SELF-CONFIDENCE AND STRATEGIC BEHAVIOR

GARY CHARNESS, ALDO RUSTICHINI, AND JEROEN VAN DE VEN

ABSTRACT. We test experimentally an explanation of over and under confidence as motivated by (perhaps unconscious) strategic concerns, and find compelling evidence supporting this hypothesis in how participants send and respond to incentivized statements of confidence. In two-player tournaments where the highest score wins, one is very likely to enter when one knows his or her stated confidence is higher than the other player's, but very unlikely when the reverse is true. Consistent with this behavior, stated confidence by males is inflated when deterrence is strategically optimal and is instead deflated by males and females when hustling (encouraging entry) is strategically optimal; this behavior is consistent with the equilibrium of the signaling game. Based on the theory of salient perturbations, we suggest that there is a strategic foundation of overconfidence. Since overconfident statements are used in familiar situations in which it is strategically effective, it may also occur in the absence of strategic benefits, provided the environment is similar.

Keywords: Self-confidence, overconfidence, salient perturbations, analogies, strategic deterrence, unconscious behavior, self-deception, hustling, experiment

JEL Classifications: A12, C91, D03, D82

1. Introduction

Belief about one's abilities is an important ingredient in many decisions, including making career choices, undertaking enterprises, and taking risks. There is considerable evidence that statements people make about their abilities often don't accurately reflect their real abilities. This phenomenon has typically manifested as overconfidence: well known studies in psychology and economics claim that people are overconfident in their ability (e.g., Svenson, 1981; Dunning, Meyerowitz, and Holzberg, 1989).[1] So an overarching question concerns the roots of such apparent overconfidence in relative ability and the corresponding benefits that might explain the persistence of the phenomenon. For example, a commonly-suggested potential personal benefit from overconfidence is the consumption value (ego utility, in Koszegi, 2006) from the belief that one is talented. In this view, people feel better with a favorable self-perception, even at the cost of being overconfident when reality intrudes.

We consider here an alternative explanation, and postulate a strategic foundation of overconfidence. Overconfident statements are often made to affect the belief of others about one's ability. This occurs in strategic situations where this effect is possible; these are common in social life. For instance, appearing more confident is likely to increase one's chances to be hired for a job or to receive a promotion, and may discourage others from competing for that same position or entering the same market. It may also elicit cooperation by others if they are in search for talented colleagues to start a joint project. In some other situations, it may pay to appear to be less confident than one actually is. Such is the case of a pool hustler, or at the workplace if this elicits help by others. We propose that the use of overconfident statements extends to a larger set of environments: Since this behavior is systematically used in familiar situations in which it is strategically effective, it may then occur even when no effect on other's behavior can be expected, provided the environment

---

[1]Moore and Healy (2008) mention a taxonomy of overconfidence, consisting of "(1) overestimation of one's actual performance, (2) overplacement of one's performance relative to others, and (3) excessive precision in one's beliefs." In this paper, we primarily consider the second of these categories, and hereafter refer to this as "overconfidence". We also note that while overconfidence is found in many studies, there is mixed evidence (see for instance Clark and Friesen, 2009) and its prevalence depends on factors such as personal experience (Weinstein, 1980) or task-difficulty (Kruger, 1999; Hoelzl and Rustichini, 2005).

in which they are made is similar to those in which such effect would be reasonable. The extension of behavior to similar environments where it may not be optimal may be facilitated if such behavior is automatic and in part unconscious. Our explanation relies upon theoretical literature on bounded rationality. The starting assumption is that cognitive limitations prevent people from calculating optimal behavior in each and every situation. Instead, it is more reasonable to assume that people make the same decisions in different situations that appear similar. This idea is presented in Myerson (1991) and further developed in a formal model in Samuelson (2001).[2] Myerson (1991) proposes that apparent suboptimal behavior can sometimes be understood by assuming that observed behavior is optimal in a related but more familiar environment, which he calls a salient perturbation.

The discipline imposed by the theory is that this environment must satisfy three conditions. First, it has to be similar to the situation the individual is really facing (has to be a *perturbation*). Second (*familiarity*), the perturbation must be more familiar to the subject than the real situation. Finally (*optimality*), the observed behavior must be optimal in the salient perturbation of the actual game. Samuelson (2001) presents a formal model showing how salient perturbations may arise as optimal planning of behavior in different environments with cognitive costs. He assumes that people plan their actions by keeping a stock of models and matching situations they encounter to the analogies they have in mind. They have to pay cognitive costs for the planning, which are higher for more complex plans. Apparent anomalies in behavior (compared to the optimal one) can be expected if situations are relatively rare but resemble a more frequently observed interaction.

Many examples can be given, for example in visual perception, in which individuals interpret what they see in terms of a salient perturbation. For example when one is sitting in a stationary train, the movement of another train is interpreted and perceived as self-motion (vection illusion (Dichgans and Brandt, 1978)). The link to Myerson's salient perturbations is apparent. The real but unfamiliar situation (you are still, the world is moving) is automatically interpreted by drawing on the more familiar situation (you are moving, the world is still) which is the salient perturbation.

---

[2]See also Gilboa and Schmeidler (1995), which provides a simple axiomatization of a decision rule that chooses a "best" act based on its past performance in similar cases.

The example of vection illustrates another property of salient perturbations. The attribution of movement to the other train is automatic, unconscious rather than a deliberate updating. Familiarity affects the degree of awareness. Behavior becomes automated and unconscious in situations that occur very frequently. By contrast, situations that look unfamiliar and have no readily available analogy are more likely to induce deliberate and reflective behavior. To illustrate, people drive automatically on a familiar route but have a heightened awareness of their environment when driving on an unknown road. A more deliberate and reflective behavior is costly in terms of the attention which has to be devoted to information processing, so the use of automatic processes can be explained, consistent with Samuelson's (2001) model, as cost reduction.

In summary we claim that biases in statements and beliefs over one's ability are important factors leading to behavior that is optimal in salient perturbations (where overconfidence is optimal) in environments in which it is not. This behavior may be in part automatic and unconscious.

Our strategy to prove the claim is to set up an experiment where the three conditions (perturbation, familiarity and optimality) are satisfied, and then show that subjects' behavior conforms to our predictions. The perturbation condition requires that we use games that are similar. This is done by introducing games that differ only in one respect (whether a signal is observed or not), but are otherwise identical. The optimality condition requires us to prove two claims. First, we have to prove that overconfidence is indeed an equilibrium behavior in the game we use. This is done by presenting a formal analysis of the equilibrium in a simple model. This analysis is necessary but not sufficient: we also have to prove that individuals in a game where the equilibrium behavior requires overconfidence are also behaving as theory predicts. We test the truth of this claim directly, observing subjects' behavior in the games. For the familiarity condition: the claim that environments where overconfidence is effective are widespread may be considered self-evident, and we do in fact claim that the statement is reasonable. But we provide an additional tests of the claim, by introducing a game where the strategically optimal behavior is instead to under-report confidence. We then prove that subjects do use underconfident behavior when such behavior is effective, as in the hustler treatment. The extension of this behavior to the case where the signal is not observed (and hustling environments are salient perturbations) is weaker than it is when

overconfidence is the optimal strategic behavior, because the hustling perturbation is much less common in real life. In this case the response to the salient perturbations is more likely to be a reflective, thoughtful, deliberate response. Let us now describe in some detail our experimental strategy and main findings.

We first use an incentive-compatible mechanism to elicit confidence in one's relative ability in a cognitive task, to detect overconfidence in a non-strategic environment. We find that the mean level of stated confidence that one is in the top half of the group is 63.4, providing evidence consistent with overconfidence.[3] We then test whether reported confidence is sensitive to social saliency and to strategic considerations. We randomly allocate participants to sender and receiver roles. We vary whether the stated confidence of a sender is shown to a paired receiver (with common information) and also vary whether the receiver is then required to compete with the sender, in a tournament in which scores on the cognition task are compared, or whether she has an available outside option. A comparison of whether or not the stated level of confidence is reported when there is no outside option (and thus no strategic relevance) shows that the awareness that another person will observe one's stated confidence has no effect on the report made.

To test whether strategic considerations affect reports in the direction of both overconfidence and underconfidence, we implemented two types of strategic treatments. In both treatments, a receiver could either enter the tournament or choose an outside option, after observing the reported confidence by the sender.[4] In the first (deterrence) strategic treatment, payoffs were such that it was in the interest of the sender to deter the receiver from entering the tournament. In the equilibrium of the game, senders over-report to appear strong and discourage receivers from entering the tournament. In the experimental data we indeed find evidence that male, but not female, senders inflate stated confidence. In the second (hustler) strategic treatment, it is in the interest of the sender in equilibrium that the receiver enters the tournament. In the data we find evidence of under-reporting by senders, both by males and by females, again consistent with the

---

[3]In an environment of incomplete information, this is not *per se* conclusive evidence of overconfidence (Benoit and Dubra, 2011; Burks, Carpenter, Goette and Rustichini, 2010).

[4]Camerer and Lovallo (1999) also study entry decisions in a tournement in the context of overconfidence. Unlike in our experiment, participants in their experiment do not observe reported confidence levels of others, and there is no strategic reason to appear under- or overconfident.

equilibrium of the game. An interesting finding in the deterrence treatment is that we also find an increase in reported confidence by male players in the role of receiver.

The hustler treatment has a double interest for our test of the theory. First, it provides a test of the idea that the confidence in the statements about one's skill follows strategic considerations, in the direction that is appropriate in the environment, not necessarily overconfidence. Second, this treatment has special interest because hustling is a less familiar situation, and so we expect that a more reflective type of behavior would be triggered as a result. Therefore, we did not expect to find that receivers in the hustler treatment would adjust their reported confidence in comparison to the baseline treatment, and the data show that indeed they do not.

Using the baseline treatment, where statements about abilities are made but no other subject could observe them, we explored the idea that overconfidence happens at an unconscious level . In this treatment, overconfidence does occur even though it can only be self-signaling. It has often been suggested that biased information processing is at the core of this. According to this view, people process information in a self-serving way, which, if true, may happen at an unconscious level. To test this, we consider an updating task where one receives a negative (but noisy) signal. We then compare behavior in an environment in which people receive feedback about own performance and in an isomorphic setting where this feedback relates to an abstract question. We find that errors in updating about own ability are far more likely than the rare errors in updating about an abstract scenario. Since the different conditions are equivalent from the point of view of a Bayesian observer, something other than Bayesian updating must be taking place.

Strategic use of confidence in statements will persist if it is effective, that is, if it really does alter receivers' behavior. In our experiment, receivers are very responsive to the reported confidence of the senders; they are highly likely to enter the tournament when their own stated confidence is higher than that of the paired sender, but highly unlikely to enter when the reverse is true. The effect on entry decisions differ across genders. Females are significantly less likely than males to enter the tournament in the deterrence treatment, despite very similar performance levels. However, this does not appear to be driven by a preference against competition (Niederle and Vesterlund, 2007) in this case, but instead by confidence level, as there is no significant difference in entry rates

when we control for confidence. In fact, the difference in confidence across gender only manifests for those people who choose to enter the tournament, with little difference in stated confidence levels for men and women who choose not to enter into the competition. In the strategic-encouragement treatment, both the stated confidence levels and entry rates for males and females were almost the same, again not providing evidence that women shy away from competition.

The remainder of this paper is structured as follows. In section 2, we provide a review of the literature. We describe our hypotheses and our experimental design in section 3. We present our experimental results in section 4, and we discuss the motivation of biased confidence in section 5. We conclude in section 6.

2. Background and literature review

The ideas of salient perturbations and analogies are consistent with some earlier experimental findings. Framing effects can be understood by assuming that different descriptions of a task trigger different analogies. A prisoner's dilemma framed as the "Community Game" elicits more cooperative behavior than if framed as the "Wall St. Game" because the label "Wall Street" is associated with more competitive behavior (see Liberman et al., 2004). Likewise, cooperation with an anonymous stranger in a one-shot game can be explained with predictions from the theory of repeated games if the repeated game is the salient perturbation of the one-shot game.

Social psychology has long considered the issues of self-esteem, overconfidence, and self-deception: for example Baumeister (1998) provides an extensive review of the overconfidence phenomenon; and further evidence and discussion on the topic of self-esteem can be found in Leary, Tambor, Terdal, and Downs (1995) and Leary (1999), where image concerns lead to a selective demand for information. Berglas and Jones (1978) and Kolditz and Arkin (1982) also study how self-handicapping is related to social saliency: Kolditz and Arkin (1982) find that subjects take performance-impoverishing drugs after receiving positive feedback about their past performance when their choice of drugs is visible to the experimenter. However, when subjects choose whether or not to take the performance-impoverishing drugs in private, no subjects take them. This suggests that performance/confidence is a social signal.

