# RATIONALITY
# & COMPETITION
## CRC TRR 190

# Deception and Self-Deception

**Peter Schwardmann** (University of Munich)
**Joel van der Weele** (University of Amsterdam)

# Discussion Paper No. 25

March 23, 2017

# Deception and Self-Deception

Peter Schwardmann [*]       Joël van der Weele [◇]

March 2017

## Abstract

Why are people so often overconfident? We conduct an experiment to test the hypothesis that people become overconfident to more effectively persuade or deceive others. After performing a cognitively challenging task, half of our subjects are informed that they can earn money by convincing others of their superior performance. The privately elicited beliefs of informed subjects are significantly more confident than the beliefs of subjects in the control condition. By generating exogenous variation in confidence with a noisy performance signal, we are also able to show that higher confidence indeed makes subjects more persuasive in the subsequent face-to-face interactions.

**JEL-codes**: C91, D03, D83.
**Keywords**: Overconfidence, self-deception, motivated cognition, persuasion, deception.

---

"When a person cannot deceive himself the chances are against his being able to deceive other people."

Mark Twain, *Autobiography of Mark Twain.*


"Show me someone without an ego, and I'll show you a loser."

Donald Trump, *How to Get Rich.*


# 1 Introduction

There is ample evidence that the average person thinks he or she is more skillful, more beautiful and kinder than others. At the same time, such overconfidence may result in substantial personal and social costs. It leads financial traders to lose money, CEOs to initiate more value-destroying mergers, and entrepreneurs to invest in projects of lower value.[1] In the face of these costs, rationalizing the persistence of overconfidence requires the existence of offsetting benefits. Social scientists commonly stress the affective benefits of overconfidence that are associated with a good self-image, ameliorated anxiety about an uncertain future, or the savoring of future successes.[2] However, it is unclear why such desires would survive natural or cultural selection.

An alternative theory posits that the benefits of overconfidence are social in nature. According to Robert Trivers, overconfidence, or self-deception more generally, 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 that "it's not a lie when you believe it", deceiving yourself may reduce both the give-away tells and the moral costs associated with lying.

We conduct a laboratory experiment to test whether overconfidence is used strategically 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 subsequent face-to-face interactions, in which they can earn money if they manage to convince other participants 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 accurate reporting.

Our main hypothesis is that participants who are informed of the profitable deception opportunity will be more confident about their performance than an uninformed control group. We

---

[1]See Barber and Odean (2001); Glaser and Weber (2007) and Biais and Hilton (2005) for overconfidence in the financial domain, Malmendier and Tate (2008) for overconfident CEOs, Koellinger et al. (2007) and Dawson et al. (2015) for overconfident entrepreneurs. Overconfidence can also lead to excess entry into competition (Camerer and Lovallo, 1999), inferior bargaining outcomes (Neale and Bazerman, 1985), and faulty forcasts (Arkes, 2001). The literature section discusses the evidence for overconfidence (and underconfidence) in more detail.

[2]See e.g. Kunda (1990); Baumeister (1998); Brunnermeier and Parker (2005); Köszegi (2006) for a discussion of the affective benefits of confidence. Taylor and Brown (1988) and Block and Colvin (1994) and subsequent papers provide discussions on the impact of positive delusions on health and well-being.

refer to this difference in subjective confidence as "strategic self-deception", as it is motivated by the consequences of holding the belief and not by actual performance or ability. We find that the shadow of future interactions significantly increases average confidence. In a follow-up experiment, we investigate the deeper motives behind this strategic self-deception. In support of a persuasion motive, we find that the treatment effect is stronger among participants who express a belief that confidence helps persuasiveness in the face-to-face interview. By contrast, we find no evidence that self-deception is the result of a desire to reduce the moral costs of lying, of wishful thinking, or of a preference for appearing consistent to the experimenter.

Our second aim is to test whether increased confidence indeed makes participants more persuasive in their attempts to convince others of their superior performance. Drawing clear causal inferences about the effects of confidence on communication strategies and social success is impossible in most datasets, due to reverse causality and unobserved covariates like beauty or extraversion. Our experimental design avoids the pitfalls of endogeneity by generating exogenous variation in privately held beliefs with a noisy feedback signal.

We demonstrate that participants' communication strategies in the face-to-face interactions reflect an appreciation of the benefits of higher confidence: participants condition their message on their privately held beliefs, with higher confidence leading to higher stated confidence. Moreover, we find that more confident participants are rated as significantly more likely to be successful by their evaluators in the face-to-face interactions. The significant impact of confidence on evaluations is present even when we control for stated confidence. This indicates that non-verbal cues are an important driver of a participant's persuasiveness.

This paper makes two main contributions. First, using fully incentivized elicitations of private beliefs, we provide evidence that people become overconfident in the pursuit of a social advantage. This result may help explain the prevalence of overconfidence. Second, by exogenously varying confidence, we demonstrate that higher confidence causes people to be more persuasive. This result provides affirmative evidence for folk wisdom and a large popular self-help literature that present confidence as the road to success, usually by pointing to correlational evidence that highly successful people are also confident.

## 2 Literature

We measure "overconfidence" as the overestimation of one's ability relative to others, or "overplacement" in the terminology of Larrick et al. (2007). Seminal studies show that a large majority of people rate themselves above average on positive, and below average on negative traits (Svenson, 1981; Dunning et al., 1989).[3] At the same time, it is well-documented that people may display overplacement in easy and underplacement in hard tasks (see Larrick et al.

---

[3]These findings have been replicated many times, including by experimental economists using belief measurements that incentivize truthful reporting (Niederle and Vesterlund, 2007; Eil and Rao, 2011; Grossman and Owens, 2012; Möbius et al., 2014).

(2007) and Moore and Healy (2008) for reviews of these findings). Moore and Healy (2008) argue that this effect results naturally from uncertainty about one's own and others' ability levels. Similarly, Benoît and Dubra (2011) show that overplacement will result from a Bayesian interpretation of feedback, if the latter comes from an easy task.[4] These questions about the prevalence of 'true' overconfidence are orthogonal to our experimental treatment, which varies incentives for self-deception but keeps our IQ test constant.

