *Com2*, $p<0.01$ in both cases, two-tailed test). In general, while messages are most influential when they are free-form, predefined messages have a larger impact in our experiment than in CD (2010). This might be due to very subtle changes in the communication protocol, such as A's unawareness of message wording or the possibility of empty talk. We conclude that the effect of communication is not robust to slight changes in the experimental design.

3.5 Conclusion

Non-binding communication is at the heart of many economic interactions, especially if cooperation cannot be contractually enforced, for example because writing fully contingent contracts is impossible or too costly, or because cooperation is not verifiable. Hence, we contribute to the literature exploring why and in which environments “cheap talk” can be influential in two-player trust games.

In this chapter we experimentally analyze whether individuals stick to their promised action, in contrast to the rational prediction, due to the aversion of being perceived as a promise breaker. While we observe slightly more cooperation of the promising party if the receiver of the promise can observe its compliance, the results are not significant. We find that 81% of subjects stick to their promise, even if their action is not observable to their interaction partners.[38] On the one hand, this result limits the scope for a further increase in cooperation with revelation. On the other hand, it highlights subjects’ preference for promise keeping per se (Vanberg, 2008), which in our experiment seems to play a more important role than social image concerns.

We find that the preference for sticking to one’s word does only exist for promises and not for statements of intent. While most of the promises are kept, statements of intent tend to be broken. In line with this result, we find that the set of available predefined messages yields different results regarding cooperation by the communicating party, the second mover. While the possibility to communicate increases cooperation by second movers if they can only choose between sending a promise or an empty message, communication has no effect on second movers’ behavior if they have the additional option of sending a statement of intent. However, the receivers of messages trust both a promise and a statement of intent in the same way. This finding allows us to exclude guilt aversion as an explanation for promise keeping, as the communicating party seems to be aware that a statement of intent does influence his partner the same way as a promise, but still does not stick to it.

[38]This even exceeds the shares reported in CD (2006) and Vanberg (2008).

To the best of our knowledge, our study belongs to one of the first economic studies analyzing the combined effect of communication and social image concerns on cooperation, suggesting a high potential and the need for further research. While we fail to prove the existence of social image concerns in our anonymous experimental set-up, one should not transfer this finding to other settings. We rather want to point out the crucial role of the design of the experiment, when trying to identify such subtle behavioral patterns. Lifting anonymity (see e.g. Tadelis, 2011) might increase the relevance of social image concerns, just like repeating the game and allowing for reputation building.

3.6 Appendix A3

A3.1 Separated Results for *Com1* and *Com2*

In all three tables the Z Stat. reflects the test of proportions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

Table A3.1: Bs' Average *Roll* Rate by Treatment and Condition

		Treatment				Z Stat.
		Com1	*Com2*	***Com***	***NoCom***	(p-value)
Condition	***Rev***	19/35 54%	26/36 72%	**45/71** **63%**	**11/30** **37%**	2.468 (0.007)
	NoRev	15/36 42%	22/36 61%	**37/72** **51%**	**16/30** **53%**	-0.179 (0.429)
Z Stat. (p-value)		1.062 (0.144)	1.000 (0.159)	1.450 (0.074)	-1.292 (0.098)	

The statistics in the last column test for the difference between *Com* and *NoCom*.

Table A3.2: Overview of Messages Sent

		Promise	Intention	Empty
Com1	*Rev*	20/35 57%	6/35 17%	9/35 26%
	NoRev	14/36 39%	13/36 36%	9/36 25%
	Z stat. (p-value)	1.540 (0.062)	-1.805 (0.036)	0.069 (0.472)
Com2	*Rev*	28/36 78%	– –	8/36 22%
	NoRev	23/36 64%	– –	13/36 36%
	Z stat. (p-value)	1.296 (0.097)	– –	1.296 (0.097)
Com	*Rev*	48/71 68%	6/71 9%	17/71 24%
	NoRev	37/72 51%	13/72 18%	22/72 31%
	Z stat. (p-value)	1.975 (0.024)	-1.692 (0.045)	0.888 (0.187)