7

Rabin and Schrag (1999) provide a model of confirmatory bias, where people misinterpret new information as supporting previously held views; in this model a confirmatory bias induces overconfidence. An agent may come to believe with near certainty in a false hypothesis, even though he or she receives an unlimited amount of information. Koszegi (2006) provides a formal economic model of overconfidence and ego utility, in which an agent derives internal benefits from positive views about his or her ability. The mechanism in this model is that each person receives an initial signal about own ability and can seek information if desired.

A number of recent papers examine overconfidence. The focus is typically on establishing overconfidence without considering the strategic value of appearing overconfident or underconfident and the response to confidence statements by others. In some recent experiments, participants receive information about the stated confidence of others (e.g., Vialle et al., 2011; Ewers, 2012), but they do not study if participants use this strategically: confidence levels were elicited before participants were told that their reported confidence would be shared with others.

Other studies directly investigate biased information processing with regard to own skills. Mobius, Niederle, Niehaus, and Rosenblat (2011) study how subjects respond to noisy feedback about their performance in an IQ test; people receive signals about their relative rank that are correct with 75% probability. They find that subjects do not update sufficiently and react more to positive feedback than to negative feedback); women are more conservative than men, but not more asymmetric. High-confidence women have a higher value for information than men. They do not compare updating to a neutral case where self-image concerns are absent.

Ertac (2011) gives feedback to subjects in two different environments.[5] In one case, feedback is non-own-performance related, while in the other, feedback is about own performance in an algebra and verbal test. She finds no systematic bias in updating in the neutral context, but systematic updating errors with the algebra/verbal test. The systematic mistakes tend to go against self-serving beliefs, as here people are more affected by bad news than by good, resulting in pessimistic beliefs. A difference from our study pertains to the type of feedback that is given. In her case,

---

[5]Our study was conducted independently, as we were unaware of this then-unpublished work when we designed our experiments.

feedback is always correct but incomplete (e.g., people are told they are not in the top 20%, but not if they are in the middle 60% or bottom 20%). Instead, we provide noisy feedback, which is not always correct. This makes it possible for participants to attribute any negative feedback to external causes, i.e., to 'bad luck' (noise).[6]

Eil and Rao (2011) study how people process and acquire objective information regarding intelligence and beauty. They rank 10 people in a group, and elicit the complete belief distribution over all possible ranks; participants receive incomplete but truthful feedback. Each person's rank was compared to an anonymous participant, with feedback about this comparison. People are found to respond much more to positive feedback than to negative feedback. Even in a non-own-performance control treatment, updating and information acquisition were unbiased. While one gets more precise information with their set-up, it is much more complicated for participants; updating seems to be quite noisy and errors appear to be common even in the control treatment.[7] The fact that updating is unbiased in the neutral task may be due to the fact all subjects start from the same symmetric (uniform) distribution in that case: Each rank is equally likely at the start. The advantage of our simple set up is that participants make very few errors on the neutral abstract task. We also ensured that the distributions of the priors are similar in both tasks, so that differences in updating cannot be due to different prior distributions.

Burks, Carpenter, Goette and Rustichini (2012), based on data in Burks et al. (2009), investigate whether concerns for self-image contribute to overconfidence and whether confidence judgments are consistent with Bayesian information processing starting from a common prior. They reject both hypotheses. Their results indicate that individuals with higher beliefs about their skills are more likely to demand information, rather than less likely. These results clearly reject self-image concerns as a mechanism that yields overconfident judgments, and are consistent with the hypothesis that overconfidence is a form of social signaling. In their experimental design there is no strategic

---

[6]Studies in psychology suggest that attributing success to one's self and attributing failures to external sources (such as 'noise') is a commonly employed strategy by people (see, e.g., Wolosin, Sherman, and Till, 1973).
[7]There is also a possible alternative interpretation of their results, in that people with high confidence on the intelligence tests are smart, and better at updating. Since they are more likely to get positive feedback, the results could reflect the fact that smarter people are better at updating. However, since they find similar results in the case of beauty, this alternative interpretation can only be correct if better-looking people are better at updating.

environment that can affect confidence, so a direct test of the hypothesis is difficult. In this paper we introduce the strategic environment explicitly, and study the strategic motivation underlying such signaling.

Grossman and Owens (2010) study how one's beliefs about own performance (on a quiz) are affected by noisy, but unbiased feedback. In the main treatment, participants overestimate their own scores, believing that they have received unlucky feedback. However, this is driven not by biased information processing, but rather by overconfident priors. In a control treatment, each participant expresses beliefs about another participant's performance, with (on average) accurate posteriors. Even though feedback improves estimates about performance, this does not lead to improved estimates of related performances. This result suggests that how people use performance feedback to update beliefs about own ability differs from how they update their beliefs about own performance, which may relate to the issue of why overconfidence persists.

3. Hypotheses and Experimental Design

3.1. **Model.** The key tool we use to test our hypotheses is a tournament game where players can send explicit statements on their ability. A simple model represents this game. There are two players. One of them (the sender, or player 1) sends a message about her ability. The other player (the receiver, player 2) then decides whether to enter a tournament with the sender. Players can have different 'types,' reflecting different abilities. The type of player i, $\theta_i \in \Theta^i$, is chosen according to some probability distribution, and is private information to the player. To simplify the exposition we assume that the set of types has only two elements, $\Theta^i \equiv \{\theta_0^i, \theta_1^i\}$ with θ_1 of better quality than θ_0 and that the prior of players is that both types are equally likely to occur. The sender moves first, and makes a claim about her ability by sending a message $t \in T = \Theta^1$. The message need not be truthful, but sending a false message costs $c > 0$. After observing the message, the receiver can choose an action from the set $\{In, Out\}$. If the receiver chooses Out, both players receive their outside option O^i. If the receiver chooses In, the two players compete in a tournament, and their payoffs are determined by their abilities and given by:

(1)

	θ_0^2	θ_1^2
θ_0^1	$0,0$	b,a
θ_1^1	a,b	d,d

We focus on the case for which a player is better off if the opponent is weaker ($b \leq 0$, $d \leq a$) and if she herself is stronger ($a \geq 0, d \geq b$). To make things interesting we assume $a \geq O^2 \geq d$, so that a strong receiver prefers playing the tournament to the outside option if he knows that the sender is a weak type, but prefers the outside option if he knows that the sender is a strong type. We also assume that $d \geq 0$, implying that a weak receiver always prefers to stay out.

This game reflects situations in which people can strategically manipulate how confident they appear to others. Under the assumptions made, a weak receiver will always opt out of the tournament, but for a strong receiver this choice will depend on his beliefs about the sender's type. The best strategy for the sender depends crucially on her outside option. If her outside option is high, she is better off if she does not have to compete with the receiver in the tournament. The sender can achieve this by appearing strong, i.e., over-report. On the other hand, if her outside option is low, she prefers that the receiver competes with her. In this case, the sender can achieve this by appearing weak, i.e., under-report. Indeed, both over- and under-reporting may occur in equilibrium (see Appendix B for details).[8]

In the experiment, we implemented two conditions. In the deterrence treatment, parameters are such that senders *over-report* in equilibrium (i.e., claim to be a weakly higher type than they really are). In the hustler treatment, senders under-report in equilibrium. While our experimental design accommodates both under and over-reporting, we argue that the hustling environment is relatively rare. In an adaptation of the game in which the payoff of players is higher if the opponent is a strong type, only equilibria with over-reporting exist for reasonable parameters. Such *mating* games are very common in nature so that overconfidence can be expected to be advantageous more

---

[8]Kartik (2009) also analyzes a sender-receiver game with lying costs. The game he analyzes has a different setup, so we cannot directly apply his results. He shows that in his setup senders almost always claim to be more confident than they really are, but the payoff structure differs from ours.

frequently than underconfidence. An example at the workplace that has the structure of a mating game is when an employee wants to convince co-workers that he or she is talented, so that he or she will be chosen to collaborate on a joint project.

3.2. **Hypotheses.** We now state our theoretical hypotheses. In light of the evidence that males and females respond differently to competitive environments (see e.g., Gneezy and Rustichini, 2004; Gneezy et al., 2003; Niederle and Vesterlund, 2007), we also consider the possibility that confidence display differs across genders. Based on these principles, we formulate the specific hypotheses below.

H1: Statements of confidence are typically social signals of intentions or private information, and individuals take them into account when they observe the self-evaluations of others. This is a first-order awareness of the social implications of self-confidence. Individuals may also anticipate this effect and adjust this signal accordingly, a second-order awareness.

Thus in our experiment stated confidence levels will be affected in the direction predicted by the equilibrium in strategic environments, where the setting is explicit and the advantages are real and clear.

H2: In our experiment, higher stated confidence levels will tend to discourage potential competitors from entering the tournament. Similarly, lower stated levels will encourage competitors to enter. So accurate reporting of confidence levels is not an optimal strategy, even taking into account the incentives for accurate reporting.

Thus both overconfidence and underconfidence can be motivated by strategic considerations, and subjects behave according to the equilibrium predictions in both types of environment.

H3: Overconfidence in statements that are useful in familiar competitive environments will extend to less familiar situations, and overconfidence is likely to occur even when no one else is watching.

Thus we predict overconfidence even when one's confidence level is unknown to other participants, in spite of the incentives provided for stating beliefs truthfully. This extension will be smaller or

even absent for stated underconfidence, because environments where stating underconfidence is optimal are less widely experienced.

H4: Males will exhibit higher stated confidence levels and are more likely to enter the tournament than females, even after controlling for performance.

Thus, in line with previous research, we also predict that males are more likely to enter the tournament than females, controlling for confidence.

3.3. **Experimental design.** Sessions were conducted in Amsterdam with 16 to 28 participants depending on the number of subjects showing up for the experimental session. Instructions were displayed on a computer screen and read aloud. Participants were told that their decisions would remain anonymous to the other people present unless explicitly indicated otherwise, and that they would receive their earnings in an envelope from a person in a different room who could only see login numbers and could not match these numbers to names or faces. Participants were paid for one task chosen at random.

We ran a total of 22 sessions with a total of 464 subjects; seven of Treatment 1 ($N = 144$), three of Treatment 2 ($N = 68$), seven of Treatment 3 (three with low outside option, $N = 60$, four with high, $N = 96$), and five of Treatment 4 ($N = 96$). Sessions lasted for 40 to 50 minutes, with an average payment of €14 (of which €7 was a show-up fee). Sessions ended with a questionnaire. Almost all participants (96 percent) were undergraduate students (average age 22 years, standard deviation 2.96; see Table 2 for details), with the majority studying economics or business; 44 percent of these subjects were female.

In every treatment, participants were randomly allocated to groups of four individuals. In each group, two players were randomly given the role of senders and the other two the role of receiver (in the instructions we always used neutral labels "A" and "B" for the two roles); each sender was randomly matched with one receiver. All participants received the same 15 questions taken from Raven's Advanced Progressive Matrices (APM), a measure of cognitive ability (Raven, 2000). Participants had eight minutes to answer as many questions as they could, and did not get any feedback after completion on the number of questions they answered correctly. The experimental

instructions can be found in Appendix A. Payments were presented in points: One point was worth one euro. In the period in which the experiment was run €1 was worth approximately between \$1.30 and \$1.40. In the exposition below we translate points directly into euro, although the instructions were strictly in terms of points.

When taking the APM test, participants only knew that they would be asked to evaluate their performance later and that every sender would be matched to a receiver with a possibility for the player with the higher rank to earn 10 points, that is, €10. Upon completion, participants were informed about all the subsequent steps in the experiment. First, one was asked to indicate one's confidence of having a score in the top two of their group, on a probability scale from 0 percent to 100 percent. They received payment for accuracy according to a quadratic scoring rule; for a stated probability p (their report divided by 100), a subject was paid €10 times $1 - (1 - p)^2$ if he really was in the top 2, and €10 times $1 - p^2$ if he was not. As can be seen in the instructions, we provided assurances that this mechanism favored accurate reporting for this part of the experiment.

