Sandra Ludwig und Carmen Thoma:

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

Munich Discussion Paper No. 2012-15

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at http://epub.ub.uni-muenchen.de/12905/

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

Sandra Ludwig† and Carmen Thoma‡

April 18, 2012

Abstract

We analyze how subjects' self-assessment depends on whether its accuracy is observable to others. We find that women downgrade their self-assessment given observability while men do not. Women avoid the shame they may have if others observe that they overestimated themselves. Men, however, do not seem to be similarly shame-averse. This gender difference may be due to different societal expectations: While we find that men are expected to be overconfident, women are not. Shame-aversion may explain recent findings that women shy away from competition, demanding jobs and wage negotiations, as entering these situations is a statement to be confident of one's ability.

Keywords: Gender, Shame, Self-confidence, Overconfidence, Experiment
JEL-Classification: C91, D03, J16

*We would like to thank Bianca Bauer, Gary Charness, Uri Gneezy, Martin Kocher, Johannes Maier, Dalia Marin, Petra Nieken, Klaus Schmidt, Marta Serra-Garcia, Sebastian Strasser, Matthias Sutter, Lise Vesterlund, Marie-Claire Villeval, Philipp Wichardt, participants of the MELESSA seminar and the Micro Workshop at the University of Munich, the European meeting of the ESA 2011 in Luxembourg and the SFB/TR 15 conference 2011 in Tutzing for helpful comments and discussions. Financial support from LMUexcellent and SFB/TR 15 is gratefully acknowledged. For providing laboratory resources we kindly thank MELESSA of the University of Munich.

†University of Munich, Department of Economics, Ludwigstr. 28 (Rgb), 80539 Munich, Germany, e-mail: sandra.ludwig@lrz.uni-muenchen.de, Tel.: +49 (0)89 2180 3677.

‡University of Munich, Department of Economics, Ludwigstr. 28 (Rgb), 80539 Munich, Germany, e-mail: carmen.thoma@lrz.uni-muenchen.de, Tel.: +49 (0)89 2180 2926.

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 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 emotion that an individual may have when she is not as able or successful as she *publicly* announced.[4] For example, 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. Similarly, an agent might have shame if subsequent to a wage (or promotion) negotiation the employer observes that the agent is not as able as she claimed to be. 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 if 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

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

[2] For example, Balafoutas and Sutter (2012), Charness et al. (2011), Datta Gupta, Poulsen, and Villeval (2011), 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 (2002), Bowles et al. (2005), and Gerhart and Rynes (1991) show that women negotiate their wage less than men.

[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. In particular, the distinction between shame and embarrassment is strongly debated. For an overview see Sabini et al. (2001).

perform an incentivized real task. Afterwards, they are randomly assigned to be principals and agents. One principal is matched with two agents. The agents estimate the relative rank of their performance in the real 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 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. Shame-aversion may also partly explain 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*.

[5]For gender differences in self-assessment see, e.g., Balafoutas, Kerschbamer, and Sutter (forthcoming), Beyer (1990), Beyer and Bowden (1997), Möbius et al. (2011), Niederle, Segal, and Vesterlund (2010), Reuben et al. (forthcoming). It is, however, difficult to compare the size of the gender difference in self-assessment between studies and whether shame or the absence of shame drives differences between studies since experimental conditions vary, in particular, how the self-assessment is elicited.

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 negatively in society. In contrast, only about 20% state that underestimating oneself is deemed negatively in society. Yet, we observe no gender difference in these statements. But 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.[6]

Regardless of the root cause for the shame to overestimate oneself, we find that women are more shame-averse than men. Women's shame-aversion may lead to more cautious behavior and self-promotion? when their performance becomes verifiable afterwards: For example, 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. Similarly, 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.

The paper is structured as follows. In Section 2, we describe our experimental design. In Section 3, we report our basic results. In Section 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 5.

[6]Evidence from the psychological literature suggests that the society evaluates the same kind of behavior differently for men and women. Bowles, Babcok, and Lai (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.

2 Experimental Design

In our experiment, subjects are randomly assigned to the roles of principals and agents.[7] 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 task. Afterwards, the agents' assess their relative performance in the real task (explained below). To not distort effort, subjects do not know about the self-assessment when completing the real task. Participants only know that another part of the experiment will follow. Participants receive separate instructions for the real 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 real task, respectively.[8]

In the real 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 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.[9] Furthermore, the task is easy to understand

[7] 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.

[8] See the appendix for the instructions.