As far as we know, there exists no direct evidence for the social origins of overconfidence, although several results are suggestive of our main hypothesis. For instance, psychologists have demonstrated a connection between self-esteem and strategies of self-presentation (Baumeister et al., 1989). Von Hippel and Trivers (2011) cite multiple studies showing that confidence plays a role in the selection of leaders and that people are impressed by confidence in others (see also Sedikides et al., 2015). Burks et al. (2013) find a correlation between overconfidence about performance on a cognitive test and a measure of social dominance in a sample of trainee truck drivers.

Charness et al. (2013) find that experimental participants strategically use public statements about their own ability to deter others from competing with them. Ewers and Zimmermann (2015) show that the desire to impress an audience causes experimental subjects to publicly state higher confidence. Thoma (2015) demonstrates that men strategically deflate their confidence in order to appear more likable, when others can observe their degree of overconfidence. Since these three studies do not measure private beliefs,[5] they cannot disentangle the role of actual confidence and stated confidence.

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

More broadly, our results relate to a large psychological literature on self-deception, motivated cognition and cognitive dissonance. Kunda (1990) and Baumeister (1998) provide extensive literature reviews of these phenomena. Classical dissonance experiments like Festinger and Carlsmith (1959) use a "forced compliance" design that creates dissonance by inducing or stimulating an aversive cognition or transgressive behavior. Private beliefs and attitudes are then shown to adapt to be consistent with the induced cognition. Our experiment differs from the cognitive dissonance literature in two crucial ways. First, we show that beliefs adapt *in*

---

[4]The authors stipulate several testable conditions to reject such rationalizations of overplacement and show that some, but not all, of the earlier evidence fails these conditions. Subsequent studies refute these conditions, providing more conclusive evidence for overconfidence in some domains (Burks et al., 2013; Benoît et al., 2015; Merkle and Weber, 2011). However, Clark and Friesen (2009) and Moore and Healy (2008) find no evidence of irrational overplacement, leaving an open debate on the cognitive origins of overconfidence.

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

*anticipation* of potentially devious actions, rather than being a (mechanical) reaction to the presence of conflicting cognitions. Second, we demonstrate the *strategic* value of biased beliefs. Whereas the cognitive dissonance literature considers dissonance reduction to be a goal in itself, we show that it can help our participants secure monetary gains in subsequent interactions.

Recently, economists have started to investigate motivated cognition and self-deception in economic contexts (see Bénabou and Tirole (2016) and Gino et al. (2016) for overviews of this nascent field). This small, mostly theoretical, literature seeks to explain overconfident or optimistic beliefs as stemming from a (possibly subconscious) optimization process. The models are often premised on people's uncertainty of their own preferences and abilities, an idea that goes back to social comparison theory (Festinger, 1954) and self-perception theory (Bem, 1967). For instance, Bodner and Prelec (2002) and Bénabou and Tirole (2011) show how, in the presence of uncertainty about their own preferences, people use actions to influence their beliefs about themselves. Economic models assume a variety of motives that cause people to invest in biased beliefs,[6] but none focus on the strategic motive that plays such a crucial role in our experiment.

Our experiment demonstrates the causal impact of confidence on persuasion. In economics, several papers report positive correlations between survey measures of self-esteem and job market outcomes from longitudinal employment data (Drago, 2011; de Araujo and Lagos, 2013). In a sample of business school students, Kaniel et al. (2010) find that self-reported optimists find jobs more easily and are promoted more quickly. They also find that optimistic people are rated by their peers as more charismatic and more likely to succeed. In psychology, there is a large literature on various strategies of "impression management" in employment interviews, both in the lab and in the field. Barrick et al. (2009) show that both verbal and non-verbal elements matter for persuasion, but do not focus on the role of confidence.

A large literature on lie-detection demonstrates that non-professionals are generally not very good at detecting lies (e.g. Vrij, 2008; 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 (2017) criticize the psychological literature for a 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 fraudulent sellers with a rate that is above chance, although contextual richness does not seem to matter much in this setting. We use a similar face-to-face interaction in our experiment, but with a highly structured verbal message.

To our knowledge our paper is the first to use exogenously induced confidence to study persuasion. Aronson and Mettee (1968) use fabricated feedback on a personality test to manipulate

---

[6]Motives for investing in biased beliefs include "ego-utility" from positive self-beliefs (Köszegi, 2006), "anticipatory utility" from the savoring of future successes (see also Brunnermeier and Parker, 2005; Schwardmann, 2017), political ideoligy Le Yaouanq (2016) and loss aversion Maier (2017). More functional motives for overconfidence include an increased personal motivation to overcome self-control problems (Bénabou and Tirole, 2002), combatting anxiety and risk-aversion (Eisenbach and Schmalz, 2015), or an increase in task performance (Compte and Postlewaite, 2004).

self-esteem and look at the incidence of cheating in a card game. Similarly, Sigall and Gould (1977) use manipulated feedback to study the effect of self-esteem on efforts in problem solving. Fischer and Sliwka (2016) use noisy feedback to study the impact of confidence on the motivation to invest in learning.

# 3   Experimental Design

Our experiment was programmed in z-tree (Fischbacher, 2007). The first set of sessions was run in March 2015 at the Munich Experimental Laboratory for Economic and Social Sciences. We ran an additional set of six sessions in October 2016. 288 subjects participated in 18 sessions of exactly 16 subjects each.[7] The experiment lasted slightly over 1 hour and the average subject earned 16.45 euros (min. 4 euros, max. 28.5 euros).[8]

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

## 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": participants indicate which probability $p$ makes them indifferent between winning a monetary prize with known probability $p$, and winning the same prize when an uncertain event $E$ occurs. After participants indicate the point of indifference in a list of probabilities, one probability is randomly drawn from the list and participants are awarded their preferred lottery for that probability. Under this mechanism, reporting the true subjective probability of $E$ maximizes the chance of winning

---

[7]We piloted our design with 32 subjects in December 2014. With 16 subjects in either condition, we found some preliminary, not statistically significant evidence for our main treatment effect (Hypothesis 1), with a difference in privately reported prior beliefs of 11 ppt: average beliefs were 54 in the control, 65 in the treatment ($p = 0.37$ MWU test).