Table A3.3: *Roll* Rates by Type of Message Sent

		Promise	Intention	Empty
Com1	*Rev*	16/20 80%	1/6 17%	2/9 22%
	NoRev	10/14 71%	4/13 31%	1/9 11%
	Z stat. (p-value)	0.580 (0.281)	-0.650 (0.258)	0.633 (0.264)
Com2	*Rev*	25/28 89%	– –	1/8 13%
	NoRev	20/23 87%	– –	2/13 15%
	Z stat. (p-value)	0.257 (0.400)	– –	-0.183 (0.427)
Com	*Rev*	41/48 85%	1/6 17%	3/17 18%
	NoRev	30/37 81%	4/13 31%	3/22 14%
	Z stat. (p-value)	0.534 (0.270)	-0.650 (0.258)	0.344 (0.365)

A3.2 Instructions (translated from German)[39]

We welcome you to this experiment. Please read these instructions carefully and follow the instructions on your screen after the start of the experiment.

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

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

[39]Passages occurring only in the communication treatments are indicated by [. . .].

questions.

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

1 Point = 25 Eurocents.

The experiment consists of two parts and a questionnaire. Part 1 will be explained below. Once all participants have finished Part 1, you will get the instructions for Part 2. A questionnaire follows after Part 2.

Instructions Part 1

At the start of the experiment, either role A or role B will be assigned randomly to each participant. You will be informed on your screen which role was assigned to you. One person A and one person B, respectively, form an interaction pair. The allocation is random and anonymous. No participant will get to know the identity of his partner during or after the experiment. Your payment in Part 1 depends on the decisions made within your interaction pair.

Decisions:

Each person A chooses between IN and OUT. If A chooses OUT, A and B get 10 points each. If person A chooses IN, the payments depend on B's decision. Every person B chooses between ROLL THE DIE and DON'T ROLL THE DIE. At the time of decision, Person B doesn't know whether A has chosen IN or OUT. But as B's decision is only relevant if A chose IN, every person B should make her decision under the assumption that A has chosen IN.

If A chose IN and B chooses DON'T ROLL THE DIE, B gets 30 points and A 0 points. If A chose IN and B chooses ROLL THE DIE, B gets 20 points and rolls a die at the end of the experiment in order to determine A's payoff. If the die shows 1, A gets 0 points, if the die shows 2,3,4,5 or 6, A gets 24 points.

The following table summarizes the payments, depending on the decisions made within an interaction pair and the result of rolling the die.

Decisions	Payoff A	Payoff B
A chooses OUT	10	10
A chooses IN, B chooses DON'T ROLL THE DIE	0	30
A chooses IN, B chooses ROLL THE DIE, Die=1	0	20
A chooses IN, B chooses ROLL THE DIE, Die=2,3,4,5,6	24	20

Please note: Every participant with role B, regardless if she chose ROLL THE DIE or DON'T ROLL THE DIE, will roll a die at the end of the experiment, such that the role of the die won't reveal the decision made by B. The result of rolling the die however is only relevant for those interaction pairs, where A chose IN and B chose ROLL THE DIE.

[**Message:**

Before A and B make their decision, B has the opportunity to choose one of three predefined messages and send it to A.]

Bonus questions:

During the experiment every participant has the opportunity to earn extra points by answering bonus questions correctly. The earnings out of these bonus questions will be displayed separately at the end of the experiment. You will get more detailed information during the experiment.

Control questions:

Before the start of the experiment control questions will appear on your screen to check that you understood the instructions. When all participants have answered these questions correctly, Part 1 of the experiment starts.

Bibliography

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

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

Altonji, J., Blank, R., 1999. Race and Gender in the Labor Market. *Handbook of Labor Economics*, Vol. 3, Eds. O. Ashenfelter and D. E. Card, 3143–3259, Amsterdam Elsevier Science.

Andreoni, J., Bernheim, B., 2009. Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5), 1607–1636.

Andreoni, J., Rao, J. M., 2011. The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism. *Journal of Public Economics*, 95, 513–520.

Apsler, R., 1975. Effects of Embarrassment on Behavior Toward Others. *Journal of Personality and Social Psychology*, 32(1), 145–153.

Ariely, D., Bracha, A., Meier, S., 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review* 99(1), 544–555.

Babcock, L., Laschever, S., 2003. *Women Don't Ask.* Princeton, NJ: Princeton University Press.

Balafoutas, L., Kerschbamer, R., Sutter, M., 2012. Distributional Preferences and Competitive Behavior. *Journal of Economic Behavior and Organization*, 83(1), 125–135.

