TI 2016-012/I
Tinbergen Institute Discussion Paper

Deception and Self-Deception

Peter Schwardmann[1]
Joël van der Weele[2]

[1] *Ludwig Maximilian University of Munich, Germany;*
[2] *Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, the Netherlands.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Deception and Self-Deception

Peter Schwardmann *   Joël van der Weele ◇

February 19, 2016

Abstract

We experimentally investigate the determinants of overconfidence and test the hypothesis, advanced by Robert Trivers, that overconfidence serves to more effectively persuade or deceive others. After performing a cognitively challenging task, half of our subjects are informed about the possibility of earning money by convincing others of their high relative performance in a structured face-to-face interaction. Privately elicited beliefs show that informed participants are 50% more overconfident than those in a control condition, and are less responsive to objective feedback on their performance. Using random variation in confidence generated by our feedback mechanism, we find that increased confidence indeed causes higher evaluations in the ensuing interactions, unless the evaluators have been explicitly instructed to watch out for lies. These results support the idea that confidence is a strategic variable in human interaction.

JEL-codes: C91, D03, D83.
Keywords: Overconfidence, belief formation, self-deception, deception.

*University of Munich. Contact: peter.schwardmann@econ.lmu.de
◇University of Amsterdam, Tinbergen Institute. Contact: vdweele@uva.nl
We would like to thank seminar audiences at the University of Amsterdam, the Toulouse Institute for Advanced Studies, Maastricht University, the University of Munich, and the TIBER symposium in Tilburg for helpful comments.

> “When a person cannot deceive himself the chances are against his being able to deceive other people.”
>
> Mark Twain

1 Introduction

There is ample evidence that the average person thinks he or she is more skillful, more beautiful and kinder than others. Hundreds of studies have found such overconfidence in different domains (Moore and Healy, 2008). At the same time, evidence shows that overconfidence is costly. It leads financial traders to make less money (Barber and Odean, 2001; Biais and Hilton, 2005), CEO’s to initiate more value-destroying mergers (Malmendier and Tate, 2008), and entrepreneurs to invest more often in low value projects (Koellinger *et al.*, 2007; Dawson *et al.*, 2015).

In the face of these costs, an important question is whether there are offsetting benefits that justify the persistence of overconfidence. Social scientists most commonly stress the affective benefits of overconfidence that are associated with holding a good self-image, ameliorating anxiety about an uncertain future, or savouring future successes (e.g. Kunda, 1990; Baumeister, 1998; Brunnermeier and Parker, 2005; Köszegi, 2006; Bénabou, 2013). However, it is unclear why such desires would survive natural or cultural selection. An alternative theory explains overconfidence as a strategy to maintain personal motivation for people who have difficulty in seeing through their plans (Bénabou and Tirole, 2002).

A third theory, which is the focus of this paper, posits that the benefits of overconfidence are social in nature. In a series of papers and a book, Robert Trivers argues that overconfidence makes it easier to convince or deceive others about one’s qualities as it obviates the need for conscious deception (Trivers, 1985, 2011; Von Hippel and Trivers, 2011). Following George Costanza’s maxim “it is not a lie when you believe it”, deceiving yourself can reduce the risk of getting caught by reducing the ‘tells’ associated with lying and reduce the moral costs of lying that have been found to play an important role in communication (Gneezy, 2005; Gibson *et al.*, 2014).

We experimentally test the hypothesis that overconfidence serves to more effectively persuade or deceive others. Participants first perform an intelligence test, and receive a small remuneration depending on their relative performance. Afterwards, a randomly chosen subset of participants is informed of the possibility to earn money by persuading another participant in a face-to-face interaction that they outperformed their peers on the intelligence test. Before this interaction takes place, we elicit private beliefs about relative performance, using monetary incentives for truthful reporting.

Our main hypothesis is that subjects who are informed of a profitable deception opportunity will be more (over)confident about their performance. We find that the shadow of future interactions indeed increases overconfidence by about 50% relative to the control group. We also find an increase in the overconfidence gap between the two groups following unbiased but noisy feedback on performance, as the control group is more responsive to the feedback. We explore the role of lying aversion and the fear of getting caught as drivers of such self-deception, using personality measures and experimental variations in the communication environment.

Our second aim is to test whether overconfidence indeed provides the social benefits conjectured by Trivers. Drawing clear causal inferences about the effects of confidence on social success is impossible in most datasets, due to potential reverse causality or unobserved covariates like beauty or extraversion (e.g. Mobius and Rosenblat, 2006). Our experiment is uniquely designed to solve this problem, as our noisy feedback mechanism generates exogenous variation in confidence. This allows clean estimation of the causal effect of confidence on the outcome of the interaction stage.

We first demonstrate that subjects' communication strategies reflect an implicit appreciation of the benefits of higher confidence, as subjects condition their message on their confidence, with higher confidence leading to higher stated confidence. We then show that confidence indeed raises evaluations in the interaction stage. Specifically, the level of self-deception observed in our treatment group increases the expected payoffs in the interaction stage by 20%, evaluated at the average confidence level.

In a treatment variation in which evaluators follow a short lie-detection tutorial, we find that the beneficial effect of confidence disappears, and actual performance becomes a more important determinant of earnings. This suggests that, naturally, self-deception is most beneficial in interactions where the other side has difficulties to detect true ability. In these settings, our results provide affirmative evidence for the claims of an enormous popular self-help literature that presents confidence as the road to success. While self-help books usually point to correlational evidence that highly successful people are also usually confident, we find that confidence causes success in our stylized interactions.

To our knowledge, this is the first paper to provide clean evidence from fully incentivized elicitation procedures that people self-deceive in pursuit of social advantage, and also the first paper to document clean causal evidence about the advantages of confidence. Understanding why people self-deceive is important to assess the welfare implications of overconfidence. In particular, whereas self-deception's affective benefits may increase welfare, its use to obtain a strategic advantage in a zero sum game may decrease it, and provide a rationale for policy initiatives that seek to curb overconfidence.

This paper proceeds as follows. In the next section we discuss the existing literature and our use of concepts like "overconfidence" and "self-deception" in more detail. Section 3 presents our experimental design, before Section 4 describes our hypotheses. The results section is split in two parts. Section 5 investigates the impact of our main treatment effect on belief formation and the degree of overconfidence. In Section 6 we ask whether overconfidence is in fact effective in deceiving others. Section 7 interprets our results and makes some proposals for future research.

2 Literature and concepts

The first part of our paper addresses the social and strategic origins of overconfidence. As far as we know, there are no behavioral studies that provide direct evidence for the social origins of self-deception, although Von Hippel and Trivers (2011) cite multiple studies showing that people are impressed by confidence in others, and that confidence plays a role in the selection of leaders. A few recent articles provide suggestive evidence for the social origins of overconfidence. Burks *et al.* (2013) find a correlation between overconfidence on a cognitive test and a measure of social dominance in a sample of trainee truck drivers. They argue that the data are consistent with overconfidence as a social bias. Charness *et al.* (2013) find that experimental participants use public statements about their own ability strategically to deter others from engaging in competition with them. Ewers and Zimmermann (2015) show that the desire to impress an audience causes experimental subjects to state higher confidence. Thoma (2015) demonstrates that when others can observe their degree of overconfidence about task performance, men strategically deflate confidence in order to appear more likable. Note that Burks *et al.* (2013) measure private beliefs, but don't vary the strategic context, whereas Charness *et al.* (2013), Ewers and Zimmermann (2015) and Thoma (2015) vary the context of the interaction, but don't measure private beliefs.[1] Our study does both.