[9] Thus, subjects may base their self-assessment not only on their performance in this task but

and the performance is easy to measure. In addition, several other studies use this task and predominantly do not observe a gender differences in performance.[10]

Each agent receives two tokens for every arithmetic problem she solves correctly.[11] Principals receive no payment for the real task. When adding the sets of numbers, however, subjects do not know their role and thus neither whether they will be paid for the real task. 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 real task. As soon as the real task is finished, subjects learn their role.

After having completed the real task, each subject is assigned (but not told) a rank between 1 and 22, which ranks her performance in the real task relative to the performance of 21 other participants. Rank 1 refers to the best performance in the group of 22 subjects and rank 22 to the worst performance. Each subject is ranked to the same 21 other participants. These 21 other participants completed the identical real task in another session ("baseline treatment") and had exactly the same instructions and incentives for this task.[12] 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

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 (cf. Section 3).

[10]See, for example, Niederle and Vesterlund (2007), Balafoutas, Kerschbamer, and Sutter (2011), and Eriksson, Poulsen, and Villeval (2009). An exception is Niederle, Segal, and Vesterlund (2010). 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.

[11]During the experiment subjects earned tokens. At the end of the experiment, tokens were converted into Euros where 1 token=0.25 Eurocent.

[12]The baseline treatment was the first session that we conducted and we only use it to determine the performance ranking. Nevertheless, we kept it as similar as possible to the subsequent sessions, and thus also included the self-assessment part, with the difference that in the baseline treatment subjects were ranked relative to the subjects of their session.

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".[13] If an agent's guessed rank is correct, i.e. equals her actual rank, she receives 50 tokens (12.50 Euros). Whereas she receives no payment if her guessed rank differs from her actual rank.[14]

In each group, the agent whose guessed rank is better, is automatically "chosen" – irrespective of the accuracy of her guessed rank.[15] 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 principal.[16] 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 stated 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

[13]To be precise, the principals also guess their rank, yet, their guessed rank has no payoff consequences and is of no further interest for us. 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.

[14]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 Grossmann (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. Finally, we want to implement steep incentives to state one's "true" guessed rank, but under a quadratic scoring rule they are rather flat as small deviations or mistakes are cheap.

[15]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.

[16]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 prinicpal'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 (cf. an application or wage negotiation setting).

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.[17]

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.

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 agent's 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

[17]For example, Charness et al. (2011) find that men tend to increase their self-assessment for strategic reasons, while women do not. The studies by Houser et al. (2010) and Lundquist et al. (2009) indicate that women are less likely to lie.

actual performance affects the principal's expected payoff.[18]

Evidence from psychology suggests that women experience emotions more intensely than men – according to self-reports (see e.g. Brody, 1997; and Grossman and Wood, 1993). 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*.

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*.

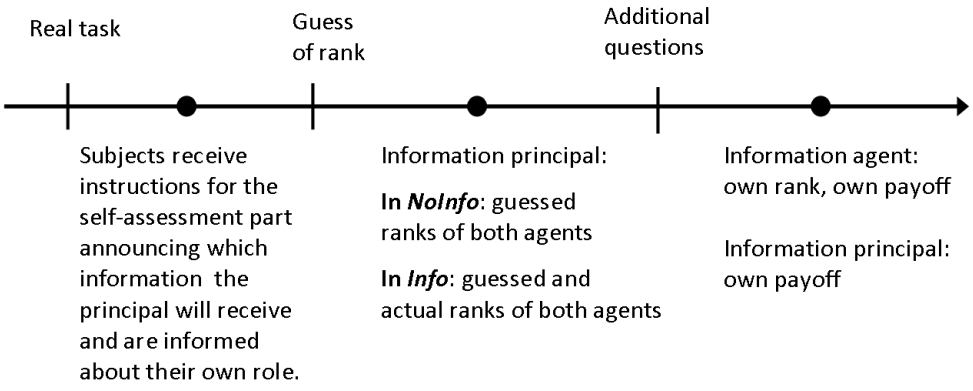
Additional questions after the elicitation of the guessed rank:
In both treatments, we elicit additional information from all subjects before informing them about the accuracy of their guessed rank. First, we ask each subject whether she thinks that her actual rank would rather be better or worse than 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

[18]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.

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.[19]

To summarize the course of the experiment, Figure 1 illustrates the order in which subjects take their actions and and receive information.

Figure 1: Course of the experiment



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 negatively 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 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

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

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.[20] Each session lasted about one hour and subjects earned 14.93 Euros on average (including a show-up fee of 4 Euros).