Table 1 gives an overview of the different treatments. In the baseline treatment, no one could see the confidence of another player, each receiver could observe the reported confidence by the paired sender in the other treatments. In all treatments there was a possible tournament between the paired sender (S) and receiver (R). In Treatments 1-3, the player with higher rank received €10 and the other received nothing. Entry by both players was mandatory in Treatments 1 and 2, but each R faced a strategic decision in Treatments 3 and 4: After observing S's reported confidence, R chose whether or not to enter a tournament. In the low-outside-option version of Treatment 3, R received €3.5 by staying out, while in the high-outside-option version of Treatment 3, R received €5.5 for doing so.[9] In these treatments, S preferred that R opted out of the tournament since that would secure €10. In Treatment 4, if R chose not to enter, R received €5.5 and S received €10. If R chose to enter and won, then R received €10 and S received €15, while if R entered and S won, then R received 0 and S received €25; thus, S preferred that R enter the tournament. In

---

[9]We initially used an outside option of €3.5, but found that 28 of 30 receivers entered the tournament. We then switched to an outside option of €5.5.

14

Treatments 3 and 4 participants must trade off honest reporting against trying to influence the opponent's entry decision.

The description we have just given, including whether or not any player could see the reported confidence of others, or whether player R was given a choice between playing in or out, was common information and known to all subjects before they reported their confidence. They were also told, in all treatments, that they would find out at the end of the game who had the higher rank between the two matched S and R players, but would learn neither their rank in the group of four nor the number of questions answered correctly.

Table 1: *Overview of treatments*

Treatment	Receiver observes Sender's reported confidence?	Payoffs if receiver opts out of tournament (S, R)	Payoffs if receiver enters tournament (S, R)	
			Sender wins	Receiver wins
1: Baseline	No	N/A	(10,0)	(0,10)
2: Social	Yes	N/A	(10,0)	(0,10)
3a. Deterrence (low)	Yes	(10, 3.5)	(10,0)	(0,10)
3b. Deterrence (high)	Yes	(10, 5.5)	(10,0)	(0,10)
4. Hustler	Yes	(10, 5.5)	(25,0)	(15,10)

Notes: S stands for Sender, R for Receiver.

Treatment 1 had some additional components, the Reports and the Machine questions:

Reports. First, after reporting their confidence, participants were sent a report telling them if they were among the top 2 of their group or not. They were told that the report was always correct when one was in the bottom 2, but that the report was incorrect in 50% of the cases when one was in the top 2. There was no deception: Reports to participants were determined as stated. After receiving the report, they were asked if they thought they were in the top 2 or not. They received €10 for a correct assessment.

Machine questions. Subsequently subjects were given an abstract scenario. Two machines in a production hall produce rings: the left machine produces 50% good rings and 50% bad rings, the right one produces 100% good rings. A mechanic inspects one of the machines every day by taking a ring. Participants were told the percentage of days the mechanic went to the left machine and were then asked if, given that the mechanic took a good ring, it was more likely that the mechanic went to the left machine or right machine. This question was asked three times, with various percentages of days that the mechanic went to the left machine. For the first two questions, the percentage was randomly drawn from a uniform distribution between 45 and 85. For the third question we used reported confidence levels by participants as an input. Participants were not told how the percentages were selected. One of the three questions was randomly selected, and for that question they received €10 for a correct assessment.

The reports and the machine questions have an identical statistical structure (see Figure B1 of Appendix B). In both, there are two possible states of nature, ω_1 and ω_2. The probability of state ω_1 occurring is p. One of two signals was sent, s_1 or s_2. If $\omega = \omega_1$, each signal was sent with probability .5; if $\omega = \omega_2$, signal s_2 was always sent. Since s_1 only occurs when state is ω_1, upon observing s_1, state ω_1 must have occurred. Upon observing s_2, ω_2 has posterior probability larger than ω_1 if and only if $p < 2/3$.[10] In the report part, the states are $\omega_1 = $ 'in top 2', $\omega_2 = $ 'not in top 2', p is the reported confidence, and $s_1 = $ 'in top 2', $s_2 = $ 'not in top 2'. In the machine part, $\omega_1 = $ 'left machine', $\omega_2 = $ 'right machine', p is given in the question, and $s_1 = $ 'bad ring', $s_2 = $ 'good ring'.

4. Experimental Results

4.1. **Confidence.** Summary statistics are reported in Table 2. The distribution of correct answers (out of 15) is approximately normal, with mean 8.75 (8.78 for males and 8.71 for females). No more than 27 percent report a confidence level below 50 percent in any of our conditions; note

---

[10]If a participant received the report 'not in top 2', the likelihood that the report was incorrect (i.e., the person is in the top 2) is $.5p/(.5p + 1 - p)$ and the likelihood that the report was correct (i.e., subjects is not in top 2) is $(1 - p)/(.5p + 1 - p)$. Conditional upon receiving the report 'not in top 2', incorrect (top 2) is thus more likely than correct if and only if $.5p > 1 - p$, that is $p > 2/3$. Similar reasoning shows that upon receiving the information that the ring is good, it is more likely that the mechanic went to the left machine than to the right machine if and only if $p > 2/3$, where p in this case is the probability that the mechanics went to the left machine.

however that the rate in the hustler treatment (27 percent) was nearly double that in the deterrence treatment (14 percent). In data pooled over the conditions, 71 percent of the people report a confidence level above 50 percent and only 20 percent report a confidence level below 50 percent; a binomial test finds this asymmetry to be highly significant ($Z = 17.00$, $p = 0.000$).

Table 2: *Summary statistics*

	Mean	Std. Error	Min.	Max.
Test				
Number correct answers	8.75	0.11	1	15
Confidence	63.75	1.02	0	100
Background characteristics				
Age	21.96	0.14	17	49
Number of siblings	1.48	0.05	0	7
Gender (fraction females)	0.44			
Member of sports club	0.49			
Took Raven test before	0.54			
Familiar with condition probs.	0.61			
Study category				
Economics/Business/Finance	0.58			
Social Sciences and Law	0.15			
Physics, Math, Computer science	0.07			
Other study or not student	0.20			
N	464			

Figure 1 shows a treatment and gender effect with respect to mean reported confidence. Men are considerably more confident in the strategic deterrence conditions compared to the involuntary-entry and hustler treatments, while no difference is found for women. There is no difference in the stated confidence level of men and women in the hustler treatment.

We did not expect to find a difference in stated confidence between senders and receivers in the baseline treatment, since their roles do not differ in that treatment, and indeed we do not find any: the confidence of males is 65 in both roles, and that of female senders and receivers is respectively

FIGURE 1. **Cumulative distribution functions of confidence, by gender and treatment.** Roles pooled. Error bars: +/- SE
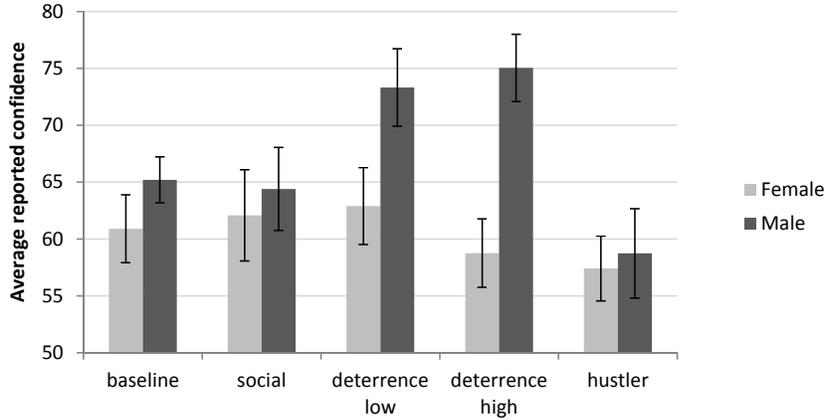
60 and 62. We therefore pool these observations by gender. Compared to the baseline treatment, male senders in the social treatment only report 3 percent points higher confidence, a difference that is not significant (WMW, $Z = -0.789$, $p = 0.430$, two tailed test). They do however report significantly higher confidence in the deterrence treatments (73 percent, low and high outside option combined, $Z = -2.342$, $p = 0.019$) and significantly lower confidence in the hustler treatment (53 percent, $Z = 2.007$, $p = 0.045$). Male receivers also report significantly higher confidence in the deterrence treatments (76), while their reported confidence in the hustler treatment is comparable to that of males in the baseline treatment (65 percent, $Z = -2.949$, $p = 0.003$). The reported confidence of females is always between 60 and 63 (in both roles) and not statistically different from that in the baseline treatment in any of the treatments, except for female senders in the hustler treatment who report a significantly lower confidence of 51 ($Z = 1.779$, $p = 0.075$)

OLS estimates of the determinants of confidence are presented in Table 3; the baseline condition reflects male behavior in the baseline treatment (Treatment 1). Specification (1) shows that the number of correct answers is a strong predictor of confidence, adding about 3 percentage points for each correct answer; this result is robust over different specifications.

Since subjects were not told their number of correct answers, the effect of correct answers on stated confidence can only be based on an estimate of one's own relative ability. In addition, we find significantly lower stated confidence in the hustler treatment, although the effect of the strategic-deterrence treatment is not quite significant. Specification (2) adds controls for the role of the participant (sender or receiver) and interaction terms for the treatment and role. Being a Sender has no effect by itself, nor is there a significant interaction effect with the Social and Deterrence treatments. However, there is a large interaction effect in the Hustler treatment, indicating that the decrease in stated confidence is entirely due to senders; in fact, the coefficient on Hustler is now actually positive, although not significant.

[Table 3 about here]

We introduce a dummy for gender and interaction effects for gender and treatment in specification (3). The results are consistent with the picture of the nonparametric tests. We find a negative but insignificant direct gender effect. However, there is a significant treatment effect: Reported confidence increases by almost 10 percentage points in the Deterrence treatment. This effect is only present for males, as the coefficient of the interaction between Deterrence and Female shows a negative coefficient of about the same size as the treatment coefficient. On the other hand, there is no such difference by gender in the Hustler treatment, indicating that both male and female sender's deflate stated confidence. There is no significant difference for sender or with the interaction of either Deterrence or Hustler and Sender, and none for the three-way interactions. Finally, specification (4) shows that people who indicated they were familiar with conditional probabilities are more confident, by more than four percentage points. This familiarity has a

significant effect even accounting for the difference in the number of correct answers (9.10 with familiarity versus 8.20 without it).

Receivers in the deterrence treatment inflate their confidence levels to about the same degree as the senders (the coefficient for the interaction variable Deterrence*Sender is small and insignificant). The receiver's inflated levels of stated confidence cannot deter senders from entry, so this cannot reflect deliberate cognitive planning. It may instead reflect unconscious motivations generated by the competitive setting, so that people may not be flexible enough to adjust their behavior to their contingent role in the deterrence environment. Their behavior can be explained by assuming that the salient perturbation of the game is the game in which both players have a strategic value of deterring the other players.

Results (Confidence)

(1) *The real performance of participants, measured by the (unknown to the participants) number of correct answers, significantly influences reported confidence in the expected directions. Those people who are familiar with conditional probabilities also report higher confidence, after controlling for correct answers.*

(2) *Men report a significant 10 percentage points higher confidence in the deterrence treatment, even though it is only known after taking the test (but before the statement is given) that there will be strategic interaction. There is no significant treatment effect for women.*

(3) *There is also a significant treatment effect in the hustler treatment, as both male and female senders deflate their stated confidence by about 15 percentage points. However, receivers do not deflate stated confidence at all.*

(4) *The only case in which there is a difference in stated confidence between Senders and Receivers is the hustler treatment, where there is a difference for both males and females. In the deterrence treatment, the similarity of the behavior of players in the two roles may reflect an automatic response to competition on an unconscious level.*

The confidence reports in the deterrence and hustler treatments will be discussed again in the analysis of the strategic behavior of participants. Here we only mention that we cannot reject

rational Bayesian updating using the Burks et al. (2009) allocation function. This may reflect our having only two intervals, either above or below the median.

4.2. **Voluntary tournament entry.** In the deterrence and hustler treatments, player R chooses whether to enter a tournament with player S, who is automatically entered into the tournament. This result contrasts with the other treatments in which both people are automatically entered into the tournament. R can take into account S's reported confidence before deciding whether to enter the tournament. Player S in turn knew that player R would observe his statement (and player R knew this, etc., since the instructions were read aloud and so were known to be identical for all participants), and could potentially anticipate the effect of the statement on player R's decision. S does not observe R's statement, so this statement could not affect R's behavior. In light of this, what determines player R's choice?

Our data show that with the high outside option in the deterrence treatment and in the hustler treatment, player R is much more likely to enter the tournament when own confidence is higher and when the opponent's confidence is lower.[11] Indeed, as is shown in Figure 2, we observe that relative confidence is a phenomenally good predictor of entry. In the deterrence case, 23 of 25 receivers (92.0%) enter when their confidence level is at least as large as the paired sender's reported confidence level, while only four of 23 receivers (17.4%) enter when their confidence level is lower than the paired sender's reported confidence level; the difference in these proportions is highly significant ($Z = 5.21$, $p = 0.000$). The corresponding data for the hustler treatment is 28 of 32 (87.5%) choosing to enter with higher stated confidence and five of 16 (31.2%) choosing to enter with lower stated confidence; the difference in these proportions is highly significant ($Z = 3.96$, $p = 0.000$). Thus, there is strong potential for senders to influence the receiver's decision.