Anderson *et al.* (2012) present an extensive study on the relation between status and overconfidence. In a series of experiments, they find that overconfident people enjoy higher status in a group. Moreover, people who are primed with a desire to achieve status, rate themselves higher relative to others on series of skills and abilities relevant for attaining higher status in a business context. Our study goes beyond this by studying the effects of an actual behavioral opportunity for social gain rather than a prime, and by providing monetary incentives for reporting accurate beliefs.[2]

In economics, a small theoretical literature has developed to explain overconfident or optimistic beliefs as stemming from a (subconscious) optimization process. Compte and Postlewaite (2004) model the optimal degree of confidence under the assumption that confidence enhances performance. Brunnermeier and Parker (2005) argue that in the presence of "anticipatory utility", some degree of overoptimism or overconfidence is optimal. Although it will lead to slightly worse decisions, this effect is second order relative to the utility from increased hopefulness. Bodner and Prelec (2002) and Bénabou and Tirole (2011) show how in the presence of uncertainty about one's own preferences, people use actions to influence their own beliefs about their deep characteristics. Bracha and Brown (2012) and Mayraz (2012) provide models where affective influences color risk perceptions, leading to suboptimal decisions. Testing these ideas empirically, Coutts (2015) and Mayraz (2012) find evidence that raising the cost of mistaken beliefs does not reduce overoptimism, a finding that is inconsistent with the theory by Brunnermeier and Parker (2005).

[1] Charness *et al.* (2013) measure private beliefs, but only in one of their conditions.

[2] Anderson *et al.* (2012) write on page 731: "[W]e cannot know with certainty whether overconfident individuals truly believed that they were highly competent, or whether they were merely reporting what they wished to believe."

The second part of our paper relates to the literature on lying and lie-detection. There is a large literature on lie detection in psychology (see Vrij, 2008, for an overview), demonstrating that non-professionals are generally not very good at detecting lies (e.g. Ekman and O'Sullivan, 1991; DePaulo, 1994). Nevertheless, there are some physiological clues that reliably indicate lying, such as pupil dilation (Wang *et al.*, 2010), or fake smiles (Ekman *et al.*, 1988). Belot and Van de Ven (2016) criticize the psychological literature for lack of incentives for deceit and detection, and the lack of contextual richness of the interaction. They conduct face to face interviews between buyers and sellers and show that buyers are able to detect fraudulous sellers with a rate that is above chance, while contextual richness does not seem to matter much in this setting. Our paper uses a similar face to face interaction, but with a highly structured verbal message.

Finally, we make some remarks on terminology. In the remainder of this paper, "overconfidence" is defined as the overestimation of one's ability relative to others, or "overplacement" in the terminology of Moore and Healy (2008). Using belief measurements that incentivize truthful reporting, several recent studies have documented such overplacement when it comes to ego-related variables like beauty or intelligence (Eil and Rao, 2011; Grossman and Owens, 2012; Möbius *et al.*, 2014), where the latter is also the focus of this paper.[3]

Throughout this paper we use "self-deception" to describe strategic shifts in confidence.[4] The use of the term "self-deception"has been criticized by Kurzban (2011) and Kahan (2016), amongst others. Kurzban argues that in a modular conception of the brain there is little space for a "self" that is being deceived, while Kahan points out that treating motivated cognition as a "bias" obscures the rationality that lies behind motivated belief structures. We do not take a strong position in this debate, but we think that "self-deception" is a useful concept for our purposes, as it captures both the strategic nature of beliefs reported by an individual (the "deception") and the fact that these beliefs refer to characteristics of that same individual (the "self").

3 Experimental Design

Our experiment was programmed in z-tree (Fischbacher, 2007) and run over the course of a week in March 2015 at the Munich Experimental Laboratory for Economic and Social Sciences. 288 subjects participated in 18 sessions of exactly 16 subjects each. The experiment lasted slightly over 1 hour and the average subject earned 16.45 Euros (min. 4 Euros, max. 28.5 Euros).

The experiment had two parts, that we will refer to as the "self-deception stage" and the "deception stage". In the self-deception stage, we investigated whether the announcement of a profitable deception opportunity had an influence on beliefs about their own performance on an

[3] While Benoît and Dubra (2011) show that overconfidence may result from a rational or Bayesian interpretation of available evidence, Burks *et al.* (2013) demonstrates that not all aspects of overconfidence can be explained by such a model.

[4] Note that our paper does not satisfy the criterion for self-deception posed in Gur and Sackeim (1979), as we do not show that subjects hold two contradicting beliefs at the same time. Yet, the use of monetary incentives for truthful reporting of beliefs in own performance shows that strategic considerations can lead subjects to firmly hold a systematically inflated belief about themselves, which we consider evidence for self-deception.

	Employers only	All subjects	Contestants only
		• Training in belief elicitation	
Self-Deception Stage		• Raven intelligence task • Elicitation of prior belief • Noisy performance signal • Elicitation of posterior belief	• Information about interview • Warning about lie detection tutorial (TW only)
Deception Stage	• Lie detection tutorial (TW and TnW only) • Elicitation of evaluations	• Instructions for interaction • Face to face interaction	• Lying aversion elicitation
		• Profit announcements • Questionnaire	

Table 1: Timeline of tasks and information by role.

intelligence test. To this end, we separated the group into "contestants" who were informed about the later deception opportunity, and a "control" group that was not. In the deception part, we investigated whether (over)confidence made the contestants more persuasive, as predicted by the theory. To this end, contestants competed to persuade "employers" of their performance of the intelligence test in a face to face interaction. Employers were the same participants who constituted the control group in the self-deception part of the experiment. They could earn money by giving accurate assessments of contestants test results. Contestants could earn money by obtaining high assessments. The sequencing of experimental tasks is depicted in Figure 3. We now explain the design of the experiment in more detail.

3.1 Self-deception stage

After coming into the lab, participants were introduced to the belief elicitation mechanism they would face later. The mechanism consisted of a variation of the Becker-DeGroot-Marshak procedure, known as "matching probabilities" or "reservation probabilities". Under this mechanism participants indicate which probability p makes them indifferent between winning a monetary prize with probability p, and winning the same prize when an uncertain event E occurs. After participants indicated the point of indifference in a list of probabilities, one probability was randomly drawn from the list and participants were awarded their preferred lottery for that probability. Under this mechanism, reporting the true subjective probability of E maximizes expected value, regardless of risk preferences (see Schlag *et al.*, 2015). We explained the procedure as well as the fact that truthful reporting maximizes expected value. This stage took about 15 minutes, and we proceeded only once all subjects correctly answered four control questions about the belief elicitation.

Participants were then divided in anonymous groups of four and proceeded to the intelligence task. The task consisted of 15 Raven matrices of varying difficulty, and participants had 10 minutes to solve as many as they could. Participants obtained 2 points for each correct answer and lost 1 point for each incorrect or omitted answer. The subjects with the two top scores in their anonymous group of four earned 2 Euros. Their earnings, and hence their ranking within the group, were only communicated to them at the end of the experiment.

We administered our main treatment after the intelligence task. Of the four groups of four in each session, two groups were given the role of "contestants" and two groups were designated as controls. The control group was not told anything about the interview stage at this point, while contestants were given the following information about the interview[5]:

- *In the next part of the experiment you will take the role of contestant.*
- *You will conduct face to face interviews with 4 subjects in the role of employers, who, after interviewing every member of your group, are paid for correctly identifying whether your test score was amongst the best 2 of your group.*
- *You are paid for one of your interviews that is selected at random*
- ***If the employer in this interview ranks you as being amongst the top 2 of your group, you will receive an additional 15 Euros.***
- *Before we move on to stage 2, we ask you to state your belief about how likely it is that you are in the top 2 of your group. Your statement is strictly confidential and will never be shown to another subject.*

While the difference in information about the interview constitutes our main treatment effect, we performed two further treatments. First, two thirds of the employers, were given a short tutorial in lie-detection during the deception stage of the experiment (see below). Second, half of the contestants facing employers who did the tutorial (so one third of overall contestants), were "warned" *before the belief elicitation* that their employer had received a "lie-detection tutorial". These treatment variations were designed to vary the interaction environment as well as the pressure on contestants to appear credible even when lying.

After contestants received information about the interview stage, participants in both treatment and control group were asked to submit their beliefs using the incentive compatible mechanism

[5] We left some ambiguity about the exact nature of the interview interaction, so that candidates would not, at this stage, have the impression that the experimenter could compare the submitted beliefs with her interview statements.

explained above. The event about which beliefs were elicited was “the probability that you are amongst the Top 2 performers in your group of four”. In the remainder of the paper, we will refer to this event as “Top 2”. The prize in the elicitation stage was 3 euros.