[8]See full instructions online: `https://www.dropbox.com/s/ftd6319tm56x6z0/Instructions.pdf?dl=0`.

| | Employers only | All subjects | Contestants only |
|---|---|---|---|
| | | • Training in belief elicitation | |
| **Self-Deception Stage** | | • Raven intelligence test<br><br>• Elicitation of prior belief<br>• Noisy performance signal<br>• Elicitation of posterior belief | • Information about interview |
| **Deception Stage** | • Lie detection tutorial (LD treatment only)<br><br>• Elicitation of evaluations | • Instructions for interaction<br><br><br>• Face-to-face interaction | • Lying aversion elicitation |
| | | • Profit announcements<br>• Questionnaire | |

Table 1: Timeline of tasks and information by role.

the prize. So unlike in other elicitation mechanisms, the incentives for truthful reporting do not depend on risk preferences.[9] We explained the procedure as well as the fact that truthful reporting maximizes the chance of winning the prize. 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 into anonymous groups of four and proceeded to the intelligence test. The test 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 test. Of the four groups of four in each session, two groups were given the role of "contestants" and two groups were designated as controls. While the control group was not told anything about the deception stage at this point, we informed contestants of the following key points (see Appendix A for the exact text):

---

[9]The veracity of this claim has been established by different authors, and the mechanism has been reinvented several times, see Schlag et al. (2015) for an overview.

Next, participants in both treatment and control group were asked to submit their beliefs using the incentive compatible mechanism 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. The instructions made it very salient that elicited beliefs were strictly confidential and would never be shown to another subject.

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 top 2 for a second time, which concluded the self-deception stage of the experiment.

## 3.2  Deception stage

For the face-to-face interactions, the 16 subjects in a session were divided into two groups, each consisting of 4 contestants and 4 "employers". The latter formed the control group in the first stage of the experiment. Before the interviews began, employers were given a paper evaluation sheet for each contestant on which to write down the contestant's message and their evaluations of the contestant's relative performance, honesty, confidence, likability and attractiveness. The interviews followed a speed dating protocol. Employers left their computer stations and sat down in front of the contestants. There were four rounds of interviews so that every employer

8

would get to interview each of the four contestants in the same group. Every interview took place behind a partition 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, each contestant verbally filled in a number between 0 and 100. In the remainder, we will refer to this number as a 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 evaluation sheets, before moving on to the next contestant. As we explain below, we kept the interaction minimalistic in order to allow for the use of the feedback signal as a *valid* instrument for a contestant's confidence.

After the four rounds of interviews, employers returned to their computer stations and entered their evaluations as well as contestants' messages into the computer. Each employer had to state the probability she attached to each of the contestant's being in the top 2. The computer program enforced that the four elicited probabilities added up to 200, so employers had to revise their original evaluations if necessary. One of these evaluations was selected at random and the employer was paid for it according to the familiar Becker-DeGroot-Marshak mechanism, with a chance to win 10 euros in case of a favorable outcome. The contestant was paid 15 euros if she was in the top 2, and 0 euros otherwise.

In the meantime, contestants participated in a task that elicited their degree of lying aversion, based on Gibson et al. (2013). 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 company's value. Again 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. After seeing their payoffs in the experiment, all subjects filled out a questionnaire about their background characteristics as well as the "assertiveness" scale from the Big 5 personality questionnaire.

Trivers (2011) hypothesizes that the strategic value of confidence consists of eliminating non-verbal give-away cues associated with lying. This value thus depends on employers' ability to spot such cues. The literature on lie detection generally finds that people in the laboratory are not great at spotting lies in others, although they do somewhat better than chance (Von Hippel and Trivers, 2011; Belot and Van de Ven, 2017). However, these studies may understate lie detection abilities outside of the lab, where we seek to deceive spouses, professional recruiters and loan officers. Recruiters and loan officers are often explicitly trained to pick up on private information and a spouse often has substantial experience in doing so. These groups can also draw on a richer interaction environment to help them spot deception.

During the deception stage, we therefore implemented a further treatment in one third of the sessions. This lie detection (LD) treatment was designed to improve employers' ability to spot

lies and to thereby erode some of the artificial advantage bestowed upon the deceiver by the laboratory context. In the treatment, employers participated in a lie-detection tutorial before they embarked on the interview. The tutorial featured on-screen written instructions and lasted 3 minutes. Contestants in the sessions of the LD treatment did not know about the tutorial.[10] If we do not find a strategic value of confidence in interactions in our baseline, but we do find it with trained employers, then we should still expect to see the effect in richer field settings.

# 4    Hypotheses

Our main hypothesis is based on Von Hippel and Trivers (2011). It postulates that contestants engage in self-deception to become more (over)confident than the control group, because this helps them convince employers of their relative success. There are two possible deeper motives for such self-deception. First, higher confidence may eliminate the non-verbal cues associated with lying (Von Hippel and Trivers, 2011). We will refer to this as the "persuasion motive". Second, higher confidence may lead to a reduction in the moral costs of lying (McKay et al., 2011). We will refer to this as the "moral motive". Both motives lead to the following hypothesis.

**Hypothesis 1** *Alerting the contestants to the possibility of deception will increase their confidence relative to the control group.*

Note that this hypothesis constitutes a stronger test than the theory in Von Hippel and Trivers (2011) requires. They maintain 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.