In the deterrence treatment, we find that males enter twice as frequently as do females, 75.0 percent versus 37.5 percent ($Z = 2.62$, $p = 0.009$, two-tailed test), as is shown in Figure 3. However, this does not reflect a difference in performance: females in the R role in the high-option condition do nearly as well as males on the Raven test (the mean score for males is 9.12 and the mean score

---

[11]We focus primarily on entry with the high outside option, since 28 of 30 receivers chose entry with the low outside option, so that statistical tests have little power.

FIGURE 2. **Entry by lower confidence.** Error bars: +/- SE
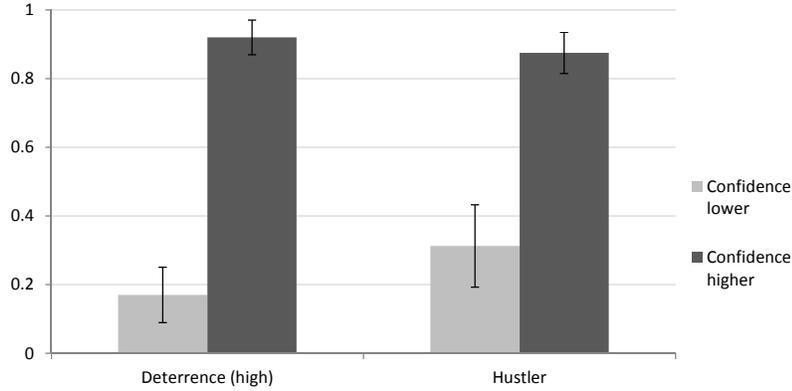
for females is 8.88; Wilcoxon ranksum test: $Z = 0.28$, $p = 0.779$, two-tailed test). At first glance, this seems to be evidence that females are *per se* averse to competition. However, female receivers state significantly lower confidence levels than do male receivers in this condition, 56.63 versus 75.83 ($Z = 3.07$, $p = 0.002$, two-tailed test). Men choose to compete more frequently, but this reflects a lower stated confidence level. This effect is only seen for people who choose to enter the tournament; the average stated confidence level for male entrants is 84.17 versus 69.89 for female entrants, while this comparison is 50.83 versus 48.67 for male non-entrants and female non-entrants, respectively.

The results are quite different in the hustler treatment, where the entry rate for males (65.22 percent) is lower than the entry rate for females (72.00 percent). Also the average stated confidence level for male entrants is 73.87 versus 69.00 for female entrants, while the average stated confidence level for male non-entrants is 49.62 versus 49.00 for female non-entrants, so we can see that the stated confidence levels for receivers in the hustler treatment is largely unaffected by gender. The

performance level was 8.82 for males and 8.56 for females, not significantly different (Wilcoxon ranksum test: $Z = 0.20$, $p = 0.843$, two-tailed test).

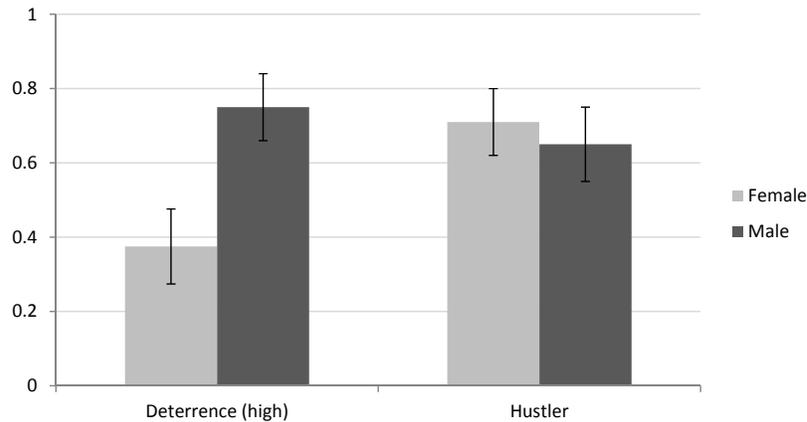FIGURE 3. **Entry by gender.** Error bars: $+/-$ SE

Table 4 reports the probit estimates of the decisions to enter the tournament.[12] The first three columns apply to the high outside-option sessions of the deterrence treatment, and the last three columns apply to the hustler treatment. Specification (1) shows that own confidence increases the likelihood of entering the tournament, and the confidence of the opponent decreases it. Each variable substantially affects the probability of entering. Specification (2) includes a dummy variable that simply compares if own confidence is higher or lower than that of the opponent. Controls in (3) for gender, number of correct answers, and risk aversion have no significant effect. Thus, the lower likelihood of entry by females seems driven by lower confidence, rather than less competitiveness.[13]

---

[12]Estimates from the Linear Probability Model are qualitatively very similar to the reported Probit marginal effects.
[13]This also suggests that males are not just reporting higher confidence, but also feel more confident. If they were just reporting higher confidence without believing it, then, controlling for confidence, males should have been less likely to enter the tournament.

The three specifications for the hustler treatment give similar results, with smaller magnitudes, but the coefficient for own confidence is not significant. Note that once again there is no evidence that women are *per se* less likely to enter.

[Table 4 about here]

Results (Tournament Entry)

(1) *When deciding whether to enter the tournament, participants are more likely to enter when their confidence is higher; they are also sensitive to the confidence reported by the opponent: If own stated confidence is lower than that of the opponent, subjects are far less likely to enter.*

(2) *Females are less likely to enter the competition in the deterrence treatment, but this effect is mainly due to the difference in confidence. Once we control for confidence, the entry rate of women is not significantly lower. There is no entry difference in the hustler treatment, regardless of whether or not we control for confidence.*

4.3. **Updating errors.** Comparing choices across the Reports and Machine questions tasks in Treatment 1 allows us to test the hypothesis that the patterns of stated confidence that we observe are only due to errors in Bayesian updating. Recall that each person received a report about his or her rank in the group, and then answered three machine questions, each with a different percentage of days that the mechanic went to the left machine.[14] With this information, we can test whether mistakes in the Reports task are due to a general inability to update posterior probabilities.

Reports of one's being in the top 2 are always true and in fact are always believed. So we focus on the response to negative feedback ("not in the top 2"). There is a threshold prior confidence level of 2/3 for whether or not one should believe one is in the bottom 2 after a negative report. Thus, an updating error for someone with initial confidence below this level would mean that one is

---

[14]Note that no feedback about actual performance is given, so that nothing is learned and the estimates given for the Machine questions seem independent of the estimate given for the Reports question. There is no obvious reason why the order of these choices should matter; if anything, one might expect the estimate for the Machine question with the same probability of visits as the level of confidence reported to be sticky, rather than changing.

more likely to be in the top 2 after receiving negative feedback; for someone with initial confidence above the threshold, the error would consist of believing the opposite. The two errors represent respectively a form of under-updating and over-updating.

FIGURE 4. **Updating error rates for reports and machine questions, by confidence interval.** An updating error is defined as an answer differing from the one provided by Bayesian updating of a prior equal to the stated confidence of the subject. Error bars: +/- SE
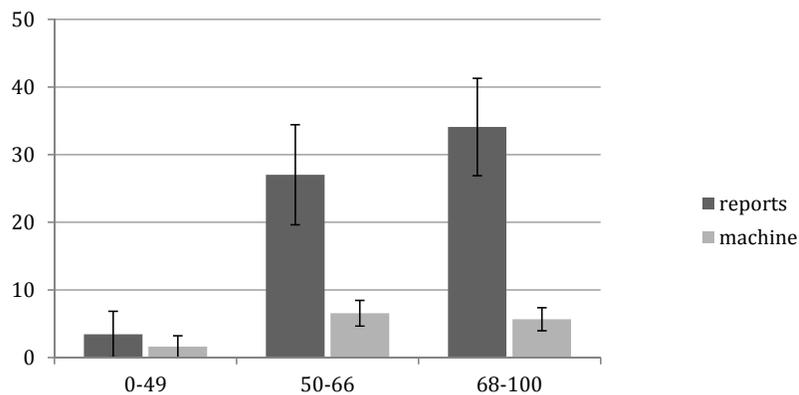
The test that errors are due to a general inability to update is provided by the comparison of the frequency of errors depending on whether or not the updated belief concerns one's own relative skill. Figure 4 compares mistakes made by subjects in each of those intervals after they receive negative feedback in the Reports and Machine settings. There are very few errors for people with stated initial confidence below 50 in either case. Updating mistakes are always rare with the Machine questions, as the percentage of mistakes ranges from less than two percent to less than seven percent. The picture is very different for the Report questions: 27 percent of subjects make under-updating errors in the interval between 50 and 67, and 34 percent of subjects

believe negative feedback when stated confidence was larger than 67 percent.[15] The error rates on the Report questions are significantly higher than with the Machine questions (Test of equal proportions gives $Z = 3.72$ and $Z = 5.46$ for the respective comparisons, both $p = 0.000$).

What is the source of errors in the Reports? It is clearly not the case that our participants are unable to do the Bayesian updating task, given their performance on the Machine questions. Thus, it seems that there is something about own ability or performance that induces these errors. Our view is that this reflects a desire (perhaps on an unconscious level) for high self-esteeem or ego utility.[16] Nevertheless, this is also modulated by intelligence, as measured by the Raven test. Participants making updating errors on the Reports seem to do worse in terms of performance. If we group over-updating and under-updating errors, we find that those making errors have fewer correct answers on the Raven test: 7.6 is the mean among participants making updating errors versus 8.3 of those who do not. Controlling for reported confidence, this difference is significant (column 1 in Table 5).[17] Thus, a lower cognitive ability is associated with making updating errors. We find a similar effect for errors on the Machine questions but the coefficient is not significant (column 2). In the case of the Reports, the difference in performance translates into a large effect on the likelihood of actually being in the top 2 (19 percent versus 41 percent, $Z = 2.00$, $p = 0.045$), but not in the case of errors on the Machine questions (32 percent versus 37 percent, $Z = 0.41$, $p = 0.680$).

[Table 5 about here]

Results (Updating Errors)

---

[15]The fractions of errors in the intervals 50-67 and above 67 are not significantly different from each other ($Z = 0.69$, $p = 0.493$).

[16]Another possibility is that participants may have misrepresented their initial confidence levels, despite the fact that they were incentivized to report truthfully.

[17]If we estimate model (1) of Table 3 separately for both types of errors, the coefficients remain similar but become insignificant on the intervals 50-67 and above 67, possibly due to a lack of power. Nonparametric tests show that participants making over-updating errors (interval above 67) do significantly worse on the Raven test (Z = 1.776, p = 0.076), but there is no significant effect for under-updating errors (interval 50-67, Z = 1.107, p = 0.269).

(1) *Errors in updating are much more likely to occur for questions involving own skill judgment than for the abstract questions; hence such errors do not merely reflect a general inability to perform Bayesian updating. Errors in response to the abstract questions are rare.*

(2) *There is no significant difference in the rates of under-updating errors and over-updating errors.*

(3) *People who make errors on the Reports do worse on the Raven test and are much less likely to be in the top 2, controlling for confidence.*

5. Self confidence and its motivations

What do these results tell us on the origin and motivation of overconfidence? One key potential motivation for being overconfident that has been suggested is the ego utility that one derives, producing an increase in self-esteem. In our data we observe substantial overconfidence even when the stated confidence level is not observed by anyone else. This finding suggests that people are either poor judges of probabilities, or that they receive some internal benefit from this inflated belief, or that think they might influence others' behavior.

The fact that people make far fewer updating errors with the Machine questions suggests that the explanation of this overconfidence is not poor ability to estimate probability of events. On the other hand errors in updating do not exhibit any bias in the direction suggested by ego-utility or preservation of self-esteem. Many subjects with a prior stated probability lower than the threshold update and receive a negative report still state a probability larger than 50% that they are in the top 2, and this might be considered a bias induced by self-image preservation. But correspondingly many subjects who should, according to their maintain a probability larger than 50% do not do so facing a negative report, a bias that goes contrary to self-image preservation.

These results are consistent with our general hypothesis that views strategic concerns as a primary source of overconfidence. In fact we see strong evidence that an increase (decrease) in a sender's reported confidence can have deterrent (encouragement) value in terms of inducing the receiver in or out of the tournament. We also see some evidence (see Figure 1 and Table 3) that

males report higher confidence in the strategic condition than in the baseline treatment. How close is this behavior to that which is optimal for senders? We take this up in the next subsection.