After participants submitted their beliefs, we gave them noisy feedback on their performance. Participants were told that they would be shown a ball drawn from one of two virtual urns containing 20 balls of different colors. If their performance was actually in the Top 2, the ball would come from an urn with 15 black balls and 5 red balls. If their performance was not in the Top 2, the ball would come from an urn with 15 red balls and 5 black balls. Thus, a black ball constituted “good news” about their performance, where the likelihood ratio was $0.75/0.25 = 3$. We will use the fact that the feedback had a random component to assess the impact of variation in confidence in the second stage. After subjects had observed the ball, they reported their belief about being in Top2 for a second time, which concluded the self-deception stage of the experiment.

3.2 Deception stage

The 16 subjects in a session were divided into two groups, each consisting of 4 contestants and 4 “employers”, where the latter were part of the control group in the first stage of the experiment. Sessions were equally divided between three experimental conditions, which differed in the tutorial for lie detection. In condition “No Tutorial” (NT), employers were not trained in lie detection. In condition “Tutorial no Warnings” (TnW), employers were trained in lie detection but contestants were not warned about this. In condition “Tutorial and Warning” (TW), employers were trained and contestants warned about this training.

In treatments TW and TnW, employers proceeded to the lie detection tutorial, which took several minutes. On the computer screen they observed four tips on how to recognize “tells” associated with lying, namely fidgeting, face-touching, fast breathing and incongruent facial expressions. In addition, the tutorial explained that avoidance of eye contact is an unreliable indicator of lying, as it is easily corrected by an experienced liar.

Before the interviews, employers were given a paper evaluation sheet for each contestant on which to write down the contestant’s message and their evaluations of the contestants relative performance, honesty, confidence, likability and attractiveness. The interviews followed a speed dating protocol. Employers left their computer stations and positioned themselves in front of the contestants. There were four rounds of interviews so that every employer would get to interview each of the four contestants in the same group. Interviews in any round took place behind partitions to assure a minimal level of privacy.

On the ring of a bell, contestants said one sentence: “I believe that my performance was in the top 2 of my group with ... percent probability”. In the blank, the contestants verbally filled in a number between 0 and 100. In the remainder, we will refer to this number as the contestant’s “message”. During the interviews, none of the 144 contestants said anything more than this sentence. After the sentence was said there were a few seconds in which employers could scrutinize contestants faces and body language, before the bell rang again to mark the end of a round. Employers were given time to fill in their evaluations, before moving to the next contestant. The reason we kept the interaction minimalistic is to keep control over the interview stage as much as possible. This allows us to exploit instrumental variable techniques for contestant's confidence, as we will explain below.

After the four rounds of interviews, employers returned to their computer stations and entered their evaluations into the computer. Meanwhile, contestants participated in a task that elicited their degree of lying aversion, based on Gibson *et al.* (2014). During this task, participants are asked to imagine themselves in the position of a company CEO, who can earn money by deceiving shareholders about the companies' value. Using a Becker-DeGroot-Marshak mechanism, participants are asked to indicate repeatedly whether they would report deceptively or not, were the amount of money they could earn from doing so was increased in four steps from 0 cents to 120 cents. Afterwards, all subjects filled out a questionnaire with background characteristics as well as the "Assertiveness" scale from the Big 5 personality questionnaire.

4 Hypotheses

Our main hypothesis is based on Von Hippel and Trivers (2011). It postulates that confidence makes deception easier, and should primarily occur in situations with deception opportunities. We therefore expect contestants to engage in more self-deception and become more overconfident than the control group. Overconfidence is measured on the sample level, as the difference between the average belief that a subject is in the top 2 of a group of 4 and 50%, the true average probability that someone is in the top 2.

Hypothesis 1 *Contestants are on average more overconfident than employers.*[6]

Our next hypothesis relates to the possible reasons for strategic self-deception. Let us define a lie as the difference between stated confidence and privately held beliefs. Suppose larger lies have a higher probability of being detected because contestants' cognitive costs and give-away cues are increasing in the size of a lie. Then overconfidence may be useful, because it allows contestants to state higher confidence levels, while keeping the size of the lie constant, or, alternatively, it enables them to state the same confidence while lying less. Here we make the implicit assumption (which is vindicated by our data) that contestants benefit from higher stated confidence because receivers put some weight on contestants' messages. As this channel relates to the cognitive technology behind deception, we call this the "technological motive" for self-deception. We conjecture that its relative importance is increasing in the probability that a lie is detected and hence, that contestants who are warned about the employer's lie detection tutorial should become more overconfident.

[6]Note that this hypothesis constitute a stronger test than the theory requires. The theory maintains that natural selection has produced the capacity to self-deceive because of its social advantages. This does not necessarily imply that this capacity will always, or only, be used when deception possibilities are present. It is compatible with the idea that self-deceivers arrive at overconfidence through heuristics that apply even when the strategic situation does not ask for it.

Even if lies are never detected, there may be moral costs from deceiving another person (Gneezy, 2005). A (subconscious) desire to avoid such costs gives a "moral motive" for self-deception (McKay *et al.*, 2011). For example, Di Tella and Pérez-Truglia (2015) shows that people will negatively bias their beliefs about other people in order to avoid feeling obliged to donate money to them. We therefore hypothesize that more lying-averse individuals, as measured by the task in Gibson *et al.* (2014), have a higher incentive to self-deceive.[7]

Hypothesis 2 *Contestants are more overconfident if*

a) they are warned about the employer's lie-detection tutorial, and

b) they are more lying averse.

If contestants do indeed become more overconfident, it is important to understand how they respond to unbiased feedback. Studies focusing on affective motives for self-deception have found that people update differently when it comes to ego-related variables (Grossman and Owens, 2012; Eil and Rao, 2011). Möbius *et al.* (2014) show that subjects attempt to maintain their overconfidence by updating "asymmetrically" (putting relatively more weight on good signals) and "conservatively" (being less responsive to both good and bad signals than Bayesian updating prescribes). We conjecture that our contestants, who have an interest in maintaining optimistic beliefs, update asymmetrically and conservatively to a greater degree than the control group.

Hypothesis 3 *Contestants are less responsive to information and place relatively higher weight on good signals than employers.*

Burks *et al.* (2013), find that socially "dominant" individuals, as measured by the assertiveness scale on the Big 5 questionnaire, exhibit more overconfidence. They go on to conjecture that this is because socially dominant individuals attribute more importance to the belief of others about their ability. Following this logic, dominant individuals should exhibit a greater treatment effect in our experiment, since contestants, unlike our controls, expect to be exposed to an audience.

Hypothesis 4 *Socially dominant individuals are more overconfident, and being a contestant increases overconfidence more if an individual is socially dominant.*

The final set of hypotheses relates to the effect of overconfidence on persuasiveness. Whether contestants find it profitable to self-deceive depends on their subjective beliefs about whether overconfidence pays. A good indication of whether overconfidence if perceived as an asset is whether contestants condition their communication strategies on their confidence. In particular, we expect higher private beliefs to result in higher stated confidence.

Hypothesis 5 *More confident contestants report a higher likelihood of being in the Top 2 to the employer.*

[7]Whether this translates into higher degrees of overconfidence depends on whether lying aversion extends also to lying to oneself. However, since self-deception seems a largely unconscious process, unlike guilt from lying to others, we conjecture that higher lying costs increase overconfidence.

Overconfidence is useful to a contestant if additional confidence helps to convince the employer that the contestant did well, as measured by the probability estimate of the employer that the contestant was in the top 2 of her group of 4. There are two reasons why contestants may value higher evaluations: pecuniary benefits and preferences over the beliefs of others about one's performance. Distinguishing between these two benefits is beyond the scope of this paper. But regardless of the source of utility, whether confidence ultimately pays and provides an advantage to the self-deceiver is one of the core elements of Trivers' hypothesis .