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

Batigalli, P., Dufwenberg, M., 2007. Guilt in Games. *The American Economic Review*, 97(2), 170–176.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M., 2013. Guilt Aversion and the Impact of Promises and Money-Burning Options. *Games and Economic Behavior*, 81, 145–164.

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

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

Bertrand, M., Hallock, K. F., 2001. The Gender Gap in Top Corporate Jobs. *Industrial and Labor Relations Review*, 55, 3–21.

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

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

Blume, A., Ortmann, A., 2007. The Effects of Costless Pre-Play Communication: Experimental Evidence from Games with Pareto-Ranked Equilibria. *Journal of Economic Theory*, 132, 274–290.

Bochet, O., Putterman, L., 2009. Not Just Babble: Opening the Black Box of Communication in a Voluntary Contribution Experiment. *European Economic Review*, 53, 309–326.

Bowles, H. R., Babcock, L., Lai, L., 2007. Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes It Does Hurt to Ask. *Organizational Behavior and Human Decision Processes*, 103(1), 84–103.

Bowles, H. R., Babcock, L., McGinn, K. L., 2005. Constraints and Triggers: Situational Mechanics of Gender in Negotiation. *Journal of Personality and Social Psychology*, 89(6), 951–965.

Bracht, J., Regner, T., 2013. Moral Emotions and Partnership. *Journal of Economic Psychology*, 39, 313–326.

Broberg, T., Ellingsen, T., Johannesson, M., 2007. Is Generosity Involuntary? *Economics Letters*, 94, 32–37.

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

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

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

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

Camera, G., Casari, M., Bigoni, M., 2011. Communication, Commitment, and Deception in Social Dilemmas: Experimental Evidence. Università di Bologna, Department of Economics, Working Paper DSE 751.

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

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

Carpenter, J., Myers, C. K., 2010. Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives. *Journal of Public Economics*, 94, 911–920.

Charness, G., Dufwenberg, M., 2006. Promises and Partnership. *Econometrica*, 74(6), 1579–1601.

Charness, G., Dufwenberg, M., 2008. Broken Promises: An Experiment. UCSB, Working Paper.

Charness, G., Dufwenberg, M., 2010. Bare Promises: An Experiment. *Economics Letters*, 107, 281–283.

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

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

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

Cooper, R., DeJong, D. V., Forsythe, R., Ross, T. W., 1992. Communication in Coordination Games. *The Quarterly Journal of Economics*, 107(2), 739–771.

Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 448–474.

Dana, J., Cain, D., Dawes, R., 2006. What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games. *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.

Dana, J., Weber, R., Kuang, J., 2007. Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory*, 33(1), 67–80.

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

DellaVigna, S., List, J. A., Malmendier, U., 2012. Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, 127(1), 1–56.

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

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

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

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

Eckel, C., Grossman, P. J. , 2008. Men, Women and Risk Aversion: Experimental Evidence. *Handbook of Experimental Economics Results*, Vol. 1, Eds. C. Plott and V. Smith, 1061–1073, New York Elsevier.

Ellingsen, T., Johannesson, M., 2004. Promises, Threats and Fairness. *The Economic Journal*, 114, 397–420.

Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G., 2010. Testing Guilt Aversion. *Games and Economic Behavior*, 68, 95–107.

Ellingsen, T., Östling, R., 2010. When Does Communication Improve Coordination? *American Economic Review*, 100, 1695–1724.

Else-Quest, N. M., Higgins, A., Allison, C., Morton, L. C., 2012. Gender Differences in Self-Conscious Emotional Experience: A Meta-Analysis. *Psychological Bulletin*, 138(5), 947–981.

Eriksson, T., Poulsen, A., Villeval, M. C., 2009. Feedback and Incentives: Experimental Evidence. *Labour Economics*, 16(6), 679–688.

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

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

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in Disguise. An Experimental Study on Cheating. *Journal of the European Economic Association*, 11(3), 525–547.

Gerhart, B., Rynes, S., 1991. Determinants and Consequences of Salary Negotiations by Male and Female MBA Graduates. *Journal of Applied Psychology*, 76(2), 256–262.

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

Glasnapp, D., Poggio, J., 1985. *Essentials of Statistical Analysis for the Behavioral Sciences.* Columbus, OH: Merrill.