5.1. **Optimality of decisions.**

Behavior of receivers. We already saw that much of the behavior of receivers can be explained by the simple rule that a receiver enters if and only if his own confidence is at least as high as the reported confidence of the sender. To test how precise this description is we check it against the data, and find that the rule correctly classifies 87.5% of the receivers' decisions in the deterrence treatment (high outside option), and 81.25% in the hustler treatment. Moreover, most of the incorrectly classified decisions are close to the cutoff level.

Optimal reporting. If we assume that receivers indeed play this strategy, and that senders anticipate this, we can analyze the best response of senders. We will model this by assuming that players have types, $\theta \in [0, 100]$, that are drawn from a continuous distribution function with density $f(\theta)$. We index players by $i = S, R$ (sender and receiver respectively). Players choose a message $t^i \in T = [0, 100]$, so that the message space is the same as the type space. The message of a player is his reported confidence, and the type is his true belief about his confidence. In our experiment a receiver has no incentives to report a confidence that differs from his type, so we assume $t^r = \theta^r$. After observing t^s, receivers choose an action in the set $\{Out, In\}$. The assumed strategy of the receiver is then to play In if and only if $t^r \geq t^s$.

Let O^S be the sender's outside option payoff if the receiver chooses Out. If the receiver chooses In, the sender's payoff is v^h if he wins and v^l if he loses. The probability that the sender wins is $\Pi(\theta^s, \theta^r)$. We will specify precise functional forms of Π. The expected payoff for the sender of the tournament is then given by:

$$(2) \qquad \int_0^{t^s} O^S dF(\theta^r) + \int_{t^s}^{100} (\Pi(\theta^s, \theta^r)(v^h - v^l) + v^l) dF(\theta^r).$$

The reason for reporting an inaccurate confidence level is to change the probability that the receiver chooses In. The optimal reported confidence for a risk-neutral sender is determined by:

28

(3) $$f(t^s)[O^S - \Pi(\theta^s, t^s)(v^h - v^l) - v^l] = c(t^s - \theta^s).$$

The RHS reflects the fact that players have the incentive provided by the quadratic scoring rule to report truthfully, creating costs when their reported confidence differs from their true belief (where $c = 2/10,000$). In the deterrence treatment, the term in brackets on the LHS is positive so that over-reporting is optimal ($O^S = v^h = 10, v^l = 0$), while in the hustler treatment this term is negative so that underreporting is optimal ($O^S = 10, v^h = 25, v^l = 10$).

We specify

(4) $$\Pi(\theta^s, \theta^r) = 1/(1 + e^{-\delta(\theta^s - \theta^r)}),$$

where $\delta = .021$ is estimated from the data of the baseline treatment in which there are no incentives to over-report. For $F(\theta)$, we assume that players believe that types are normally distributed (truncated at 0 and 100) with mean 50 and standard deviation 21. The value of the standard deviation is estimated from the data and we take a mean of 50 to reflect that players do not believe that other players are on average overconfident.

Figure 5 plots the optimal reporting for senders under the assumptions made. The thin solid line represents truthful reporting. The thick solid line represents the optimal report in the deterrence treatment, and the dashed line for the hustler treatment. In both treatments it is optimal for senders to deviate substantially from their true belief. For instance, the optimal report for a sender with a confidence of 60 is 80 in the deterrence treatment, and 20 in the hustler treatment. In both cases the optimal deviation from truthful reporting is substantially larger than what we see in the data. Based on reported confidence levels in the baseline treatment, we should expect that the average confidence is about 20 higher in the deterrence treatment (we find roughly 0 for females and 10 for males), and 45 lower in the hustler treatment (we find about 10-15 lower). We should also find no reported confidence below 60 in the deterrence treatment (because even for type 0 the optimal report is above 60), or above 30 in the hustler treatment (because even for type 100 the optimal report is below 30), but we see quite a few examples in the data.

FIGURE 5. **Optimal reporting by receivers given the behavior of senders.**
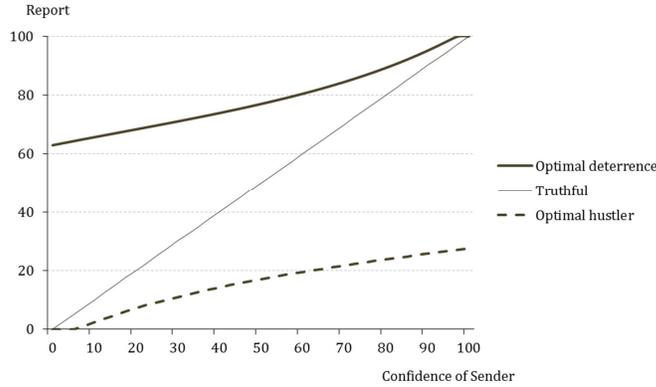
We conclude that the behavior of senders goes in the right direction, but they under inflate and deflate.[18] The fact that they do not exactly match the estimated optimal levels for reports is not surprising. They must form expectations about several parameters, e.g., those related to the distribution of types, and they only play the game once.

Receivers, on their turn, appear to take the confidence statements at face value instead of deflating them in the deterrence treatment or inflating them in the hustler treatment, as they simply seem to compare their own confidence to the reported confidence by the sender. This behavior is an indication that receivers also do not anticipate a level of over- and underinflating as high as our estimated optimal reports. This again supports the idea that overconfidence is automatic and not well anticipated, although with our data we cannot rule out that receivers anticipate over- and under-reporting to some extent.

---

[18]The exact magnitude depends on the assumptions we make. In particular, the distribution of types matters. We have also estimated optimal reporting for alternative distributions (assuming a different mean or a uniform instead of normal distribution) but the underinflating and deflating seems robust to different specifications.

6. Conclusion

Our experiments examined the determinants of self-confidence, and the degree to which it reflects strategic concerns about social image (utility from the perceptions of others). Our main conclusion is that levels of stated con fidence are likely to be influenced by strategic interest, perhaps unconsciously processed. We see evidence that people will inflate or deflate incentivized statements of confidence levels when doing so is strategically beneficial. We suggest that inflating confidence when doing so is not strategically beneficial reflects the notion of salient perturbations. In familiar situations overconfidence has strategic value, so that we may observe it in non-strategic environments that are similar to the familiar situation.

Our novel strategic environment (in which another party observes the stated confidence level of another and then chooses whether or not to enter a tournament with this other person) allows a direct test of the strategic-interest hypothesis. First, the social signal is perceived and has consequences: subjects in our experiment do respond to statements about confidence made by others, taking that information into account when choosing whether or not to enter. In the deterrence treatment, male (but not female) participants on average report significantly higher confidence levels than in the non-strategic treatments. Inflated confidence serves as an effective deterrent. Interestingly, males (but not females) do so in both roles, even when deterrence is impossible; this suggests processing on an unconscious level.

In the presumably less-familiar hustler environment, we observe deflated confidence for both men and women in the role of senders, which serves to encourage entry. We argue that conscious cognition is present in this less-familiar environment. Strategic deterrence and hustling are consistent with the equilibrium we characterize; the degree to which one engages in costly strategic distortion depends on the values of the parameters in the game.

We also find a substantial proportion of updating errors about one's own ability. However, this is not a consequence of a generic error in Bayesian updating because people are much more accurate in updating their beliefs when the information is about an impersonal issue then they are when new information concerns their own ability. In fact, they are hardly ever wrong in the abstract

setting. This shows that information processing when own reputation is at stake is of a different nature than abstract, neutral information processing.

When inflated reported confidence is strategic, it is natural to find gender differences in our participants' behavior, given the evidence of other gender differences such as with respect to financial risk preferences (Charness and Gneezy 2010, 2012), competition (Gneezy, Niederle, and Rustichini 2003), and even shame (Ludwig and Thoma 2012). But since hustling is a much less familiar environment and strategic distortion is presumably driven by cognitive ability (which is the same for men and women on the Raven test), we see men and women engaging equally in this behavior.

We also find no evidence that women shy away from competition. While men choose to enter a tournament much more frequently than women do in the deterrence treatment, there is no difference when one controls for confidence; there is no difference in entry rates or stated confidence in the hustler treatment. So women are not less competitive than men in our data.

There are a number of directions for future research. Is some of the observed behavior truly unconscious? To what extent is self-deception present? An additional question is whether the behavior we observe in the voluntary-entry treatment would persist over time, as would be predicted by the equilibrium we have described.

References

Baumeister, R. F. (1998), "The Self," in D. Gilbert, S. Fiske and G. Lindzey (Eds.), *The Handbook of Social Psychology*, Boston: McGraw-Hill.

Benoit, Jean-Pierre and Juan Dubra (2011), "Apparent Overconfidence," *Econometrica*, forthcoming.

Berglas, S. and E. Jones (1978), Drug choice as a self-handicapping strategy in response to non-contingent success, *Journal of Personality and Social Psychology* 36, 405–417.

Burks, S., J. Carpenter, L. Goette, and A. Rustichini (2009), "Cognitive abilities explain economic preferences, strategic behavior and job performance, " *Proceedings of the National Academy of Sciences* 106, 19, 7745-7750.

Burks, S., J. Carpenter, L. Goette, and A. Rustichini (2012), "Overconfidence and Social Signaling," *The Review of Economic Studies*, forthcoming.

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

Charness, G. and U. Gneezy (2010), "Portfolio Choices and Risk Attitudes," *Economic Inquiry* 48, 133-146.

Charness, G. and U. Gneezy (2012), "Strong Evidence for Gender Differences in Risk-taking," *Journal of Economic Behavior and Organization* 83, 50-58.

Cho, I.K, and D. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics* 102, 179-221.

Dichgans, J., and T. Brandt (1978), "Visual-vestibular interaction: effects on self-motion perception and postural control," In: *Handbook of Sensory Physiology*, edited by R. Held, H. Leibowitz, and H-L. Teuber. Berlin: Springer, 755-804.

Dunning, D., J.A. Meyerowitz, and A.D. Holzberg (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.

Eil, D. and J. Rao (2011), "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics* 3(2), 114-138.

Ertac, S. (2011), "Does Self-Relevance Affect Information Processing? Experimental Evidence on the Response to Performance and Non-Performance Feedback," *Journal of Economic Behavior & Organization*, 80(3), 532-545.

Ewers, M. (2012), "Information and Competition Entry," *IZA Discussion Paper* 6411.

Gilboa, I. and D. Schmeidler (1995), "Case-based Decision Theory," *Quarterly Journal of Economics* 110(3), 605-639

Gneezy, U., M. Niederle, and A. Rustichini (2003), "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics* 118(3), 1049-1074.

Gneezy, U. and A. Rustichini (2004), "Gender and Competition at a Young Age," *American Economic Review* 94(2), 377-381.

Grossman, Z. and D. Owens (2010), "An Unlucky Feeling: Overconfidence and Noisy Feedback," mimeo.

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

Kartik, N. (2009), "Strategic Communication with Lying costs," *Review of Economic Studies* 76, 1359-1395.

Kolditz, T. A., and R. M. Arkin (1982), "An impression management interpretation of the self-handicapping strategy," *Journal of Personality and Social Psychology* 43, 492-502.

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

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

Leary, M. (1999), "The social and psychological importance of self-esteem," in R. M.Kowalski and M. R. Leary, eds., *The social psychology of emotional and behavioral problems: interfaces of social and clinical psychology*, 197-222.

Leary, M. R. Tambor, S. Terdal, and D. Downs (1995), "Self-esteem as an interpersonal monitor," *Journal of Personality and Social Psychology* 68, 518-530.

Liberman, V., S. Samuels, and L. Ross (2004), "The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves," *Personality and social psychology bulletin* 30, 1175-85.

Ludwig, S. and C. Thoma (2012). "Do Women Have More Shame than Men: An Experiment," unpublished manuscript.

Möbius, M., M. Niederle, P. Niehaus and T. Rosenblat (2011), "Managing Self-Confidence: Theory and Experimental Evidence, mimeo.

Moore, D. and D. Cain (2007), "Overconfidence & Underconfidence: When and Why People Underestimate (and Overestimate) the Competition," *Organizational Behavior and Human Decision Processes* 103(2), 197-213.

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

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

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.

Rabin, M. and J. Schrag (1999), "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics* 114, 37-82.

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

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

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

Vialle, I., L. Santos-Pinto, and J.L. Rullière (2011), "Self-Confidence and Teamwork: An Experimental Test, "Mimeo.