Hypothesis 6 *For a given level of performance, more confident contestants will obtain better evaluations.*

5 Results of the self-deception stage

5.1 Do deception opportunities increase confidence?

Our first and main hypothesis is that the possibility of deception will lead to greater overconfidence, as measured by the difference between average beliefs and 50%, the actual probability that the average subject is in the top half. We first look at subject's beliefs before they received feedback, i.e. "Prior beliefs", and then turn to post-feedback "Posterior beliefs".

Figure 1 shows that prior overconfidence amongst contestants (12.2 ppt) is about 50% higher than amongst subjects in the control group (8.3 ppt). Put differently, roughly one third of contestants' overconfidence can be attributed to a strategic motive. The difference in beliefs is significant at the 5% level ($p = 0.049$, Mann-Whitney U test).[8]

Table 2 presents a series of OLS regressions investigating the determinants of prior beliefs. The first column shows that the treatment effect is significant at the 5% level, controlling for individual test scores. In the second column, we restrict our dataset to contestants, and look at the effects of a warning that employers received a lie detection tutorial as well as the impact of lying aversion. The latter variable is a dummy that takes the value of 1 if the the participant's aversion to lying is higher than the median, and zero otherwise. In line with Hypothesis 2, both coefficients are positive. However, neither is close to statistical significance.

The third column confirms the finding in Burks *et al.* (2013) that socially dominant individuals are more overconfident.[9] Here, "Dominant" is a dummy variable that takes a value of 1 if the dominance score on the personality scale is higher than the median. The coefficient is similar in size to that of being a contestant, but only significant at the 10% level. In the final column, we investigate the hypothesis that dominant individuals exhibit a more pronounced treatment effect. The insignificant and negative interaction term allows us to reject Hypothesis 4 and Burks *et al.* (2013)'s conjecture that dominant individuals are more driven to impress others. Instead, this

[8]While present our results in terms of overconfidence to highlight the direction of the bias in average beliefs, confidence intervals and test statistics are computed based on individual measures confidence.

[9]For this regression we have 16 fewer observations than in Column 1, because a technical glitch caused us to loose the data of one session's questionnaire.

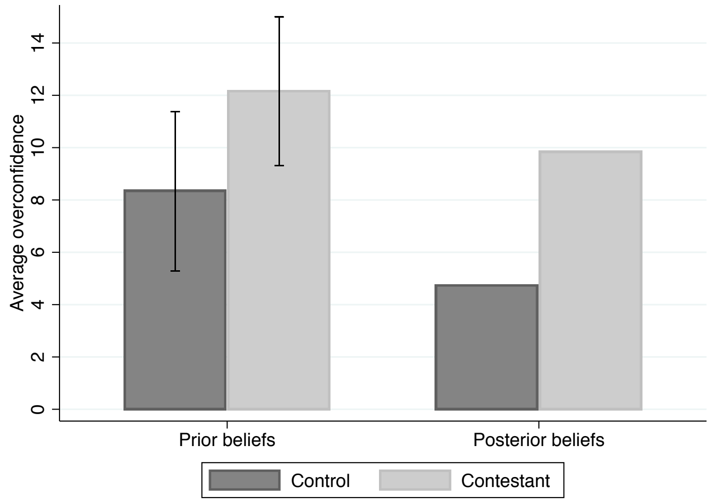


Figure 1: Average overconfidence by experimental role, as measured by the average stated probability that the contestant is in the top half of performers, minus 50%. Bars represent 95% confidence intervals for our confidence measure, and are omitted for posterior beliefs, which are not normally distributed.

suggests that social dominance reflects a preference over one's own rather than others' beliefs about one's relative standing. This point receives corroboration in Section 6.2, where we find that the dominance dummy is not correlated with stated confidence, once we control for private beliefs.

5.2 Does confidence converge after feedback?

Next we investigate the effect of unbiased feedback on overconfidence in control and treatment group. Figure 1 shows that after an unbiased noisy signal overconfidence goes down in both groups. However, contestants' post-feedback overconfidence (9.7 ppt) is now more than twice as high as the control group's (4.7 ppt). Due to the binary nature of the feedback signal, the distribution of posterior beliefs becomes quite dispersed, and the difference in means of our confidence measure is only significant at the 10% level ($p = 0.066$, Mann-Whitney U test).[10]

Figure 2 explains why posterior beliefs in the different conditions do not converge. The bars show the average shortfall of the belief update relative to the rational Bayesian update, where the latter is computed for each participant on the basis of the prior belief and the strength of the noisy signal. After every signal, participants update less than Bayesianism prescribes, so both control and treatment group exhibit conservatism. In line with Hypothesis 3, conservatism is higher in the contestant's sample, with marginal statistical significance ($p = 0.099$, two-sided MWU). Contestants are therefore slightly less responsive to information than the control group, which allows them to better preserve their optimistic beliefs. When we separately graph the belief

[10] Appendix A shows the distribution of posteriors. Posteriors are clearly not normally distributed. Therefore, we omit confidence intervals in Figure 1, which would seem artificially small.

	(1)	(2)	(3)	(4)
Score	0.993***	0.954***	0.987***	0.985***
	(0.111)	(0.174)	(0.119)	(0.119)
Contestant (d)	3.737**		4.401**	5.265*
	(1.885)		(1.972)	(2.861)
Warned (d)		1.843		
		(2.867)		
Lie av. (d)		1.917		
		(2.717)		
Dominant (d)			3.937*	4.770
			(2.017)	(2.929)
Contestant * Dominant (d)				-1.641
				(3.951)
Constant	45.47***	48.05***	43.42***	42.94***
	(2.053)	(3.206)	(2.246)	(2.598)
Observations	288	144	272	272
R^2	0.211	0.192	0.222	0.223

Table 2: OLS regressions of confidence on task scores and treatment dummies. “Dominant” and “Lie av.” are dummy variables that takes the value 1 if the personality score is above the median. Robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

updates following positive and negative signals, we observe the same difference, although it is more pronounced for positive signals ($p = 0.0264$, two-sided MWU) than negative signals ($p = 0.635$, two-sided MWU). This latter result is not in line with the part of Hypothesis 3 that states that contestants place relatively more weight on positive signals. Note however, that for both the contestant and control group individually, average conservatism is not significantly different between good and bad signals.[11]

A potential caveat to these results is that differences in updating reflect differences in priors for the two groups, or accidental differences in the proportion of good and bad signals received, which, based on the experimental design, should be equal across groups. To check this we run logistic regressions, i.e. linearized versions of Bayes’ rule, which allow us to control for the prior belief and the signal’s direction (see Möbius *et al.* (2014)). The results, reported in Appendix B, show that contestants are indeed more conservative than the control group. Over the whole sample, we do not replicate the result by Möbius *et al.* (2014) that participants update asymmetrically. In fact, we find the opposite, as participants put slightly more weight on low signals, a tendency that is somewhat more pronounced for contestants. When we replicate Table 2 for the posterior, we find

[11] Part of the difference in conservatism between the two groups is driven by subjects who update irregularly. There is a small number (17, or 6%) of participants who update in the wrong direction, and a larger number (65, or 22.6%) who don’t update at all. Combined, these “irregular” updates are more frequent amongst the contestants than amongst the control group ($p = 0.089$, 2-sided Fisher Exact test), driving part, but not all, of the finding that contestants are more conservative. Note that the failure to update may itself be motivated and consistent with our hypotheses.

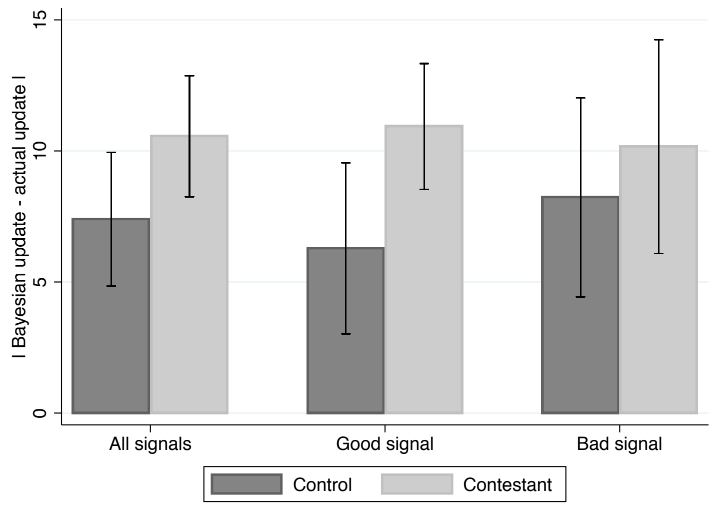


Figure 2: Conservatism by information condition and signal. Conservatism is measured as the average of the absolute value of the Bayesian update minus the actual update.

that the quantitative effects of both being warned of the employers' tutorial and being lying averse on overconfidence have increased, with the latter almost achieving significance.