3 Main Experimental Results

Table 1 shows for each gender and treatment the number of observations, the average number of correctly solved problems (average number correct), the average guessed rank and the corresponding standard deviations.

Table 1: Summary Statistics

	Treatment	# observations	average number correct (s.d.)	average guessed rank (s.d.)
Women	*NoInfo*	25	16.20 (5.28)	7.80 (3.64)
	Info	24	15.87 (5.03)	10.25 (4.59)
Men	*NoInfo*	25	17.20 (5.24)	6.80 (3.28)
	Info	26	16.84 (6.81)	6.38 (4.48)

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

Neither women's nor men's performance differs between treatments: Female agents solve on average 16.20 problems correctly in *NoInfo* and 15.87 in *Info*, whereas male agents solve on average 17.2 problems correctly in *NoInfo* and 16.84 in *Info* (two-sided Mann-Whitney-U-tests (MWU-test) yield $p = 0.779$ and $p = 0.799$ for women and men, respectively). 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 (MWU-test, $p = 0.035$, two-sided): In *NoInfo*, women's average guessed rank is 7.80, while it is 10.25 in *Info*. This observation confirms our first hypothesis.

[20] 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.

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 (MWU-test, p=0.394, two-sided), which confirms our second hypothesis. If anything, the effect even goes in the opposite direction: In *NoInfo*, their average guessed rank is 6.80, while it is 6.38 in *Info*. Possibly men want to show the principal that they are very convinced of their ability by stating a better rank especially if the principal can observe the true performance.

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.

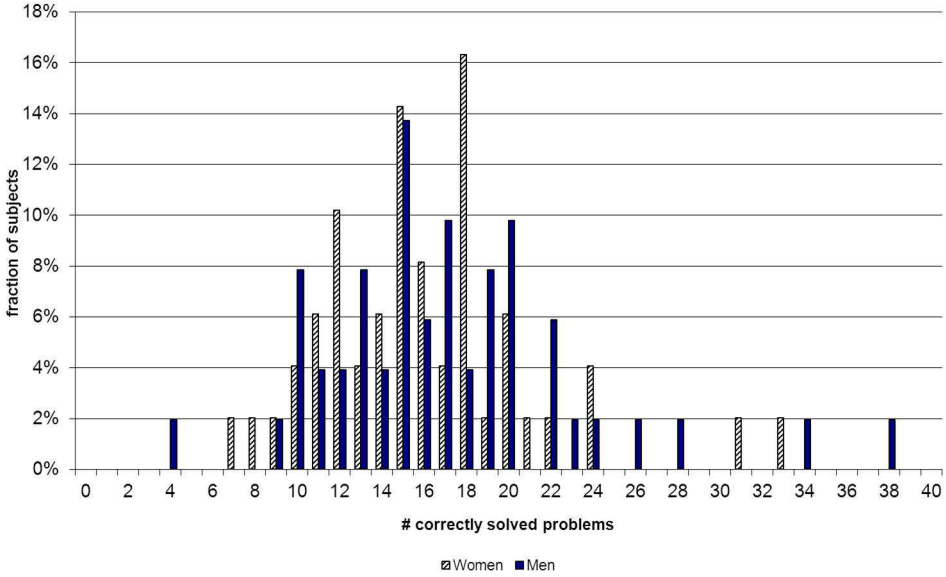
From these observations we conclude that women feel ashamed if they state a better rank than they actually have in case their actual rank is observable. To avoid this shame they downgrade their self-assessment. Men do not have this kind of shame or are at least less shame-averse than women such that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

Since only women but not men downgrade their beliefs in *Info*, shame-aversion may at least partly explain the frequently observed gender gap in self-assessment. Next, we therefore 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 (MWU-test, $p = 0.004$, 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 (cf. Table 1) does not differ significantly (two-sided MWU-tests yield $p = 0.613/0.520/0.520$ for *Info*/*NoInfo*/both treatments pooled).[21] Figure 2 illustrates this observation for the pooled sample of both treatments.[22] Thus, in treatment *Info*, we confirm earlier evidence on women stating lower beliefs about their performance than men. Yet, we cannot confirm

[21]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 (cf. Figure 2).

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

Figure 2: Distribution of correctly solved problems



the finding that women state lower beliefs in treatment *NoInfo* (cf. Table 1). In *NoInfo* the gender difference in guessed ranks is not significant (MWU-test, $p = 0.349$, 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.