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

Wolosin, R. J., S. Sherman, and A. Till (1973), "Effects of Cooperation and Competition on Responsibility Attribution After Success and Failure," *Journal of Experimental Social Psychology* 9, 220-23

The comments in square brackets are meant to illustrate instructions to the reader and were not part of the instructions.

General instructions

Introduction Welcome to our experiment. You will receive €7 for showing up, regardless of the results. The instructions are simple. If you follow them carefully, you can earn a substantial amount of money in addition to your show up fee. Throughout the stages we will ask you to answer questions. At each stage, you will receive more detailed instructions.

You will be part of a group of 4 persons. You don't know who the other persons are, and you will remain anonymous to them. All your choices and the amount you will earn will remain confidential and anonymous, except if explicitly indicated otherwise. You will receive your earnings in an envelope. The person that puts the money in the envelopes can only see the login number that has randomly been assigned to you, and cannot match any names, student numbers, or faces with the login numbers and the decisions made.

Payments There are several items in the experiment for which you can earn points. At the end of the experiment, one item is randomly chosen and your points for that item are paid in addition to the show-up fee (1 point is worth € 1). One of the participants is randomly chosen to be an assistant during the experiment. There is a random component in the experiment. The task of the assisting person will be to throw a dice which will determine the outcome.

No deception Remember, we have a strict no deception policy in this lab.

Questions Please remain seated and raise your hand if you have any questions, and wait for the experimenter. Please remain silent throughout the experiment.

Part 1.
In the first stage, all group members receive the same 15 questions. You will see a matrix with one missing segment at the bottom right. Your task is to identify the segment that would logically fit at the position of the missing segment, by choosing from the suggested answers. You can make your choice by clicking the corresponding number on the right of your screen. [A screenshot with an example question was provided.]

You can go back and forth between the questions. There is a time limit of 8 minutes. The time remaining is indicated on your screen.

After the time limit, we will rank all 4 people in your group depending on the number of questions answered correctly. The person with the highest score will get rank 1, and the person with the lowest score will get rank 4. In case of ties, the computer will randomly determine who gets the higher rank. After this, you will get some questions regarding how well you think you did.

We then randomly divide the group in 2 players A and 2 players B. Every player A will be matched

against a player B. If your rank is higher than the player with which you are matched, you can receive 10 points.

Part 2.
All four group members have now finished with the questions, and we have determined the rank of every person.

We now ask you to indicate how likely you think that you are among the top 2 of your group. You can indicate this on a scale from 0 to 100%. Indicating 0% means that you are sure you are not among the best 2 of your group, while indicating 100% means that you are sure you are among the top 2 of your group. Similarly, 50% indicates that you think it is equally likely that you are among the best 2 of your group, or that you are not among the best 2 of your group.

We will pay you for the accuracy of your estimate. You earn more points for this item if your estimate is more accurate. The formula that is used to calculate the amount of money you earn is chosen in such a way that your expected earnings are highest when you report to us what you really believe. Reporting any value that differs from what you believe decreases your expected score for this item. If you are interested, you can find some detailed examples of this to see how this works. [An explanation with examples was available to participants, see below.]

The role of player A and player B We matched you with one other randomly chosen person from your group. You are either Player A or B, and this is randomly determined.

[*baseline*] None of the players can see the other player's estimate of being in the top 2.

[*social*] Player A will not see the estimate by player B that he or she is among the best two in the group, but player B will see the estimate by Player A that he or she is among the best two of the group.

[*baseline and social*] Later on in the experiment, we will compare the rank of player A with the rank of player B, and for that item the player with the highest rank receives 10 points, the other nothing. Both of you will see who has the highest rank, and this ends the stage.

[*strategic deterrence and hustler*] Player A will not see the estimate by player B that he or she is among the best two in the group, but player B will see the estimate by Player A that he or she is among the best two of the group.

Later on in the experiment, after player B has observed the estimate of player A, player B will choose between two options: IN or OUT.

If player B chooses OUT, then for that item player B receives 3.5 [5.5] points and player A automatically receives 10 points. Both players will see who has the highest rank, and this ends the stage.

[*deterrence*] If player B chooses IN, we will compare the rank of player A with the rank of player B, and for that item the player with the highest rank receives 10 points, the other nothing. Both

of you will see who has the highest rank, and this ends the stage.

[*hustler*] If player B chooses IN, we will compare the rank of player A with the rank of player B. For this item, Player A receives 25 points if (s)he is the highest ranked player, and 15 points if (s)he is not the highest ranked player. Player B receives 10 points if (s)he is the highest ranked player, and 0 points if (s)he is not the highest ranked player. Both of you will see who has the highest rank, and this ends the stage.

You can see what role you have on the top left of your screen (see the example below). [Participants could see their role on the next screen.]

Determination of your score What follows is a brief explanation about the determination of your score, showing that it is in your interest to report truthfully what you believe in order to maximize your expected earnings.

The score is determined as follows. You start with 10 points. We subtract points depending on how close your reported belief is to the outcome. The outcome is set to 1 if you are in the top 2, and to 0 if you are not.

For instance, if you report 70% (.7), and you are in the top 2 (outcome is 1), you are .3 away from the outcome, while if you are not in the top 2 (outcome is 0), you are .7 away from the outcome.

The difference with the outcome is squared and multiplied by 10, and then subtracted from the 10 points that you start with. Thus in the example with 70%: if you are in the top 2, this gives you $10 - 10(.3)^2 = 9.1$. If you are not in the top 2, this gives $10 - 10(.7)^2 = 5.1$. You would weight these two scores by your belief about the likelihood of each occurring.

Larger differences between your reports and the outcome decrease your score proportionally more than small differences. To minimize the expected difference, and maximize your expected score, you should report what you believe.

The following examples illustrate that your expected score is highest when you report your true beliefs. *All numbers used are for illustrations only and are no indication for the decisions for you to take.*

Example 1
You believe 50% and report 50%. As a simple example: if you believe there is a 50% chance you are in the top 2, and you report 50%, then there is always a difference of .5 with the outcome, and since this is squared we always subtract 10 times $(0.5)^2$ points from your score, i.e. 2.5 points. Your expected score is 7.5.

You believe 50% but you report 100%. If you report 100%, then in one case there is no difference (if you are in the top 2) and no points are subtracted. But in the other case the difference is 1 (if you are not in the top 2), and then we subtract 10 times $(1)^2$ from your score. If you believe the likelihood of being in the top 2 is 50%, you expect this to happen in 50% of the cases, so the amount subtracted would be $10(0.5) = 5$. This gives you an expected score of 5, which is lower than if you report your belief of 50%.

Example 2
You believe 70% and report 70%. As another example, suppose that you think there is a 70%

likelihood that you are among the best 2. If you report 70%, your score will be either 9.1 (if you are in the top 2) or 5.1 (if you are not in the top 2). You believe that with 70% chance your score will be 9.1, and with 30% your score will be 5.1. So your expected score is $0.7(9.1) + 0.3(5.1) = 7.9$.

You believe 70% and report 100%. Now suppose that, instead of reporting this belief of 70%, you report another number. For instance, you report 100% (1). This means that if you are in the top 2, the outcome is as predicted, and you get $10 - 10(0)^2 = 10$ points. If you're not in the top 2, you are 1 away from the outcome, and your score will be $10 - 10(1)^2 = 0$. Since you actually expect to be in the top 2 with 70% chance, your expected score is 7. This is lower than if you would have reported 70%.

You believe 70% and report 20%. The same is true if you report a number below your belief, for instance 20% (.2). If you are in the top 2, your score would be $10 - 10(0.8)^2 = 3.6$ points. If you're not in the top 2, your score will be $10 - 10(0.2)^2 = 9.6$. Since you actually expect to be in the top 2 with 70% chance, your expected score is $0.7(3.6) + 0.3(9.6) = 5.4$, again lower than if you would have reported 70%.

The table below shows the expected scores for some more possible beliefs you may have and reports you give. As you can see, expected scores are highest when the reported belief is equal to the true belief (the cells on the diagonal that are highlighted in green).

Expected scores

Your report (%)	Your Belief (%)										
	0	10	20	30	40	50	60	70	80	90	100
0	10	9	8	7	6	5	4	3	2	1	0
10	9.9	9.1	8.3	7.5	6.7	5.9	5.1	4.3	3.5	2.7	1.9
20	9.6	9	8.4	7.8	7.2	6.6	6	5.4	4.8	4.2	3.6
30	9.1	8.7	8.3	7.9	7.5	7.1	6.7	6.3	5.9	5.5	5.1
40	8.4	8.2	8	7.8	7.6	7.4	7.2	7	6.8	6.6	6.4
50	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5
60	6.4	6.6	6.8	7	7.2	7.4	7.6	7.8	8	8.2	8.4
70	5.1	5.5	5.9	6.3	6.7	7.1	7.5	7.9	8.3	8.7	9.1
80	3.6	4.2	4.8	5.4	6	6.6	7.2	7.8	8.4	9	9.6
90	1.9	2.7	3.5	4.3	5.1	5.9	6.7	7.5	8.3	9.1	9.9
100	0	1	2	3	4	5	6	7	8	9	10

[*baseline*] **Part 3.**
Based on your true ranking in the group, we will send you a report. The report will say if you are among the two best of your group, or if you are not among the two best of your group.

However, *sometimes the report will be incorrect.* The way this works is as follows.

If you *are not* among the top two of your group, then the report will always be correct and inform you that you are not among the best two of your group.

If you *are* among the top two of your group, the report is mistaken in half of the cases. That is, in half of the cases, the report correctly informs you that you are among the top two of your group. In the other half of the cases, the report is incorrect and says you were not among the top two of your group, even if you were.

Whether or not the report you receive is correct when you are among the top two of your group, depends on the outcome of a dice throw by the assistant. You will not see the outcome, but if the assistant throws 1, 2, or 3, you will receive a correct report when you are among the top two. If the assistant throws 4, 5, or 6, you will receive an incorrect report when you are among the best two of your group. (For some groups, the incorrect report is sent after different values of the dice, but in any case the report is incorrect in half of the cases when you are among the best 2.)

After you see the report, we will ask you if you think the report is more likely to be correct or incorrect.

You earn 10 points if you are right.

[*baseline*] **Part 4**
In this part, we ask you some questions about the scenario below. The first part is always the same, but some additional information is given in the question, so please read it carefully. For this part, we randomly choose a question and this is treated as a single item.

Scenario Consider two machines placed in two sides of a large production hall, left side = L and right side = R. The two machines produce rings, good ones and bad ones. Each ring that comes from the left machine, L, has a 50% chance of being a good ring and a 50% chance of being a bad ring. Each ring that comes from the right machine, R, is good. Both machines produce 100 rings every day.

The mechanic visits the production hall every day, and randomly examines one of the machines by taking one ring. On some days, he takes a ring from the left machine, and the other days he takes a ring from the right machine. Suppose the ring he takes is *good*.

We will ask you if it is more likely that the mechanic went to the left or right machine.

Example question 'On 50% of the days, the mechanic takes a ring from the left machine, and the other 50% of the days from the right machine. Of the rings that come from the left machine, on average half are good and half are bad. Each ring that comes from the right machine is good.

Imagine the ring he takes is good. Is it more likely to come from the left or right machine?'

You will get 3 questions like this one. We vary the percentage of days that the mechanic goes to the left or right machine, but everything else remains the same.

You earn 10 points if you are right.

Part 5.

[*baseline and social*] In this part, you are informed if player A or B has the highest rank.

[*strategic deterrence and hustler*] Player A will not see the estimate by player B that he or she is among the best two in the group.

Player B will see the estimate by Player A that he or she is among the best two of the group, and then gets the choice between two options: IN or OUT.

[We repeated the instructions of Part 2 in which the payoffs were given for this item].