Summary 1 *Participants who are informed about a profitable deception opportunity are initially about 50% more overconfident about their performance on an intelligence task. The provision of feedback does reduce overconfidence across the board, but contestants are less responsive to feedback than the control group. After the self-deception stage of the experiment, contestants are on average more than twice as overconfident as the control group.*

6 Results of the deception stage

In the previous section we showed that the shadow of profitable deception opportunities increases overconfidence amongst participants. We now investigate how confidence influences contestants' communication strategies and whether higher confidence pays off by raising the probability that a contestant gets hired.

6.1 Identification strategy

We cannot rely on mere correlations to establish a causal relationship between contestant confidence and the messages sent by the contestant or the evaluations by the employer. Such correlations may be spurious as contestants' confidence may correlate with many unobserved determinants of contestants' messages and employers' evaluations. For example, more beautiful people may be more confident, experience positive discrimination in evaluations (Mobius and Rosenblat, 2006) and, other things equal, be less likely to lie because they are more able to achieve favorable social outcomes without inflating their stated confidence.

Instead, we use the random, exogenous variation in contestants' confidence resulting from the noise component of the feedback signal. This variation can not affect evaluations in any other way than through confidence, making it a valid instrument. As far as we are aware, the exogenous variation inherent in our signal allows us to provide the first unambiguously causal evidence linking confidence to evaluations in a competitive social environment.

In the first stage of the 2SLS regressions that follow, we regress contestants' posterior beliefs about their relative performance on a dummy variable that takes a value of 1 if the subject saw a black ball (the favorable signal) and a dummy that takes a value of 1 if the contestant was in fact in the top 2. We assure that our instrument is valid by controlling for actual performance. First stage regressions (which we omit to save space) feature an instrument that is significant at the 0.5 percent level and R-squared values of 0.3. The first stage F-statistic is reported in our second stage regression tables.

In the following analysis, we distinguish between three conditions: the NT condition where employers were not trained in lie-detection and contestants were not warned, the TnW condition where there was training but no warning, and the TW condition in which there was training and a warning to contestants.

6.2 Do more confident contestants send higher messages?

A prerequisite for a social advantage of self-deception is that increased confidence is used to communicate a high performance more convincingly to the employer. Thus, contestants should increase the verbal estimate of their performance to the employer ("the message") if this raises their evaluation, a fact that we will verify below. Alternatively, subjects may choose to more convincingly communicate a message of a given size, in which case we would expect to see a lower probability of lying when confidence increases (this strategy is explored in Appendix C). The relative attraction of these two strategies depends on how both persuasion and lying costs are influenced by the size of a lie, an interesting but difficult question which is beyond the scope of this paper.

Table 3 shows the result of two stage least square regressions of the message on the posteriors, instrumented by the noisy signal, and a series of control variables. Column 1 indicates that over all conditions, a 1ppt increase in posterior beliefs increases the message by 0.31 ppt. Interestingly, receiving a warning increases the message by about 5ppt. This is a large difference, that we explore in more detail by running separate regressions for unwarned (Column 2) and warned subjects (Column 3). Since unwarned contestants cannot ascertain whether or not employers were trained, observations from the NT and TnW conditions can be pooled in Column 2.

One plausible interpretation for higher messages in the TW condition is that the warning causes the participants to perceive the interview as a "game", in which lying is seen as an optimal strategy rather than a norm violation. This interpretation is supported by the fact that subjects condition their message on their beliefs when they are unwarned but not when they are warned. In addition,

Condition	All contestants	NT & TnW	TW
	(1)	(2)	(3)
Posterior (Signal)	0.308***	0.486***	0.053
	(0.10)	(0.14)	(0.15)
Top 2	-1.280	-3.148	-0.312
	(2.53)	(3.67)	(3.40)
Lying aversion (d)	-2.013	-5.768*	2.161
	(2.29)	(3.24)	(2.91)
Dominant (d)	0.702	-1.214	1.240
	(1.98)	(2.87)	(2.92)
Female	1.656	1.446	1.037
	(2.41)	(3.51)	(2.68)
Round	0.243	0.283	0.169
	(0.20)	(0.28)	(0.25)
Warned (d)	4.630**		
	(2.09)		
Constant	55.078***	48.557***	73.295***
	(6.21)	(8.24)	(8.58)
First-stage F-statistic	32.42	25.91	12.93
Observations	544	352	192

Table 3: Determinants of contestant message. IV regressions with Posterior instrumented by Black Ball, standard errors (in parentheses) are clustered at the subject level, $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

higher lying aversion decreases messages for unwarned subjects, but not for warned subjects. In general, the large difference in communication strategy between warned and unwarned subjects suggests that unwarned subjects have moral constraints of lying. Since contestants who were warned do not seem to send higher messages when they have higher beliefs, the strategic benefits of having overconfidence are likely to be limited in this condition. It is then not surprising that the warning failed to significantly increase beliefs in the first place.

Summary 2 *Unwarned subjects condition their message on confidence. We find suggestive evidence that moral concerns constrain their degree of lying. Warned subjects generally choose high messages, independent of their confidence.*

These results suggest that contestants are aware that a) higher messages are strategically advantageous, and b) such messages are constrained by confidence for unwarned subjects. Together, this provides a motive for the self-deception that we observed in the first stage of the experiment.

Condition	NT		TnW		TW	
	(1)	(2)	(3)	(4)	(5)	(6)
Posterior (signal)	0.600**	0.553*	-0.065	-0.272	-0.249	-0.269*
	(0.25)	(0.30)	(0.21)	(0.18)	(0.16)	(0.15)
Top 2	-8.492	-7.977	16.068**	14.446***	4.759	4.712
	(5.07)	(5.34)	(6.02)	(4.81)	(3.71)	(3.61)
Female	-8.507*	-8.588**	3.269	1.699	3.507	3.179
	(4.31)	(4.16)	(3.47)	(3.37)	(4.02)	(3.95)
Round	0.312	0.303	-2.487**	-2.675**	0.958	0.886
	(1.27)	(1.26)	(1.12)	(1.14)	(1.33)	(1.33)
Message		0.130		0.515***		0.426**
		(0.21)		(0.11)		(0.17)
Constant	22.664	15.621	49.722***	26.801***	58.239***	26.228**
	(15.39)	(9.71)	(10.43)	(8.20)	(9.54)	(9.98)
First-stage F-Stat	14.00	20.39	28.98	24.14	18.93	22.52
Observations	192	192	192	192	192	192

Table 4: Determinants of employer's evaluations of candidates. IV regressions with Posterior instrumented by Black Ball, standard errors (in parentheses) are clustered at the subject level, ${}^{*}p < 0.10, {}^{**}p < 0.05, {}^{***}p < 0.01$.

6.3 Does higher confidence cause better evaluations?

There are two ways in which higher contestant posteriors could lead to higher employer evaluations. First, higher confidence may lead contestants to send higher messages. If employers do not discount messages appropriately this may lead to higher evaluations. Second, more confident individuals may be able to more convincingly deliver a message of a given size, e.g. by avoiding suspicious non-verbal cues. While we can pin down the effect of posteriors on evaluations, we will not be able to disentangle the relative importance of these channels with any certainty, because we lack independent variation between the message and the posterior.[12] When we report regressions that control for the message a contestant sent, these should therefore be interpreted with caution.