The previous results are confirmed by ordered probit regressions (see Table 2), where we regress the guessed rank on performance (number of correctly solved problems), a female dummy, a dummy for treatment *Info*, and risk attitude.[23] 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, math-

[23]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. (2005) show that this general risk question is a good predictor of actual risk-taking behavior.

ematics, natural sciences), self-esteem, number of siblings, and a dummy that is one if an agent's mother is not working.[24] 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. Table 2 summarizes the regression results. The reference category is man in *NoInfo*.

Table 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.

The number of correctly solved problems (performance) is a strong predictor for the guessed rank. The result is robust over both specifications.[25] In specifications

[24]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.

[25]Note that the performance in the number adding task is neither correlated with the final

(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 and, moreover, the coefficient of the female dummy becomes insignificant. This confirms Results 1 - 3.

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.[26] 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 – 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 guessed rank and her "modal rank", which is the rank most likely assigned to her given the observed performance distribution of all participants in all sessions.[27] If an agent's guessed rank coincides with her modal rank, her accuracy is 0, whereas a negative (positive) 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 (MWU-test, $p = 0.922$, two-sided). Yet, women tend to overes-

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.

[26] See e.g., Benoit and Dubra (2011), Burks et al. (2010), Hoelzl and Rustichini (2005), Klayman et al. (1999), and Pulford and Colman (1997).

[27] To determine the modal rank for every possible performance level, we ran a Monte-Carlo simulation. 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 modal rank equals the mode of all 500,000 simulated ranks.

timate 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 (MWU-test, $p = 0.077$, two-sided).[28]

Given that many women overestimate their relative performance but guess worse 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 modal rank in *Info*. Yet, the difference is not significant (MWU-test, $p = 0.292$, 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.

4 What causes 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 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.

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 (Chi^2, $p = 0.059$, two-sided). This finding suggests that

[28]If we use the agents' actual instead of modal 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.

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 (Chi^2, $p = 0.311$, two-sided). Nevertheless, one might take these findings as a (weak) indication that men consciously state a better rank in *Info*.

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 negatively in society. 85% of all subjects say "yes". In comparison, only 21% say that underestimating oneself is deemed negatively in society.[29] These observations can explain, why people have shame when they overestimate themselves, but not if they underestimate themselves. People expect their social image to suffer, when overestimating themselves. Thus, people who care about their social 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 social 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 perceive emotions stronger than men. Hence, women might experience shame 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. 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

[29]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 negatively when it is the second question). We pool both treatments as the results are very similar.

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.[30] 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 guessed rank of women/men minus her estimated average 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 positive (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 negative (-0.25 and -2 ranks in *Info* and *NoInfo*, resp.).[31] The expectations about men and women differ significantly in the sense that women are rather expected to be underconfident than men.[32] Figure 3 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 women'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.

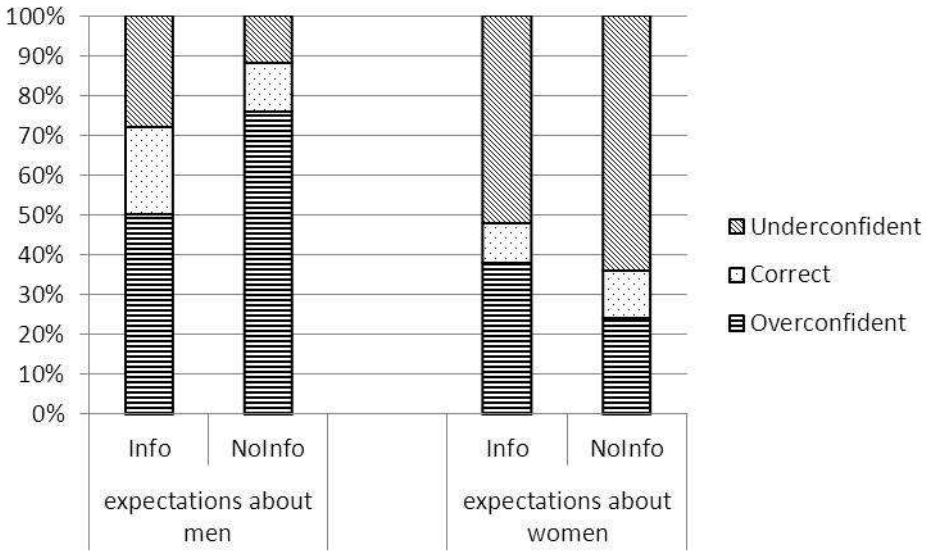
Fourth, educational differences may foster women's shame: Girls may rather be taught to be modest and reticent, while boys are taught to be self-confident and tough – which in turn may shape society's expectations. Women might have internalized these principles and may feel shame in case they do not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

[30] 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.