Self-deception is only strategic if contestants employ it in the pursuit of a social advantage. A good indication of whether confidence 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 2** *An exogenous increase in confidence will increase contestants' stated likelihood to the employer of being in the top 2.*

One of the core elements of Trivers' theory is that increased confidence ultimately provides an advantage to the self-deceiver. In our experiment, this means that additional confidence

---

[10]The tutorial used four lie-detection tips from the wiki `http://www.wikihow.com/Detect-Lies` on how to recognize "tells" associated with lying, namely fidgeting, face-touching, fast breathing and incongruent facial expressions. In addition, it explained that avoidance of eye contact is an unreliable indicator of lying. These tips are based on a large body of psychological research grounded in psychological research on lie detection, which has shown that the indicators we have diagnostic power for lies, even if the effects are often not very strong (e.g. Ekman and Friesen, 1969; Zuckerman et al., 1981).
A working paper version of this article also featured a treatment where contestants were warned about the lie-detection tutorial of the employer (Schwardmann and Van der Weele, 2016).
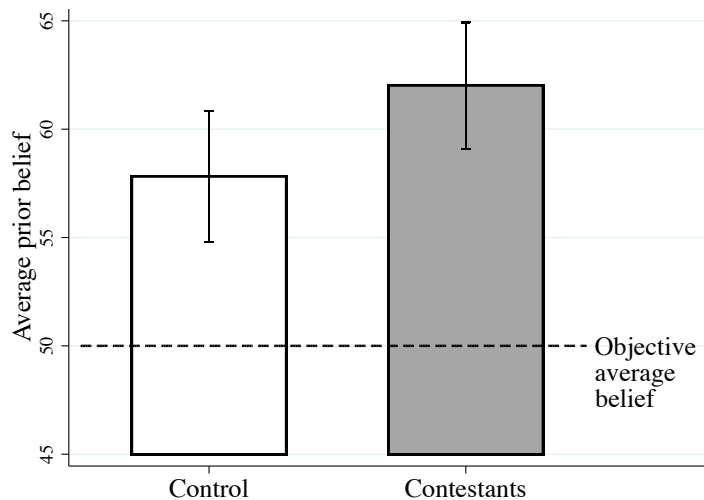
Figure 1: Average beliefs in treatment (contestants) and control group.

increases interview performance, which leads to pecuniary benefits or, more generally, utility from being seen as capable.

**Hypothesis 3** *An exogenous increase in contestant confidence will increase evaluations by the employer.*

## 5   Results of the self-deception stage

Our main hypothesis is that providing some subjects with a profitable opportunity to impress or persuade others will increase their privately elicited beliefs. To test this hypothesis we compare the beliefs before feedback of the contestants to the prior beliefs of the control group. We find that average beliefs among contestants (mean 62.0 percent, s.d. 17.74) are 4.2 percentage points higher than among subjects in the control group (mean 57.8 percent, s.d. 18.4, two sided t-test $p = 0.049$). Figure 1 depicts the average beliefs in treatment and control group relative to the the objective average belief of 50 percent, the true proportion of subjects in the top 2. If we take overconfidence to be the difference between the average belief and the true objective percentage, we find a difference of 12 percentage points in the sample of contestants. More than a third of this can be attributed to strategic self-deception.

Table 2 presents the results of four OLS regression models. The first column shows the raw treatment effect on prior beliefs in percent. We control for individual test scores by adding the score in the second column, and a dummy for every possible test score in the third column. The positive coefficient of test scores in the second regression model tells us that beliefs correlate with actual performance and thus appear at least loosely tethered to reality.

| Dependent variable: | (1)<br>Prior belief | (2)<br>Prior belief | (3)<br>Prior belief | (4)<br>Belief bias |
|---|---|---|---|---|
| Contestant (d) | 4.201** | 3.770** | 4.098** | 5.648** |
| | (2.129) | (1.871) | (1.874) | (2.393) |
| Score | | 1.149*** | | -3.289*** |
| | | (0.124) | | (0.159) |
| Dummy for each score | No | No | Yes | No |
| Constant | 57.81*** | 42.32*** | 60.24*** | 52.04*** |
| | (1.505) | (2.135) | (2.538) | (2.731) |
| | | | | |
| Observations | 288 | 288 | 288 | 288 |
| $R^2$ | 0.013 | 0.241 | 0.280 | 0.602 |

Table 2: OLS regressions of the determinants of private prior belief (column 1-3), and our measure of overconfidence (column 4); robust standard errors in parentheses; $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

The dependent variable in regression model 4 is a constructed measure of individual over- and underconfidence. The variable *belief bias* is constructed by subtracting from a subject's prior belief the average likelihood that a subject with the same test score is in the top 2 of his or her group of four. The treatment effect on belief bias is also positive and significant.[11] All results in Table 2 are robust to the inclusion of the control variables gender or age.

After eliciting beliefs for the first time we provide subjects with an unbiased noisy signal. Posterior beliefs go down in both groups, but the differences between contestants' post-feedback confidence (58.3 percent, s.d. 24.5) and the control group's (54.7 percent, s.d. 24.1) remains rather stable. Due to the binary nature of the feedback signal, the distribution of posterior beliefs becomes more dispersed and, as a result, the difference in means of our confidence measure is no longer significant ($p = 0.21$, two sided t-test). In Appendix B we investigate the updating behavior of subjects and find no meaningful differences between treatment and control group.

**Summary 1** *Participants that are informed about the profitable opportunity to convince another participant of their superior performance are about 4 percentage points more confident. This strategic self-deception explains about one third of the average overconfidence in this group.*

In Section 7.1 we present the results of a replication experiment with 400 subjects, where we find a slightly larger treatment difference of 5 percentage points.

## 6 Results of the deception stage

In this section, we will use the privately elicited, posterior beliefs of the contestants of the self-deception stage as a measure of confidence. We investigate how it influences their communication

---

[11]The coefficient of test score in regression 4 is negative. To see why, note that for very high scores, a subject is always in the top 2 and therefore mechanically exhibits a weakly negative belief bias. The reverse is true for very low test scores.

strategies and evaluations by employers, who were the same participants that formed the control group in the self-deception stage.