FIGURE 6. **Illustration of the signal structure in the Report and the Machine question.** This figure was not provided to subjects, and is only meant as illustration.
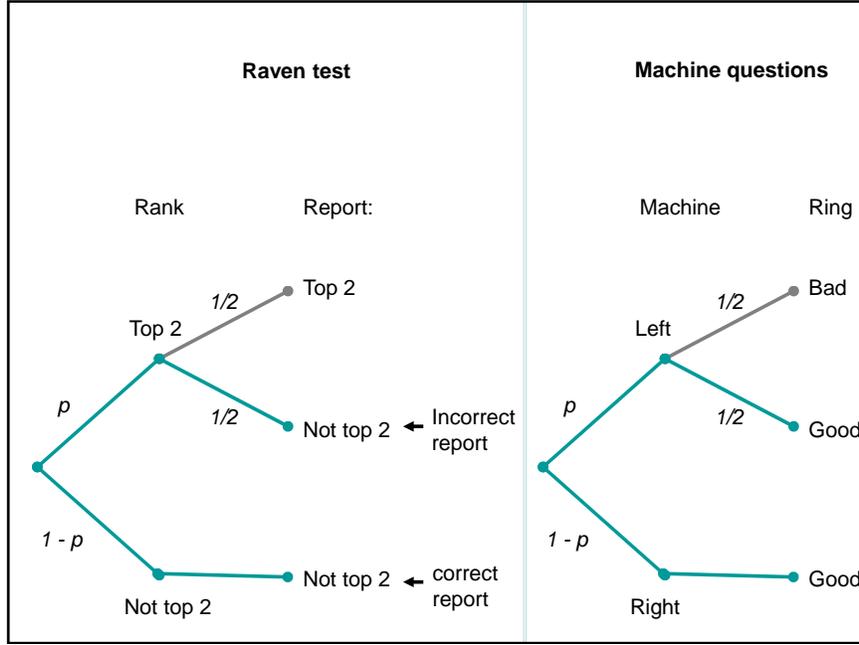
APPENDIX B. MODEL

In this appendix we characterize the equilibrium set of the extensive form game described in section 3.1. Our aim is to formally show that the following intuition is equilibrium behavior: (i) In the deterrence treatment, the sender has a relatively high outside option and will therefore overreport to appear strong and deter the receiver from entering the tournament, (ii) In the hustler treatment, the sender has a relatively low outside option and will therefore underreport to appear weak and encourage the receiver to enter the tournament.

B.1. **Outline.** We start with repeating the payoff structure and introducing some notation and definitions, making precise what we mean by over- and underreporting. We then analyze the equilibrium set when the payoffs are symmetric, as in the deterrence treatment.

After establishing the conditions under which any type of equilibrium may occur (pooling, separating, partial separating) in section B.3, we summarize in section B.4 the entire characterization of the equilibrium set (see Theorem B.9).

In section B.5 we compute the predictions of the model for the payoff used in the experiment. We first (section B.5) plug in the parameters that we used in the deterrence treatment of our experiment to show that over-reporting is part of the equilibrium in that treatment. We then (section B.5) extend the analysis to asymmetric payoffs and show that underreporting is equilibrium behavior with the parameters of the hustler treatment.

B.2. **Preliminaries.** Recall that θ_j^i denotes player $i \in \{1,2\}$ of type $j \in \{0,1\}$. The sender is indexed as player 1 and the receiver as player 2. A weak player is indexed as 0, a strong player as 1. We assume that the payoffs of players are symmetric as in the following payoff table:

(5)

	θ_0^2	θ_1^2
θ_0^1	$0,0$	b,a
θ_1^1	a,b	d,d

This is the case for the deterrence treatment. We will return later to the hustler treatment which has asymmetric payoffs. We consider:

(6) $$a \geq O^2 \geq d \geq 0 \geq b,$$

where O^2 is player 2's outside option. This implies that Out strictly dominates In for a weak player 2. Strategies of player 1 are functions from type to probability on signals: $\sigma^1(\theta^1; \cdot) \in \Delta(T)$. Strategies of player 2 are functions from type and signal of player 1 to probability on actions: $\sigma^2(\theta^2, t; \cdot) \in \Delta(\{In, Out\}))$. To lighten notation we call in the following $\sigma(\theta_1^2, t_0; In) = r, \sigma(\theta_1^2, t_1; In) = s$, where t_0 and t_1 are the low and high message respectively.

We now make precise what we mean by over- and underreporting.

Definition B.1. *(Over- and underreporting) We call an equilibrium in our game under-reporting if:*

(7) $$\sigma^1(\theta_0^1; t_0) = 1; \sigma^1(\theta_1^1; t_0) \equiv \tau \in (0,1).$$

that is if the low type only reports a low type, and the high type reports a low type with positive probability. We call an equilibrium over-reporting equilibrium if:

(8) $$\sigma^1(\theta_1^1; t_1) = 1; \sigma^1(\theta_0^1; t_1) \equiv \sigma \in (0,1).$$

B.3. **Types of equilibria.**

B.3.1. *Monotonic equilibria.* We first examine monotonic equilibria, i.e., those where the function $\theta^1 \to \sigma^1$ is increasing (higher types give higher signal). We then show that non-monotonic equilibria do not exist. In our simple model an equilibrium is monotonic if:

(9) $$\sigma^1(\theta_1^1; t_1) \geq \sigma^1(\theta_0^1; t_1)$$

and we say it is strictly monotonic if the inequality (9) is strict.

The equilibrium set is easily characterized if we take into account the following. Take the $(\sigma, \tau) \in [0,1]^2$ pairs describing as in (7) and (8) the strategy of player 1.

Lemma B.2. *If (6) holds, then for generic payoffs the only monotonic equilibria are either the fully revealing truthful, or the two pooling (at the low and high type respectively) or the under or over-reporting equilibria.*

Proof. Suppose that an equilibrium exists with $(\sigma, \tau) \in (0,1)^2$. This implies that the two signals are indifferent for both types of player 1. This is equivalent to

$$(s - r)(b - O^1) = (r - s)(d - O^1) = 2c$$

which can only hold if $s \neq r$ and which in turn implies

(10) $$O^1 = (b + d)/2.$$

43

So except for non-generic cases in which the equality (10) holds, there is no equilibrium where $(\sigma, \tau) \in (0,1)^2$, so equilibria are on the boundary of the unit square. Monotonicity requires $1 - \tau \geq \sigma$, which excludes the strategies with $\sigma + \tau > 1$ (such as the "reverse fully revealing" equilibrium $(\sigma, \tau) = (1,1)$). So we have either one of the three residual corners of the square, or a point in the two sides $\{0\} \times (0,1)$ (under-reporting equilibria) and $(0,1) \times \{0\}$ (over-reporting equilibria). $\quad\square$

B.3.2. *Fully pooling and fully separating equilibria.*

Lemma B.3. *If (6) holds a fully revealing equilibrium exists if and only if:*

$$(11) \qquad\qquad 2c \geq \max\{O^1 - b, d - O^1\}$$

Proof. Suppose that player 1 is following the truthful strategy. At the best response of player 2, he chooses *Out* when he is low type for any signal, and when of high type chooses *In* if and only if the signal is t_0, because:

$$m(\theta_0^1 | t_0) = 1, m(\theta_0^1 | t_1) = 0,$$

where $m(\cdot, t)$ denotes the posterior belief upon observing message t. So the best response of 2 to the truthful strategy of player 1 has $r = 1, s = 0$. To determine the set of parameters for which the fully revealing strategy of player 1 is part of an equilibrium we determine now when this strategy is a best response to $r = 1, s = 0$. With this strategy of 2, type θ_0^1 prefers t_0 to t_1 if and only if $2c \geq O^1 - b$; and type θ_1^1 prefers t_1 to t_0 if and only if $2c \geq d - O^1$. These conditions are equivalent to (11). $\quad\square$

As intuitively clear this equilibrium exists for all costs large enough.

For the high signal pooling we have (if we ignore the case $(a+d)/2 = O^2$):

Lemma B.4. *If (6) holds a pooling equilibrium at the high signal exists if and only if either:*

$$(12) \qquad\qquad (a+d)/2 < O^2 \text{ and } O^1 - b \geq \max\{2c, d - O^1\}$$

or:

$$(13) \qquad\qquad (a+d)/2 > O^2 \text{ and } b - O^1 \geq \max\{2c, O^1 - d\}$$

Proof. Remember that we are considering all Nash equilibria. Take a pooling equilibrium at the high signal. The best response of player 2 is determined by r and s given the posterior belief at t. Note that $m(\theta_0^1 | t_1) = 1/2$, and we can take $m(\theta_0^1 | t_0) \in [0,1]$ since the event t_0 has probability zero in this equilibrium. At the best response given these posteriors,

$$(14) \qquad\qquad r \in [0,1] \text{ and } s \in sign\left(\frac{1}{2} - \frac{O^2 - d}{a - d}\right)$$

Consider now the best response of player 1 to such pairs (r, s). The condition that t_1 is preferred to t_0 by both types of player 1 is equivalent to

$$(15) \qquad\qquad (s - r)(b - O^1) \geq 2c \geq (r - s)(d - O^1).$$

Thus, a pooling equilibrium at high type exists if and only if there is a pair (r, s) that satisfies both (14) and (15). We consider the two cases.

    (1) If $(a+d)/2 < O^2$ then $s = 0$ and (15) is now equivalent to $r(O^1 - b) \geq 2c \geq r(d - O^1)$ which is equivalent to (12);

    (2) If $(a+d)/2 > O^2$ then $s = 1$ and (15) is now equivalent to $(1 - r)(b - O^1) \geq 2c \geq (1 - r)(O^1 - d)$ which is equivalent to (13).

$\square$

For the low signal pooling we have:

Lemma B.5. *If (6) holds a pooling equilibrium at the low signal exists if and only if either:*

(16) $$O^2 \leq (a+d)/2 \text{ and } d - O^1 \geq \max\{2c, O^1 - b\}$$

or:

(17) $$O^2 \geq (a+d)/2 \text{ and } O^1 - d \geq \max\{2c, b - O^1\}$$

Proof. With r, s denoting as usual the probability that the high type Player 2 chooses *In* at t_0 and t_1 respectively, an equilibrium pooling on the low type exists if and only if with $r \in sign((a + d)/2 - O^2), s \in [0, 1]$ the inequality

$$(r - s)(d - O^1) \geq 2c \geq (s - r)(b - O^1)$$

holds. These hold if and only if (16) hold. $\square$

In the following we can then focus on the under and over reporting equilibria; remember that we have excluded by definition the fully pooling or fully separating equilibria from this set.

B.3.3. *Under-reporting equilibria.*

Lemma B.6. *An under-reporting equilibrium exists if and only if the inequalities in 6 and*

(18) $$d - O^1 \geq \max\{2c, O^1 - b\}$$

for player 1 and

(19) $$O^2 \geq (a + d)/2$$

for player 2 hold.

Proof. We already know that at all equilibria, player 2 chooses *Out* at θ_0^2 irrespective of the signal. We check whether an equilibrium exists with $\tau \in (0, 1]$. At θ_1^2 with such a strategy of player 1, player 2 chooses *In* at t in a best response if and only if the posterior $m(\theta_0^1|t) \geq \frac{O^2 - d}{a - d}$. This is never the case if the signal is t_1, and it holds at t_0 when $\frac{1}{1+\tau} \geq \frac{O^2 - d}{a - d}$. So the strategy of player 2 has only one indeterminate value $r \equiv \sigma(\theta_1^2, t_0; In)$ and we know that $r \in sign(\frac{1}{1+\tau} - \frac{O^2 - d}{a - d})$, where the *sign* correspondence is 1, 0 at positive and negative values, and the unit interval at 0.

Our last step is to check when the best response of player 1 to such strategy has the under-reporting form, with $\tau \in (0, 1]$. Since as we have seen player 2 exits at t_1, choosing t_1 gives $O^1 - c$ to the type θ_0^1 and gives O^1 to the type θ_1^1. For a given r, t_0 is better than t_1 for type θ_0^1 if $O^1 - c < (1 - (r/2))O^1 + (r/2)b$, and t^1 is indifferent to t_0 for type θ_1^1 if $O^1 = (1 - (r/2))O^1 + (r/2)d - c$. These two conditions are satisfied if for some $r \in (0, 1)$:

$$r(d - O^1) = 2c > r(O^1 - b)$$

which is equivalent to the additional condition (18). $\square$

The equilibrium is based on the fact that player 1 may be willing to pay the cost of signaling t_0 to hustle player 2 in the tournament to get the payoff d; the gain is $(r/2)(d - O^1)$ and is equal to the cost c. Player 2 at θ_1^2 may choose *In* at the low signal t_0 because may get the high payoff a from θ_0^1 or the lower payoff d from θ_1^1, but overall this is the same as the *Out* payoff O^2.

B.3.4. *Over-reporting equilibria.* A similar analysis yields:

Lemma B.7. *An over-reporting equilibrium exists if and only if the inequalities in 6 and*

(20) $$O^1 - b \geq \max\{2c, d - O^1\}$$

for player 1 and

(21) $$O^2 \leq (a+d)/2$$

for player 2 hold.