Table 4 shows the determinants of employer evaluations in the NT, TnW and TW conditions respectively. Column 1 tells us that higher confidence resulted in a large increase in evaluations when employers did not participate in the lie detection tutorial (NT condition). In addition, untrained employers appear to discriminate against women. Under the above caveats, Column 2 adds contestants' messages to the set of regressors and suggests that messages do not have a significant effect on evaluations in the NT condition.

Contrary to what we expected, Columns 3 and 4 reveal that employers who were trained in lie

[12]We implemented a treatment designed to generate such variation by inducing an exogenous shift in messages. We informed half of the contestants that higher messages lead to better payments on average, information that we based on two pilot sessions for our experiment that are not reported in this paper. However, this treatment failed to influence contestants' messages. Since we did not have any independent hypotheses about this treatment, it is not included in the analysis.

detection and faced contestants who were not warned, were not responsive to contestants' confidence. Instead, evaluations are positively correlated with a contestants actually being in the top 2, i.e. her objective relative ability. These results are robust to using the contestants absolute scores in the quiz as a measure of ability and to including a host of other controls which we omitted here to save space, such as the messages, beliefs and ability of all contestants an employer sees.

The results indicate that employers became better at detecting true ability. This effect is surprisingly large given the minimalistic nature of the tutorial. We conjecture that the tutorial inadvertently made employers better at spotting actual top performers because the characteristics the tutorial associates with truthful individuals are more likely to be exhibited by high ability individuals. Given that employers see all four contestants and can compare contestants appearances, messages and body language, they have access to a lot of information that could potentially help them get at who the best performers in the quiz are. The tutorial may have helped employers to make use of this information more effectively.

When contestants were warned about the employers' tutorial, higher confidence did again not lead to better evaluations. In fact, negative coefficients in Columns 5 and 6 suggests that it may backfire. Furthermore, when we reduced the informational advantage of employers by warning contestants, this lowered employers' ability to gage contestants true ability. Finally, whether candidates were warned or not, the trained employers in Columns 4 and 6 seem to put more stock in the contestants' messages. Higher-order polynomials of posterior and message do not have significant coefficients if we include them in the regressions of Table 4. Results are also robust to including employers' beliefs about their own performance.

Summary 3 *Controlling for actual ability, an increase in confidence by 10ppt increases evaluations with about 6ppt. This is not longer the case when employers are trained, in which case true ability is a more important determinant of evaluations. When contestants are warned, neither true ability nor confidence has a significant effect on evaluations.*

Our data are somewhat suggestive of an "arms race" between employers and contestants. As long as employers have no training, they can be fooled by overconfidence of contestants. As soon as employers are aware of potentially deceptive strategies, they become better at spotting true ability. When warnings to contestants give away the employer's training, neither side is able to gain an advantage over the other.

6.4 Does self-deception pay?

To understand whether self-deception pays, we investigate whether increased confidence leads to a higher chance of getting "hired" by employers. Once again, we use the exogenous variation in confidence to provide clean evidence for this question. Table 5 shows the result of probit regressions where the posterior is instrumented by the signal. The table reports marginal effects, evaluated at the average participant's confidence level.

Condition	NT (1)	TnW (2)	TW (3)	NT&TnW (4)
Posterior	0.027**	0.005	-0.012	0.019**
	(0.01)	(0.01)	(0.01)	(0.01)
Top 2	-0.420	0.781**	0.178	0.045
	(0.32)	(0.33)	(0.26)	(0.23)
Female	-0.401	0.254	0.059	-0.206
	(0.27)	(0.21)	(0.24)	(0.16)
Round	-0.029	-0.210***	-0.052	-0.110**
	(0.06)	(0.07)	(0.08)	(0.05)
Constant	-1.068	-0.331	0.764	-0.725
	(0.83)	(0.57)	(0.59)	(0.48)
Observations	192	192	192	384

Table 5: Determinants of employer's hiring decision (a dummy that takes a value of 1 if a contestant is hired). IV probit regressions with Posterior instrumented by Black Ball, standard errors (in parentheses) are clustered at the subject level, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Column 1 of Table 5 shows that the marginal effect of higher confidence on the probability of being hired is positive in the case of untrained employers. An increase in the posterior by 1 percentage point raises the probability of being hired by 2.7 percent. Given that the average increase in confidence amongst contestants relative to the control group was about 4ppt, this means that the average contestant increased her probability of getting hired by roughly 10% as a consequence of self-deception. This increases expected payoffs by about 0.1*15=1.5 euros, or 20% of expected interview payoffs (7.50 euros). Thus, if the employer has not followed a lie detection training, self-deception is quite profitable.

However, in line with the results on employers' evaluations, this effect disappears when employers have followed a lie-detection tutorial (Column 2). In that case, it is the actual performance that determines the probability of being hired. When employers were trained and contestants were warned (Column 3), neither true ability nor confidence have an effect on the hiring decision. Column 4 pools observations for all unwarned subjects, regardless of employer training. Confidence, but not actual performance is profitable. This is again in line with Hypothesis 6 and shows that the benefits of self-deception holds against both trained and untrained employers, as long as contestants are not warned.

Summary 4 *An increase in confidence by 4ppt, our average treatment effect, increases the chance of getting hired for the average contestant by about 10ppt and raises expected payoffs of the interaction stage by 20%. This effect disappears when employers are trained in lie detection.*

7 Discussion and conclusion

The results of our experiment demonstrate that the strategic environment matters for cognition about the self. We observe that deception opportunities increase average overconfidence relative to others, and that, under the right circumstances, increased confidence can pay off. Our data thus support the the idea that overconfidence is strategically employed for social gain.

Our results do not allow for decisive statements about the exact cognitive channels underlying such self-deception. While we find some indications that an aversion to lying increases overconfidence, the evidence is underwhelming.[13] When it comes to the ability to deceive others, we find that even when we control for the message, confidence leads to higher evaluations in some conditions. This is consistent with the idea that self-deception improves the deception technology of contestants, possibly by eliminating non-verbal give-away cues.

However, we are unable to say with certainty whether or not this technology motive causes contestants to self-deceive in the first place. Warnings about the employer's training in lie-detection do not induce additional self-deception. But this may be because the warnings also reduce moral objections to lying, rather than because the technology motive is absent. To see this, recall that, unlike their unsuspecting counterparts, warned contestants do not condition their messages on their confidence and their lying aversion. The moral and technology motives for self-deception appear to be intertwined and isolating their marginal influence on overconfidence will require a subtle empirical strategy.

Our results have ramifications in the fields of both economics and psychology. In economics, we provide new evidence to a debate about the sources of overconfidence, for example among entrepreneurs. While Koellinger *et al.* (2007) and Dawson *et al.* (2015) provide evidence that overconfident individuals are more likely to become entrepreneurs, our results suggest that overconfidence may be triggered by the business environment, even if we start from a set of similar individuals. That is, entrepreneurs may be overconfident because overconfidence serves to more effectively persuade investors and customers. We conjecture that overconfidence is likely to be more pronounced in industries where it is likely to be an asset, for example because measures of true ability are noisy and job competition is high. According to this argument, politics and finance are areas where we would expect overconfidence to be rife.

Our results also provide some indications that asymmetric information can spur arms races between detection and bias, not dissimilar to the investments in deception and detection we observe between different species in the animal world (Trivers, 2011). If people deceive themselves to better deceive others, mechanisms devised by a principal that try to screen high-ability agents may themselves trigger cognitive investments in overconfidence. It is not clear whether principal or agent would win out in such an arms race; in our experiment screening trumped self-deception in

[13] Our measure of lying aversion is negatively correlated with the message sent by unwarned contestants. It is also, positively, though not significantly, correlated with prior overconfidence. Posterior beliefs were a sizable 5.5 percentage points higher for participants above the median split for our measure of lying aversion, a difference that just falls short of statistical significance. Future work could vary the moral motive experimentally and achieve more power than we are able to with our noisy proxy.

the condition with trained employers and unwarned employees.