## 6.1   Identification strategy

We cannot rely on mere correlations to establish a causal relationship between contestant confidence and the messages stated by the contestant or the evaluation of the contestant by an employer. Such correlations may be spurious, as contestants' confidence is likely to correlate with unobserved determinants of contestants' messages and employers' evaluations. For example, more attractive people may be more confident, experience positive discrimination in evaluations (Möbius and Rosenblat, 2006) and therefore have less need to lie. Alternatively, an arrogant personality may drive both overconfidence and lower evaluations in face-to-face interactions.

To avoid the pitfalls of endogeneity, we make use of the exogenous variation in contestants' confidence that results from the noise component of the feedback signal. Recall that the feedback signal aligns with the true state of the world, i.e. whether a contestant really was in the top 2, in 75 percent of cases. Since the signal is informative, it shifts contestants' beliefs. But, conditional on the true state of the world, the feedback signal a contestant receives is completely random and exogenous. Since we observe and can perfectly control for whether a subject was in the top 2, we can use the feedback signal as an instrument for contestants' posterior beliefs.

Figure 2 shows, for each state of the world, the histograms of posterior beliefs separated by whether a subject received good or bad news. Subjects are very reactive to the signal, making it a strong instrument. This is confirmed in Appendix C, which contains the first stage regressions of the instrumental regressions in this section. By keeping the interview interactions minimalistic, we also assured that subjects were not able to communicate their feedback signal during the interviews. The signal therefore can only influence outcomes through its impact on private beliefs. This guarantees the validity of our instrument.

## 6.2   Do more confident contestants send higher messages?

A good indication of contestants' appreciation of the strategic value of confidence is that verbally stated confidence increases with (exogenous shocks to) their private confidence.

Figure 3 shows that most subjects lie: the average message they publicly state in the interviews exceeds their privately held beliefs. The fitted line indicates that the correlation between confidence and average message is positive. Furthermore, we observe that the distance between average message and the truth decreases with increasing confidence, while the likelihood of observing subjects telling the truth (an observation on the 45 degree line) increases with confidence. Part of this pattern may reflect a "ceiling effect", as privately confident subjects have less room to exaggerate their confidence in public. In total, 38% of subjects provide a consistent message that is equal across all interview rounds.
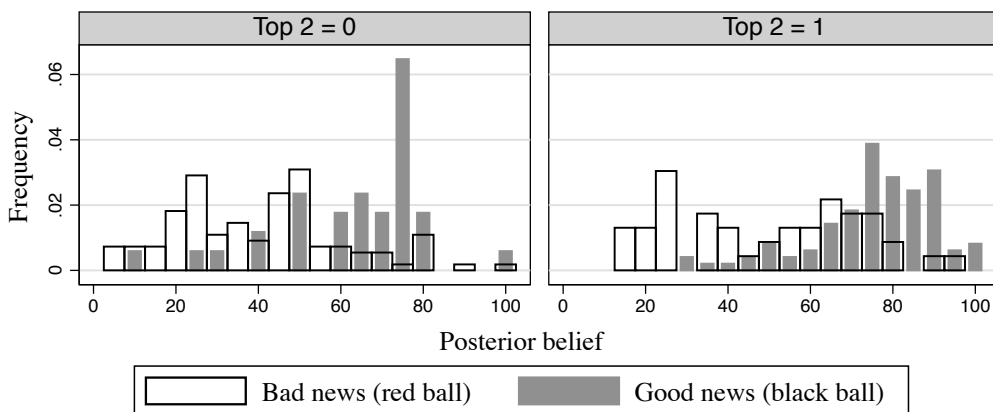
Figure 2: Histograms of posterior beliefs for those not in the top 2 (left panel) and those in the top 2 (right panel).

The first regression model in Table 3 confirms the positive correlation between confidence and message. The OLS regression in the second column shows that a good-news signal increases messages. Since we control for whether the contestant truly was in the top 2 of her group of 4, the coefficient of *black ball* demonstrates the causal effect of confidence on messages.

Our favored specification is the instrumental variable regression in column 3, where confidence is instrumented by *black ball*. It shows that a 1 percentage point increase in the posterior belief, our measure of confidence, increases the message by 0.26 percentage points. Since the coefficient is smaller than 1 (and since the bulk of contestants state messages that lie above their privately held beliefs), an increase in confidence is also associated with a decrease in the size of the lie. Results in Appendix D indicate that lying also decreases along the extensive margin as confidence increases.

The coefficient in the IV regression (column 3) has the same size as the OLS estimate (column 1). Testing for endogeneity based on the specification in column 3, we cannot reject the null hypothesis that posterior beliefs can be treated as an exogenous variable ($p=0.96$).[12]

In column 4, we show that the effect size of the IV regression is robust to controlling for lying aversion, gender and test scores. As expected, more lying averse subjects send lower messages. All regressions in Table 3 indicate that the average contestant increases her message from earlier to later interview rounds. This may be due to contestants getting more practiced at lying.

**Summary 2** *A one percentage point increase in confidence increases the message by 0.26 percentage points.*

---

[12]We use a variant of a standard Hausman test that allows for clustered standard errors (e.g. see Cameron and Miller (2015), Section 7.B)
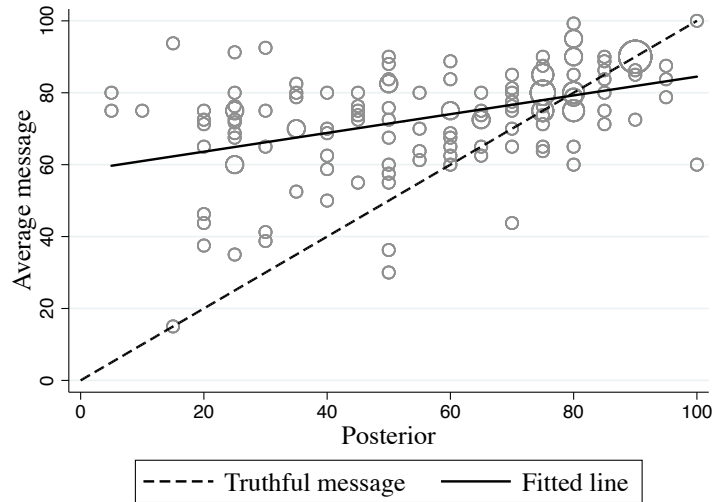
14

Figure 3: Scatter plot of average messages and confidence, as measured by privately elicited posterior beliefs.