Goldin, C., Rouse, C., 2000. Orchestrating Impartiality: The Impact of Blind Auditions on Female Musicians. *American Economic Review*, 90(4), 715–741.

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

Grossman, Z., 2010a. Self-Signaling Versus Social-Signaling in Giving. UCSB, Working Paper.

Grossman, Z., 2010b. Strategic Ignorance and the Robustness of Social Preferences. UCSB, Working Paper.

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

Güth, W., Huck, S., Ockenfels, P., 1996. Two-Level Ultimatum Bargaining with Incomplete Information: An Experimental Study. *The Economic Journal*, 106(436), 593–604.

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

Holt, C. A., 1986. Scoring-Rule Procedures for Eliciting Subjective Probability and Utility Functions. *Bayesian Inference and Decision Techniques*, Eds. P. Goel and A. Zellner, 279–290, Amsterdam Elsevier.

Houser, D., Vetter, S., Winter, J., 2012. Fairness and Cheating. *European Economic Review*, 56, 1645–1655.

Hurkens, S., Kartik, N., 2009. Would I Lie to You? On Social Preferences and Lying Aversion. *Experimental Economics*, 12(2), 180–192.

Ismayilov, H., Potters, J. M., 2012, Promises as Commitments. CentER Discussion Paper No. 2012-064.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., Barlas, S., 1999. Overconfidence: It Depends on How, What, and Whom You Ask. *Organizational Behavior and Human Decision Processes*, 79(3), 216–247.

Koch, A. K., Normann, H.-T., 2008. Giving in Dictator Games: Regard for Others or Regard by Others? *Southern Economic Journal*, 75(1), 223–231.

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

Kriss, P. H., Blume, A., Weber, R. A., 2011. Coordination, Efficiency and Pre-Play Communication with Forgone Costly Messages. University of Zurich Department of Economics Working Paper No. 34.

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

Lacetera, N., Macis, M., 2010. Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme. *Journal of Economic Behavior and Organization*, 76, 225–237.

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

Larson, T., Capra, C. M., 2009. Exploiting Moral Wiggle Room: Illusory Preference for Fairness? A Comment. *Judgment and Decision Making*, 4(6), 467–474.

Lewis, M., 1995. Embarrassment: The Emotion of Self-Exposure and Evaluation. Eds. J. P. Tangney and K. W. Fischer, *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*, New York, NY, US: Guilford Press, xvii, 198–218.

Linardi, S., McConnell, M. A., 2008. Volunteering and Image Concerns. California Institute of Technology, Social Science Working Paper No. 1282.

Lundquist, T., Ellingsen, T., Gribbe, E., Johannesson, M., 2009. The Aversion to Lying. *Journal of Economic Behavior and Organization*, 70, 81–92.

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

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

Mazar, N., Amir, O., Ariely, D., 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6), 633–644.

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

Mohlin, E., Johannesson, M., 2008. Communication: Content or Relationship? *Journal of Economic Behavior and Organization*, 65, 409–419.

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

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

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

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

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

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

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

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

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

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

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

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

Rosenberg, M., 1965. *Society and the Adolescent Self-Image.* Princeton, NJ: Princeton University Press.

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

Sabini, J., Garvey, B., Hall, A. L., 2001. Shame and Embarrassment Revisited. *Personality and Social Psychology Bulletin*, 27(1), 104–117.

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

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

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

Savage, L. J., 1971. Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, 66(336), 783–801.

Scheff, T. J., 1988. Shame and Conformity: The Deference-Emotion System. *American Sociological Review*, 53(3), 395–406.

Smith, R. H., Webster, J. M., Parrott, W. G., Eyre, H. L., 2002. The Role of Public Exposure in Moral and Nonmoral Shame and Guilt. *Journal of Personality and Social Psychology*, 83(1), 138–159.

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

Tadelis, S., 2011. The Power of Shame and the Rationality of Trust. Mimeo.

Tangney, J. P., 1995. Shame and Guilt in Interpersonal Relationships. Eds. J. P. Tangney and K. W. Fischer, *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*, xvii, 114–139, New York: Guilford Press.

Tangney, J. P., Dearing, R. L., 2002. *Shame and Guilt.* New York: Guilford Press.

Vanberg, C., 2008. Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*, 76(6), 1467–1480.