In psychology, our results support the idea that cognition did not evolve only to generate accurate representations of the world (McKay and Dennett, 2009). Instead, faulty representations can be optimal reactions to an individual's environment. In the physical world, our senses probably evolved to gather a version of reality that serves us in our daily activities. A similar argument can be applied to the social world, where our cognition has evolved to generate representations that serve us in our daily interactions. Self-confidence appears to be a form of cognition where inaccurate representations can provide higher benefits than accurate ones.

When it comes to welfare implications, overconfidence has been associated with personal and social costs.[14] In deciding whether policy should actively attempt to curb overconfidence, an understanding of its benefits is crucial. If overconfidence serves to reduce anxiety or to provide motivation, it may well be socially beneficial. If, as we document in this paper, overconfidence is the result of a privately optimal cognitive strategy that can lead to the misallocation of talent in a zero sum game, it is more likely to be socially harmful. The exact welfare implications of strategic self-deception are likely to be subtle and depend on the environment as well as agents' private information. Characterizing the welfare implications of self-deception constitutes an important research agenda.

Future work could also look into the robustness of our affirmative results across settings in the field and lab. In particular, future experimental research could vary the strategic environment (e.g. by considering bargaining or contests) and the variable subject to self-deception (e.g. by considering absolute past performance, future performance or non-ego-related variables) to pin down the scope of strategic self-deception.

References

Anderson, C., Brion, S., Moore, D. a. and Kennedy, J. a. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, 103 (4), 718–35.

Barber, B. and Odean, T. (2001). Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics*, 116 (1), 261–292.

Baumeister, R. (1998). The Self. In D. Gilbert, S. Fiske and G. Lindzsey (eds.), *The Handbook of Social Psychology*, McGraw-Hill.

Belot, M. and Van de Ven, J. (2016). How private is private information? The ability to spot deception in an economic game. *Experimental Economics*, In press.

Bénabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, 80 (2), 429–462.

[14] In an interview in 2015, Kahneman says that overconfidence is probably the most damaging cognitive bias and the one he would eliminate first if he had a magic wand. See `http://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview`.

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

— and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics*, 126 (2), 805–855.

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

Biais, B. and Hilton, D. (2005). Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of Economic Studies*, 72 (2), 287–312.

Bodner, R. and Prelec, D. (2002). Self-signaling and diagnostic utility in everyday decision making. In I. Brocas and J. D. Carillo (eds.), *Collected Essays in Psychology and Economics*, Oxford University Press.

Bracha, A. and Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75 (1), 67–80.

Brunnermeier, M. and Parker, J. (2005). Optimal expectations. *American Economic Review*, 95 (4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L. and Rustichini, a. (2013). Overconfidence and Social Signalling. *The Review of Economic Studies*, 80 (3), 949–983.

Charness, G., Rustichini, A. and van de Ven, J. (2013). Self-Confidence and Strategic Behavior. *SSRN Electronic Journal*, pp. 1–52.

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

Coutts, A. (2015). Testing Models of Belief Bias : An Experiment. *Mimeo, Nova School of Business and Economics.*

Dawson, C., Meza, D. D., Henley, A. and Arabsheibani, R. (2015). The Power of (Non) Positive Thinking: Self-Employed Pessimists Earn More than Optimists. *IZA working paper*, (9242).

DePaulo, B. M. (1994). Spotting Lies: Can Humans Learn to Do Better? *Current Directions in Psychological Science*, 3 (3), 83–86.

Di Tella, R. and Pérez-Truglia, R. (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs About Others. *American Economic Review*, 105 (11), 3416–3442.

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

Ekman, P., Friesen, W. and O'Sullivan, M. (1988). Smiles when lying. *Journal of personality and social psychology*, 54 (3), 414.

— and O'Sullivan, M. (1991). Who can catch a liar? *American psychologist*, 46 (9), 913.

Ewers, M. and Zimmermann, F. (2015). Image and misreporting. *Journal of the European Economic Association*, 13 (2), 363–380.

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

Gibson, R., Tanner, C. and Wagner, A. F. (2014). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103, 532–548.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95 (1), 384–394.

Grossman, Z. and Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior and Organization*, 84 (2), 510–524.

Gur, R. and Sackeim, H. (1979). Self-Deception: A Concept in Search of a Phenomenon. *Journal of Personality and Social Pscyhology*, 37 (2).

Kahan, D. M. (2016). The Expressive Rationality of Inaccurate Perceptions. *Behavioral and Brain Sciences*, In press.

Koellinger, P., Minniti, M. and Schade, C. (2007). I think I can, I think I can : Overconfidence and Entrepreneurial Behavior. *Journal of economic psychology*, 28 (4), 502–527.

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

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108 (3), 480–498.

Kurzban, R. (2011). Two problems with self-deception: No self and no deception. *Behavioral and Brain Sciences*, 34 (1), 32–33.

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

Mayraz, G. (2012). The optimism bias : model and evidence. pp. 1–40.

McKay, R., Mijovic-Prelec, D. and Prelec, D. (2011). Protesting too much: Self-deception and Self-signaling. *Behavioral and Brain Sciences*, 34 (1), 34–35.

McKay, R. T. and Dennett, D. C. (2009). The evolution of misbelief. *The Behavioral and Brain Sciences*, 32 (6), 493–510; discussion 510–561.

Mobius, M. and Rosenblat, T. (2006). Why beauty matters. *The American Economic Review*, 96 (1), 222–235.

Möbius, M. M., Niederle, M., Niehaus, P. and Rosenblat, T. S. (2014). Managing Self-Confidence. *Mimeo*.

Moore, D. a. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115 (2), 502–17.

Schlag, K., Tremewan, J. and Van der Weele, J. J. (2015). A Penny for Your Thoughts : A Survey of Methods for Eliciting Beliefs. *Experimental Economics*, 18 (3), 457–490.

THOMA, C. (2015). Under- versus overconfidence: an experiment on how others perceive a biased self-assessment. *Experimental Economics*, 19 (1), 218–239.

TRIVERS, R. (1985). Deceit and self-deception. In *Social Evolution*, Benjamin/Cummings, pp. 395–420.

— (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. Basic Books.

VON HIPPEL, W. and TRIVERS, R. (2011). The evolution and psychology of self-deception. *The Behavioral and brain sciences*, 34 (1), 1–16; discussion 16–56.

VRIJ, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley and Sons.

WANG, J. T.-Y., SPEZIO, M. and CAMERER, C. F. (2010). Pinocchio ' s Pupil : Using Eyetracking and Pupil Dilation To Understand Truth-telling and Deception in Games. *The American Economic Review*, 3 (June), 984–1007.

A Additional Figures

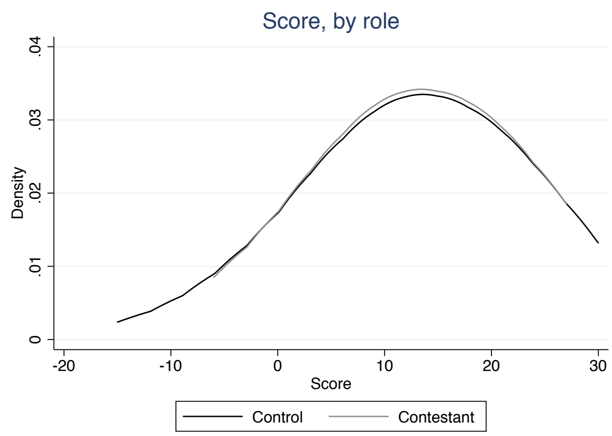
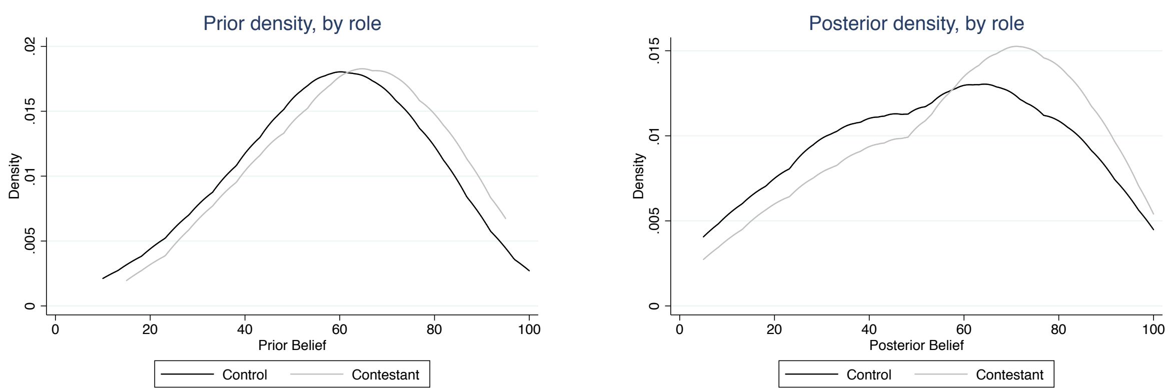


Figure A.1: Prior density by information condition. This figures whose that both control and treatment had identical scores.