## 6.3 Does higher confidence cause better evaluations?

There are two ways in which higher confidence could lead to higher employer evaluations. First, as we have seen, higher confidence leads contestants to state higher messages. If employers do not discount the truth-content of messages by too much, for instance because they expect some subjects to be honest, higher messages may lead to higher evaluations. Second, more confident individuals may be able to more convincingly deliver a message of a given size by avoiding suspicious non-verbal cues. These channels reflect the two advantages of higher confidence that respectively give rise to the moral motive and the persuasion motive of self-deception.

The first four regression models in Table 4 focus on our baseline treatment. The OLS regression in column 1 shows that more confident contestants, i.e. those with a higher posterior belief, receive higher evaluations from employers. Column 2 tells us that contestants who receive a good signal obtain higher evaluations, controlling for whether the contestant actually was in the top 2. This demonstrates that contestant confidence has a positive causal effect on evaluations by employers.

In the IV regression in column 3, the posterior belief is instrumented by the signal, again controlling for *top 2*. We see that a one percentage point increase in the posterior belief leads to a 0.32 percentage point increase in the evaluation a contestant receives. The fact that the coefficient in the IV regression exceeds the coefficient in the OLS regression is indicative of the OLS regression being biased by endogeneity.[13] We cannot be sure about the exact reason for

---

[13]We use a variant of a standard Hausman test that allows for clustered standard errors (e.g. see Cameron and Miller (2015), section 7.B) to test for endogeneity. Whether we can reject the null hypothesis that posterior beliefs are exogenous depends on the specification. It can not be rejected in the specification of column 3 (p=0.29). But

|  | OLS regressions | | IV regressions | |
| | (1) | (2) | (3) | (4) |
| **Dependent variable:** | Message | Message | Message | Message |
| Posterior | 0.261*** | | 0.260*** | 0.299*** |
| | (0.0483) | | (0.0972) | (0.0963) |
| Round | 0.416** | 0.416** | 0.416** | 0.416** |
| | (0.200) | (0.200) | (0.197) | (0.197) |
| Black ball (d) | | 5.381** | | |
| | | (2.487) | | |
| Top 2 (d) | | 3.403 | 0.538 | -7.335** |
| | | (2.732) | (3.138) | (2.987) |
| Lying aversion | | | | -9.056* |
| | | | | (4.829) |
| Female (d) | | | | -0.869 |
| | | | | (2.212) |
| Score | | | | 0.670** |
| | | | | (0.279) |
| Constant | 57.35*** | 68.38*** | 57.12*** | 53.69*** |
| | (3.645) | (2.290) | (5.189) | (5.140) |
| Observations | 576 | 576 | 576 | 576 |
| $R^2$ | 0.158 | 0.061 | 0.158 | 0.226 |

Table 3: Determinants of contestant's messages; in the IV regressions the posterior belief is instrumented by *black ball*; standard errors (in parentheses) are clustered at the subject's group of 4 contestants and 4 employers; $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

the endogeneity; one possible cause is an omitted personality trait, such as arrogance, that is positively correlated with confidence and negatively correlated with evaluations.

To investigate how much of the effect of beliefs on evaluations works through the messages contestants sent to employers, column 4 adds messages to the set of regressors.[14] When doing so, the coefficient of posterior beliefs decreases slightly but remains significant. Messages themselves also have a strong effect on evaluations. Column 4 therefore suggests that higher confidence leads to higher evaluations through both verbal and non-verbal channels. Quantitatively, these two channels appear to be of roughly equal importance.

**The lie detection treatment.** In column 5 we study the determinants of employers' evaluations in the lie detection (LD) condition. We hypothesize that, controlling for messages, employers in the LD condition might be more responsive to privately held beliefs than employers in the baseline condition. The data show that the opposite appears to be true, as employers

---

it can be rejected in the specification in column 4 (p=0.079).

[14]We saw in section 6.2 that there appears to be no omitted variable that jointly determines confidence and messages. Nonetheless, there may exist omitted variables that correlate with a contestant's message and the evaluation she obtains, independent of confidence. Since we cannot completely rule out this particular source of endogeneity, the specifications that feature the contestant's message as an explanatory variable should be interpreted with caution.

|  | OLS regressions | | IV regressions | | |
|---|---|---|---|---|---|
|  | (1) | (2) | (3) | (4) | (5) |
| **Dependent variable:** | Evaluation | Evaluation | Evaluation | Evaluation | Evaluation |
| Posterior | 0.221*** | | 0.320** | 0.290** | -0.252 |
|  | (0.0635) | | (0.127) | (0.132) | (0.179) |
| Round | -0.792 | -0.792 | -0.792 | -0.917 | -2.676* |
|  | (1.275) | (1.277) | (1.245) | (1.249) | (1.579) |
| Black ball (d) | | 5.650** | | | |
|  | | (2.308) | | | |
| Top 2 (d) | | 3.907 | -1.295 | -1.864 | 13.78*** |
|  | | (4.090) | (4.815) | (4.629) | (4.541) |
| Message | | | | 0.284*** | 0.518*** |
|  | | | | (0.0923) | (0.150) |
| Constant | 39.02*** | 47.38*** | 33.85*** | 15.13*** | 26.93*** |
|  | (4.978) | (4.022) | (6.533) | (5.058) | (6.495) |
| Condition | Baseline | Baseline | Baseline | Baseline | LD |
| Observations | 384 | 384 | 384 | 384 | 192 |
| $R^2$ | 0.070 | 0.043 | 0.057 | 0.112 | 0.243 |

Table 4: Determinants of employers' evaluations of contestant; in the IV regressions the posterior belief is instrumented by *black ball*; standard errors (in parentheses) are clustered at the subjects' group of 4 contestants and 4 employers, $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

in the LD treatment appear not to be responsive to contestants' confidence. Instead, the significant coefficient of *top 2* indicates that they are able to identify true ability. This is somewhat surprising, because spotting true ability is different from spotting lies, which we define as increases of stated confidence over truly held confidence. Unfortunately, we cannot put too much weight on these interpretations. It turns out that our random variation in feedback produced a skewed distribution of signals and the results are sensitive to outlier observations.[15] Thus, the LD treatment does not provide any definite evidence on the importance of confidence when facing a trained audience, which remains an interesting and open question.