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

[32] 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 3: Expectations about Subjects' Self-Assessment Accuracy



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 to 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 be worse than they claimed to be. 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 if others observe them being unsuccessful. Our experiment provides first evidence of women's shame-aversion. Yet, further research is needed to pin down the impact of shame-aversion in more complex settings.

References

[1] Altonji, J., and R. Blank (1999), "Race and Gender in the Labor Market," Ashenfelter, Orley, and David E. Card (Eds.), *Handbook of Labor Economics*, Vol. 3, Elsevier: 3143–3259.

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

[3] Balafoutas, L., R. Kerschbamer, and M. Sutter (forthcoming), "Distributional Preferences and Competitive Behavior," *Journal of Economic Behavior and Organization*.

[4] Benoit, J.-P., and J. Dubra (2011), "Apparent Overconfidence," *Econometrica*, **79**(5) ,1591-1625.

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

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

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

[8] Bowles, H. R., L. Babcock, and L. Lai, (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.

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

[10] Brody, L. R. (1997), "Gender and Emotion: Beyond Stereotypes," *Journal of Social Issues*, **53**, 369–394.

[11] Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2010), "Overconfidence is a Social Signaling Bias," IZA Discussion Paper 4840.

[12] Charness, G., A. Rustichini, and J. van de Ven (2011), "Overconfidence, Self-esteem, and Strategic Deterrence," mimeo.

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

[14] Datta Gupta, N., A. Poulsen, and M. C. Villeval (2011), "Gender Matching and Competitiveness: Experimental Evidence," *Economic Inquiry*, forthcoming.

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

[16] Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2005), "Individual Risk Attitudes: New evidence from a Large, Representative, Experimentally Validated Survey," Discussion Papers of DIW Berlin No. 511, German Institute for Economic Research, Berlin.

[17] Eagly, A.H. (1987), *Sex Differences in Social Behavior: A Social-Role Interpretation,* Hillsdale, NJ: Erlbaum.

[18] Eckel, C., and P. J. Grossman (2008a), "Differences in the Economic Decisions of Men and Women: Experimental Evidence," *Handbook of Experimental Economics Results*, Volume 1, Eds. Charles Plott and Vernon Smith, 509–519. New York Elsevier.

[19] Eckel, C., and P. J. Grossman (2008b), "Men, Women and Risk Aversion: Experimental Evidence," *Handbook of Experimental Economics Results*, Volume 1, Eds. Charles Plott and Vernon Smith, 1061–73, New York Elsevier.

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

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

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

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

[24] Greiner, B. (2004), "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63, ed. Kurt Kremer and Volker Macho. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen: 79–93.

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

[26] Hoelzl, E. and A. Rustichini (2005), "Overconfident: Do You Put Your Money on It?," *Economic Journal*, **115**, 305–318.

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

[28] Houser, D., S. Vetter, and J. Winter (2010), "Fairness and Cheating," mimeo, University of Munich.

[29] Johansson Stenman, O., and K. Nordblom (2010), "Are Men Really More Overconfident than Women? A Natural Field Experiment on Exam Behavior," University of Gothenburg Working Papers in Economics 461.

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

[31] Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011), "Managing Self-Confidence: Theory and Experimental Evidence," National Bureau of Economic Research Working Paper W17014

[32] Niederle, M., C. Segal, and L. Vesterlund (2009), "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness," National Bureau of Economic Research Working Paper 13923.

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

[34] Niederle, M., and A. H. Yestrumskas (2008), "Gender Differences in Seeking Challenges: The Role of Institutions," National Bureau of Economic Research Working Paper 3922.

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

[36] Reuben, E., P. Rey-Biel, P. Sapienza, and L. Zingales, "The Emergence of Male Leadership in Competitive Environments," *Journal of Economic Behavior and Organization*, forthcoming.

[37] Rosenberg, M. (1965), *Society and the Adolescent Self-Image,* Princeton, NJ: Princeton University Press.

[38] 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.

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

Appendix

General Instructions [33]

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

[33] Original instructions were in German and are available from the authors upon request.

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 depends 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

[Information of participant B in treatment *NoInfo*:
Participant B learns:

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