(a) (b)

Figure A.2: Belief distribution density by treatment group. Panel a) depicts prior beliefs, Panel (b) depicts posterior beliefs.

B Bayesian updating regressions

We reproduce the regressions in Möbius *et al.* (2014), based on a linearized version of Bayes' formula. The model is given by

$$logit(\mu_{i,post}) = \delta logit(\mu_{i,prior}) + \beta_H I(s_i = H)\lambda_H + \beta_L I(s_i = L)\lambda_H + \epsilon_i. \tag{1}$$

Here, $\mu_{i,prior}$ and $\mu_{i,post}$ represent the prior and the posterior belief of person i respectively. λ is the likelihood ratio, which in our case is $0.75/0.25 = 3$. $I(s_i = H)$ and $I(s_i = L)$ are indicator variables for a high and low signal respectively.

As explained in more detail in Möbius *et al.* (2014), perfect Bayesian updating implies $\delta, \beta_H, \beta_L = 1$. Several different deviations from perfect Baysian updating can be distinguished. If both $\beta_H < 1$ and $\beta_L < 1$, this implies that subjects are "conservative", i.e. they place too little weight on new information. If $\beta_H > \beta_L$, this implies "asymmetry", i.e. subjects place more weight on good signals compared to bad signals. Finally, $\delta < 1$ implies what we call "centrism", an updating bias towards 50%.

	Full sample			Restricted sample		
	All	Control	Contestants	All	Employers	Contestants
δ	0.824***	0.696***	0.937	0.886*	0.791**	0.988
	(0.0517)	(0.0688)	(0.0673)	(0.0585)	(0.0829)	(0.0742)
β_H	0.513***	0.627***	0.407***	0.689***	0.719***	0.635***
	(0.0513)	(0.0691)	(0.0689)	(0.0461)	(0.0620)	(0.0626)
β_L	0.591***	0.619***	0.582***	0.914	0.916*	0.929
	(0.0626)	(0.0885)	(0.0877)	(0.0616)	(0.0849)	(0.0884)
Obs.	279	137	142	199	104	95
R^2	0.685	0.643	0.740	0.800	0.755	0.852

Table B.1: Belief updating regressions. Columns (1)-(3) reflect the entire sample, whereas Columns (4)-(6) exclude subjects with "irregular updates", i.e. either zero updates or updates in the wrong direction. At the top of each column we indicate the sample (control, contestant, all) under consideration. H_0: coefficient equals 1. Standard errors in parentheses $^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table B.1 shows the results of OLS regressions, where stars indicate rejections of the Null hypothesis that the coefficient is 1 at different levels of confidence. Columns (1)-(3) reflect the entire sample, whereas Columns (4)-(6) exclude subjects with "irregular updates", i.e. either zero updates or updates in the wrong direction. At the top of each column we indicate the sample (control, contestant, all) under consideration. Table B.2 shows the same regressions with interaction terms for the treatment group, making it possible to directly compare treatment effects between the two groups.

	(1) Full Sample	(2) Restricted
δ	0.696***	0.791***
	(0.0688)	(0.0830)
β_H	0.627***	0.719***
	(0.0691)	(0.0620)
β_L	0.619***	0.916***
	(0.0885)	(0.0850)
δ * contestant	0.241**	0.198*
	(0.0962)	(0.111)
β_H * contestant	-0.220**	-0.0838
	(0.0976)	(0.0881)
β_L * contestant	-0.0370	0.0131
	(0.125)	(0.123)
Observations	279	199
R^2	0.694	0.803

Table B.2: Belief updating regressions. Column (1) includes the entire sample, whereas Column (2) exclude subjects with "irregular updates", i.e. either zero updates or updates in the wrong direction. H_0: coefficient equals 1. Standard errors in parentheses $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Our main findings are the following:

- **Centrism:** We find $\delta < 1$, for employers, but not for contestants. This implies a bias towards 50%. Since the average belief is over 50%, this implies employers have a downward updating bias compared to contestants.

- **Conservatism:** we reject $\beta_H = 1$ for all specifications, and reject $\beta_L = 1$ for the full sample only. Subjects are conservative on average.

- **(Anti-)asymmetry**: in the restricted sample, $\beta_H < \beta_L$. I.e. both groups respond more to bad signals. This effect is stronger for contestants.

In Table B.3 we repeat the regressions of Table 2 with posterior as the dependent variable. The statistical relationships we see are similar. It is noteworthy that, in Column 2, the coefficients of the warned and lying aversion dummies have increased. Especially, lying aversion is almost significant at the 10% level and actually achieves significance upon including controls such as age. The dominance dummy in Columns 2 and 4 is no longer significant.

	(1)	(2)	(3)	(4)
Score	1.218***	1.121***	1.169***	1.166***
	(0.149)	(0.209)	(0.153)	(0.154)
Contestant (d)	5.003*		5.028*	6.996*
	(2.637)		(2.755)	(4.090)
Warned (d)		4.435		
		(3.828)		
Lying av. (d)		5.511		
		(3.656)		
Dominant (d)			2.780	4.677
			(2.776)	(4.009)
Contestant * Dominant (d)				-3.737
				(5.535)
Constant	38.94***	40.75***	38.80***	37.73***
	(2.751)	(4.344)	(3.213)	(3.688)
Observations	288	144	272	272
R^2	0.172	0.160	0.164	0.165

Table B.3: OLS regressions of confidence after signal on task scores and treatment dummies. "Dominant" and "Lie av." are dummy variables that takes the value 1 if the personality score is above the median. Robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

C The incidence of lying

We define a lie as a message that is higher than the posterior. Thus, we do not count untruthful understatements of one's beliefs as lies. Instrumental variable probit regressions on the incidence of lying are reported in Table C.1. The results in Column 1 clearly indicate that a higher posterior reduces the incidence of a lie. This is consistent with the idea that higher confidence obviates the need for lying. This effect occurs only for the warned subjects, which suggests that the effect is likely to be mechanical: Since warned subjects all choose relatively high messages independent of the posterior, a higher posterior simply reduces the scope for lying. Interestingly, unwarned subjects do not lie significantly less when they have a higher posterior, consistent with the finding in Table 3 that these subjects increase their messages with their posterior.

Condition	All contestants	NT & TnW	TW
	(1)	(2)	(3)
Posterior (signal)	-0.039***	-0.026	-0.080***
	(0.01)	(0.02)	(0.02)
Top 2	-0.034	-0.419	0.616
	(0.32)	(0.38)	(0.54)
Lying aversion (d)	0.070	-0.266	0.566
	(0.25)	(0.30)	(0.41)
Dominant (d)	-0.147	-0.201	-0.119
	(0.23)	(0.27)	(0.43)
Female	0.539**	0.616*	0.258
	(0.26)	(0.33)	(0.41)
Round	0.049*	0.023	0.117**
	(0.03)	(0.03)	(0.06)
Warned (d)	0.252		
	(0.23)		
Constant	2.388***	2.008**	4.932***
	(0.79)	(0.93)	(1.67)
Observations	544	352	192

Table C.1: Determinants of whether contestant lies. IV probit regressions with Posterior instrumented by Black Ball, standard errors (in parentheses) are clustered at the subject level, $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.