**Summary 3** *An increase in confidence by 1 percentage point increases evaluations by about 0.22 percentage points. We find suggestive evidence that both higher interview messages and a more convincing non-verbal performance by contestants are the drivers of this effect.*

In Appendix E, we establish that the results in Table 4 are robust to more flexible model specifications for messages and the inclusion of control variables related to characteristics of the

---

[15] The effect of the exogenous shift in beliefs (*black ball*) and true ability (*top 2*) on evaluations is identified by looking at those contestants for whom these two independent variables do not take the same value. Unfortunately, only 2 out of 48 contestants in the LD condition were not in the top 2 and received a good signal. These subjects also turned out to be much more optimistic (an average posterior of 80 percent) compared to the baseline treatment. Furthermore, only 5 out of 48 contestants were in the top 2 and received a bad signal. These subjects were much more pessimistic (an average posterior of 40 percent) compared to the baseline. This makes it impossible to adequately disentangle the relative importance of beliefs and ability in shaping evaluations.

contestant, the employer, and other contestants in the same group.

# 7 Discussion

In this section we present the results of a follow-up experiment. We provide a replication of our self-deception result, and disentangle the potential motives behind it. We also discuss the terminology of self-deception.

## 7.1 Replication

To replicate our self-deception result and delve deeper into the motivation behind it, we ran a separate experiment. We conducted the sessions in October 2016 in the Munich Experimental Laboratory for Economic and Social Sciences, where 400 subjects (200 contestants and 200 employers) participated. The self-deception stage was the same as in the first experiment, but we made a slight change to the composition of the control group. In each experimentally matched group of 4 contestants, two contestants were informed about (their role in) the interviews before the belief elicitation, and two were informed afterwards (but before the interview). Contestants were not aware that some of their group members were in a different condition. Thus, while 100 treated contestants had knowledge about the interview at the time of the elicitation, the control group consisting of 100 untreated contestants and 200 employers did not.[16]

We replicate our main result on self-deception. On average, the informed group is more confident (mean 60.8, s.d. 18.2) than the control group (mean 54.4, s.d. 20.5, two-sided t-test $p = 0.0063$). Table 5 presents OLS regressions, where column 1 is analogous to column 2 of Table 2. Controlling for score, contestants who were aware of the interview are almost 5ppt more confident than the control group.

## 7.2 Motives for self-deception

**Persuasion motive.** Trivers' theory is based on the idea that self-deception occurs because of an (implicit) belief that confidence helps to convince others. To test whether this persuasion motive is at play in driving the strategic self-deception we observe, our follow-up experiment included a task to elicit the belief that confidence aids persuasion in the deception stage. This task took place after the interviews, but before payoffs were announced. We incentivized people to state beliefs about the probability that each of three anonymous contestants from the first

---

[16]We made these changes to test whether self-deceiving contestants outperform non-self-deceiving contestants. We hypothesized that belief elicitation would function as an anchor for the beliefs of the control-contestants, so that treated contestants would go into the interview with higher confidence, and outperform the control-contestants. However, in a post-interview measure of confidence we found that the difference in confidence between the two groups had almost entirely dissipated during the interview. This suggested that the belief the control group increased their confidence upon learning about the interview and belief elicitation did not function as anchor. In line with this, we found no evidence that the treated group did better in the interview, although we did see a statistically highly significant correlation between confidence and evaluations.

|  | Baseline (1) Belief | Persuasion (2) Belief | Lying av. (3) Belief | Wishful (4) Belief | Consistent (5) Belief | Dominant (6) Belief |
|---|---|---|---|---|---|---|
| **Dependent variable:** | | | | | | |
| Score | 0.884*** | 0.863*** | 1.008*** | 0.881*** | 0.880*** | 0.930*** |
|  | (0.131) | (0.132) | (0.135) | (0.133) | (0.206) | (0.0883) |
| Treatment | 4.839** | 0.549 | 5.018 | 3.880 | 5.620* | 12.54** |
|  | (2.049) | (3.032) | (4.370) | (2.476) | (2.964) | (5.098) |
| Confidence Pays (CP) |  | 0.0543 |  |  |  |  |
|  |  | (0.0512) |  |  |  |  |
| Treatment × CP |  | 0.159* |  |  |  |  |
|  |  | (0.0857) |  |  |  |  |
| Lying Aversion (LA) |  |  | -1.866 |  |  |  |
|  |  |  | (8.684) |  |  |  |
| Treatment × LA |  |  | -0.287 |  |  |  |
|  |  |  | (9.678) |  |  |  |
| Ability Pays (AP) |  |  |  | -0.0453 |  |  |
|  |  |  |  | (0.0531) |  |  |
| Treatment × AP |  |  |  | 0.0705 |  |  |
|  |  |  |  | (0.0913) |  |  |
| Consistency (C) |  |  |  |  | 6.339 |  |
|  |  |  |  |  | (6.244) |  |
| Treatment × C |  |  |  |  | 2.928 |  |
|  |  |  |  |  | (7.290) |  |
| Dominance (D) |  |  |  |  |  | 17.20*** |
|  |  |  |  |  |  | (5.395) |
| Treatment × D |  |  |  |  |  | -14.68* |
|  |  |  |  |  |  | (8.422) |
| Constant | 47.98*** | 46.83*** | 45.37*** | 48.61*** | 44.55*** | 36.69*** |
|  | (1.574) | (1.871) | (4.172) | (1.824) | (2.861) | (3.280) |
| Observations | 400 | 400 | 344 | 400 | 200 | 672 |
| $R^2$ | 0.129 | 0.146 | 0.178 | 0.131 | 0.159 | 0.186 |

Table 5: Determinants of privately stated confidence; OLS regressions; robust standard errors in parentheses; $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

experiment, denoted $X$, $Y$, $Z$, were selected in the decisive interview round. Subjects were told (truthfully) that contestants $X$ and $Y$ differed from each other only in ability, while contestants $X$ and $Z$ differed from each other only in privately stated confidence.[17] Thus, the differences in these pairwise beliefs give us an individual measure of the subjective importance of ability and confidence for interview success.