Proof. In this case at equilibrium $\sigma^1(\theta_1^1; t_1) = 1$, and $\sigma^1(\theta_0^1; t_1) \equiv \sigma$. The posterior beliefs are $m(\theta_0^1|t_0) = 1$ and $m(\theta_0^1|t_1) = \frac{\sigma}{1+\sigma}$. Player 2 weakly prefers *In* to *Out* if and only if $\frac{\sigma}{1+\sigma} \geq \frac{O^2 - d}{a - d}$, which is equivalent to $\sigma = \frac{O^2 - d}{a - O^2} = 1/\tau$. His strategy is to choose *Out* at θ_0^2 and $\sigma^2(\theta_1^2, t_0; In) = 1; \sigma^2(\theta_1^2, t_1; In) = s$. The variables determining the equilibrium are s and σ; s is constrained to:

$$s \in sign(\frac{\sigma}{1 + \sigma} - \frac{O^2 - d}{a - O^2}),$$

and σ is determined by the best response of player 1. He prefers t_1 to t_0 at θ_1^1 and is indifferent between t_1 and t_0 at θ_0^1 if for some $s \in (0,1)$, $2c > (1-s)(d - O^1)$ and $2c = (1-s)(O^1 - b)$. Together these conditions are equivalent to (20) above. □

To complete the full characterization of the equilibrium set, we examine non-monotonic equilibria.

Lemma B.8. *Non-monotonic equilibria do not exist.*

Proof. An equilibrium is non-monotonic if $\sigma(\theta_1^1; t_1) < \sigma(\theta_0^1; t_1)$, i.e., if $1 - \tau < \sigma$. It is easy to see that the fully revealing non-monotonic equilibrium (both players 1 are always dishonest) does not exist. In that case, $\tau = \sigma = 1$. In that case, a strong player 2 chooses *Out* after t_0 and *In* after t_1. It is easy to verify that player 1 does not deviate at θ_0^1 if $O^1 \leq b - 2c$, and does not deviate at θ_1^1 if $O^1 \geq d + 2c$. This requires $d + 2c \leq b - 2c$ which is incompatible with $b \leq d$ for any $c > 0$.

Consider next the case with $\tau \in (0,1)$ and $\sigma = 1$. A strong player 2 chooses *Out* following t_0 as this can now only come from a strong player 1. If player 1 is indifferent between t_0 and t_1 after θ_1^1, and prefers t_1 after θ_0^1, we must have:

$$s(b - O^1) \geq 2c = s(O^1 - d),$$

but these can never be simultaneously satisfied for $d \geq b$ and $c > 0$. Intuitively, indifference by a strong player 1 requires that he is willing to incur costs c to encourage a strong player 2 to choose *Out*. This requires that he values the outside option more than competing with a strong player 2, so $O^1 > d$. But then a weak player 1 must surely also like player 2 to choose *Out*, since for him the payoff from competing with a strong player 2 is even worse: $b < d$, and would then deviate to t_0.

Consider next the case with $\tau \in 1$ and $\sigma \in (0,1)$. A strong player 2 chooses *In* following t_1 as this can now only come from a weak player 1. If player 1 is indifferent after θ_0^1 and prefers t_0 after θ_1^1, we must have:

$$r(O^1 - d) \geq 2c = r(b - O^1),$$

but this can never be satisfied for $d \geq b$ and $c > 0$. Intuitively, if even a weak player 1 is willing to pay a cost to encourage a strong player 2 to choose *In*, then a strong player 1 also prefers the strong player 2 to choose *In* and would deviate to t_1. □

B.4. Summary. We can summarize the previous sections characterizing the equilibria. A verbal description of the equilibrium set may be helpful. When the cost is high compared to the other payoffs, the only equilibrium is the fully revealing. With smaller cost, the equilibrium correspondence separates into two branches, under and over reporting respectively, depending on whether $(d+b)/2 \geq O^1$ (in which case we have the under-reporting branch, see equation (23)) or the opposite holds. The behavior in these two branches is similar. In the under-reporting, when $O^2 \geq (a+d)/2$ we have a truly under-reporting equilibrium, where the high type reports the low type with positive probability. When O^2 is smaller, the equilibrium becomes a pooling equilibrium. The behavior of the over-reporting branch is similar.

The following theorem gives a complete characterization of the equilibrium.

Theorem B.9. *For generic payoffs, in the interesting case $a \geq O^2 \geq d \geq 0 \geq b$,*

(a) *(**Fully revealing truthful branch**): There is a fully revealing equilibrium if*

$$(22) \qquad 2c \geq \max\{O^1 - b, d - O^1\}$$

(b) *(**Under-reporting branch**): There is an under-reporting equilibrium if and only if:*

$$(23) \qquad d - O^1 \geq \max\{2c, O^1 - b\}$$

$$(24) \qquad O^2 \geq (a+d)/2$$

with strategies $\sigma^1(\theta_0^1; t_0) = 1, \sigma^1(\theta_1^1; t_0) = \tau = \frac{a - O^2}{a - d}$ and

$$\sigma^2(\theta_1^2, t_0; In) = r = \frac{2c}{d - O^1}, \sigma^2(\theta_1^2, t_1; In) = s = 0.$$

There is a pooling equilibrium at the low signal if (i) condition (23) holds and $O^2 \leq (a+d)/2$, sustained by $\sigma^2(\theta_1^2, t_0; In) = 1, \sigma^2(\theta_1^2, t_1; In) = 0$, or (ii) condition (24) holds and $O^1 - d \geq \max\{2c, b - O^1\}$ sustained by $\sigma^2(\theta_1^2, t_0; In) = 0, \sigma^2(\theta_1^2, t_1; In) = 1$.

(c) *(**Over-reporting branch**): There is an over-reporting equilibrium if and only if:*

$$(25) \qquad O^1 - b \geq \max\{2c, d - O^1\}$$

and

$$(26) \qquad O^2 \leq (a+d)/2$$

with strategies $\sigma^1(\theta_0^1, t_1) = \sigma = \frac{O^2 - d}{a - d}, \sigma^1(\theta_1^1, t_0) = 0$ and

$$\sigma^2(\theta_1^2, t_0; In) = r = 1; \sigma^2(\theta_1^2, t_1; In) = s = 1 - \frac{2c}{O^1 - b}.$$

There is a pooling equilibrium at the high signal if (i) condition (25) holds and $O^2 \geq (a+d)/2$, sustained by $\sigma^2(\theta_1^2, t_0; In) = 1, \sigma^2(\theta_1^2, t_1; In) = 0$, or (ii) condition (26) holds and $b - O^1 \geq \max\{2c, O^1 - d\}$ sustained by $\sigma^2(\theta_1^2, t_0; In) = 0, \sigma^2(\theta_1^2, t_1; In) = 1$.

Remark B.10. *Note that some of the pooling equilibria are counterintuitive: they are sustained by the belief that a strong player 2 chooses Out after the low signal but In after the high signal. While these types of pooling equilibria are Perfect Bayesian Equilibria, it is easy to show that they do not survive equilibrium refinements such as "D1" (Cho and Kreps, 1987).*

B.5. The experimental design. We finally consider a game which is closer to the one used in our experimental sessions.

The strategic deterrence treatment. In the strategic deterrence treatment payoffs were symmetric, so the analysis of above applies. Specifically, if we subtract 5 from all payoffs, and make use of the fact that ties are broken by random assignment of the *win* outcome, the payoffs were:

(27)

	θ_0^2	θ_1^2
θ_0^1	$0, 0$	$-5, 5$
θ_1^1	$5, -5$	$0, 0$

and outside options $O^1 = 5$ for player 1 and $O^2 = 0.5$. This is the case $O^1 - b = 10 \geq \max\{2c, d - O^1\} = 2c$ because $d - O^1 = -5$, and $(a + d)/2 = 2.5 > O^2 = 0.5$, hence (provided costs c are sufficiently low) we are in the over reporting branch (see theorem B.9).

The Hustler treatment. In the hustler treatment the payoffs from entering the tournament were not symmetric, and we represent them as follows:

(28)

	θ_0^2	θ_1^2
θ_0^1	e, d	g, f
θ_1^1	a, b	e, d

where $a = 25, b = 0, d = 5, e = 20, f = 10, g = 15$, and $O^1 = 10, O^2 = 5.5$ (again using the fact that ties are broken by random assignment of the *win* outcome). Note that

(29) $$\text{for all } \theta, O^1 < v^1(\theta); f > O^2 > d > b$$

We look for equilibria of the under-reporting form, as in lemma (B.6).

Lemma B.11. *An under-reporting equilibrium of the game with tournament payoffs (28) exists if 30 and*

(30) $$e - O^1 \geq 2c$$

hold.

The proof is a simple computation. The intuition for the equilibrium is clear. Player 2 of the low type prefers *Out* to *In* irrespective of what outcome he is expecting in the tournament. The high type will choose *In* if he gives enough weight to the event that he is facing a low type. At a monotonic under-reporting equilibrium the high signal t_1 reveals that the type is high, so player 1 of low type will not incur the cost of lying when by doing so he could only tempt player 2 to choose *Out*. The high type player 1 may be made indifferent between telling the truth and thus forcing player 2, out or hustling him by stating the low type, paying the cost, and getting with enough probability to reap the benefit of a match with a low type player 2. He can be made indifferent between these two options when the extra gain $r/2(e - O^1)$ can be made equal to the cost $2c$, for some r the probability that the high type player 2 plays *In* after a low signal. This is what condition (30) insures.

Table 3: *Determinants of confidence* (scale 0 – 100)

Dependent var. : $Confidence$	(1)	(2)	(3)	(4)
Number of correct answers	3.50***	3.52***	3.52***	3.32***
	(0.39)	(0.39)	(0.38)	(0.40)
Social	0.29	0.05	0.08	-1.74
	(2.94)	(4.15)	(4.07)	(4.12)
Deterrence (low and high)	3.54	3.06	9.21**	9.26**
	(2.31)	(3.25)	(4.02)	(4.07)
Hustler	-5.55**	1.06	0.57	1.06
	(2.64)	(3.70)	(4.77)	(4.79)
Sender		0.26	0.65	1.02
		(3.31)	(3.27)	(3.26)
Social × Sender		0.47	0.16	1.82
		(5.88)	(5.77)	(5.82)
Deterrence × Sender		0.96	-0.00	-0.57
		(4.59)	(5.43)	(5.48)
Hustler × Sender		-13.21**	-13.21**	-15.18**
		(5.24)	(6.47)	(6.46)
Female			-3.06	-2.57
			(2.76)	(2.78)
Deterrence × Female			-11.00**	-12.30**
			(5.20)	(5.24)
Deterrence × Sender × Female			-0.71	-0.58
			(6.26)	(6.31)
Hustler × Female			1.96	-0.04
			(6.27)	(6.28)
Hustler × Sender × Female			-1.00	0.51
			(7.97)	(8.04)
Familiar with conditional probs.				4.46**
				(1.93)
Constant	33.05***	32.78***	33.76***	31.54***
	(3.75)	(4.16)	(4.22)	(8.97)
Observations	464	464	464	462
R-squared	0.17	0.19	0.23	0.26
Adj. R-squared	0.16	0.18	0.21	0.23

Notes: OLS estimates. Other control variables in model (4) are: familiarity with
Raven test, study category, age, number of siblings, birth order, member of
sports club, entity theory question. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 4: *Determinants of entering*

Dependent var: choice is In	Deterrence (high outside option)			Hustler		
	(1)	(2)	(3)	(4)	(5)	(6)
Own confidence	0.044***	0.015**	0.014*	0.012***	0.007	0.008
	(0.012)	(0.007)	(0.008)	(0.004)	(0.005)	(0.005)
Opponent's confidence	-0.029***			-0.005		
	(0.010)			(0.003)		
Confidence is lower‡		-0.619***	-0.621***		-0.412**	-0.414**
		(0.138)	(0.141)		(0.181)	(0.190)
Female			-0.096			0.040
			(0.203)			(0.149)
Number correct answers			-0.019			0.016
			(0.042)			(0.046)
Risk aversion†			-0.030			0.044
			(0.080)			(0.044)
Observations	48	48	48	48	48	48
Pseudo R-squared	0.59	0.56	0.57	0.26	0.30	0.32

Notes: Probit estimates, reporting marginal effects. ‡Lower confidence is equal to 1 if receiver's confidence is lower than the paired sender's, and 0 otherwise.†Eight missing observations were replaced by the mean in model (3). St. err. in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: *Determinants Updating errors and performance on the Raven task*

Dependent var.: *Number of correct answers*	(1)	(2)
Report error	-1.09**	
	(0.52)	
Machine error		-0.90
		(0.59)
Confidence	0.04***	0.03***
	(0.01)	(0.01)
Constant	6.09***	6.49***
	(0.68)	(0.72)
N	112	112
R-squared	0.11	0.10

Notes: Sample is participants in baseline treatment that received negative feedback. Over-updating and under-updating errors combined. St. err. in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.