To test the impact of this belief, column 2 in Table 5 shows OLS regressions that include *Confidence pays*, our measure for this belief, as well as an interaction with the treatment.

---

[17]Specifically, let $a$ denote privately stated confidence in percentages, $b$ confidence stated in the interview, and $c$ being in the top 2 (1) or not (0). Subjects saw the following information in a table: $(a_X, b_X, c_X) = (75, 75, 1)$, $(a_Y, b_Y, c_Y) = (75, 75, 0)$ and $(a_Z, b_Z, c_Z) = (35, 75, 1)$. The subjects were asked to give their best guess of the probability that each of the three candidates would get selected. Each guess was rewarded with the BDM mechanism with a prize of 30 eurocents.

The belief that confidence enhances interview performance increases the treatment effect at marginal significance levels. Moreover, the coefficient for the treatment dummy becomes small and statistically indistinguishable from zero. These results are direct evidence that self-deception is motivated by a desire to be more persuasive. On the basis of our data, we cannot say whether subjects are aware of this motivation when we elicited their confidence either during the self-deception stage or afterwards.

**Lying aversion.** A further deep motive consistent with strategic self-deception is that 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). Di Tella et al. (2015) provide evidence for such self-deception from a somewhat different context, showing that people will negatively bias their beliefs about other people in order to justify taking money from them. We hypothesize that more lying averse individuals, as measured by the task in Gibson et al. (2013, see Section 3), are more likely to self-deceive.[18]

Column 3 of Table 5 introduces our measure of lying aversion, as well as an interaction with the treatment. We elicited lying aversion from contestants in both our original and our replication experiment, so we combine the data to achieve maximum power. We find that lying aversion does not decrease the size of the treatment effect, although the treatment dummy now becomes insignificant.[19]

## 7.3 Alternative motives

We now discuss two further motives that could plausibly drive our results in the self-deception stage, but are not consistent with strategic self-deception.

**Wishful thinking.** Wishful thinking can be a force behind self-deception (Mayraz, 2011; Heger and Papageorge, 2013). A wishful thinker derives "anticipatory utility" from the belief that things will turn out well for her in the future. In the context of our experiment, this may inspire her to inflate her belief that she is a "top 2 type", either because she thinks that her higher belief makes her more persuasive, or because she believes that employers can actually spot who is in the top 2, regardless of contestants' beliefs. While the former mechanism is consistent with the persuasion motive we hypothesize, the latter is a potential confound to the result that self-deception is strategic.

We investigate whether this latter form of self-deception plays a role in our experiment, using contestants' belief that ability helps interview performance. The measurement of this

---

[18]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.

[19]The coefficient for lying aversion is also insignificant in the follow-up experiment only. Note that since we only have the lying aversion measure for contestants, the treatment dummy is estimated using only the control group of uninformed contestants in the follow-up experiment.

belief was described above in the discussion of the persuasion motive. Column 3 of Table 5 includes OLS regressions where we include *ability Pays*, our measure for the belief that ability enhances interview performance, as well as an interaction term with the treatment dummy. We find no evidence that anticipatory utility drives self-deception in our experiment. The coefficient for the interaction is not close to conventional significance levels, and although the treatment effect declines slightly, we cannot accept the null hypothesis with any confidence.

**Appearing consistent.** Another potential confound for our results is a motivation to appear consistent. At the belief elicitation stage, subjects may anticipate that they will lie in the interview, even though they have not yet received the exact instructions for this stage. Subjects may wish to avoid appearing like a liar to the experimenter, the only person who observes both private and public statements, albeit in an anonymized and impersonal manner. This wish may lead them to exaggerate their private belief statement and incur the costs of decreased accuracy.

To test whether such concerns play a role we implemented a new belief elicitation *after* the deception stage. This elicitation was designed to maximize consistency concerns for two reasons. First, because subjects have just made their verbal interview statements, consistency concerns should be more salient in this elicitation. Moreover, we did not incentivize this elicitation for accuracy, making it costless to submit a report that is consistent with the interviews.

We find that 48 out of 200 subjects (24%) report a belief that is the same as the message in the last interview round. This identifies the subjects who are potentially (most) concerned with consistency. In Table 5, column 5, we include a dummy for such consistency in the regressions for confidence, as well as an interaction term between consistency and the treatment. We observe no significant coefficient for the latter and a slight *increase* in the treatment effect. Statistical significance declines a bit in what is now a smaller sample, as consistency is defined only for contestants and not for employers. We obtain very similar results if we define consistency as the equality between the initial (incentivized) belief elicitation and the message in the first or any other interview round. We conclude that consistency concerns do not explain the treatment effect.

**Dominance as a trait.** Burks et al. (2013) find that socially "dominant" individuals exhibit more overconfidence. They conjecture that this is because socially dominant individuals attribute more importance to the belief of others about their ability. If this is indeed the case, dominant individuals should exhibit a boost in confidence when told about the interview, as this introduces an audience for their confidence. We thus expect a stronger treatment effect for dominant individuals in our experiment.

To test this hypothesis, we collected a measure of "Assertiveness", based on the "Big Five Aspect Scales" (BFAS, DeYoung et al., 2007).[20] DeYoung et al. (2007) validate this scale and

---

[20]The BFAS divides each of the Big 5 domains into two different "aspects", measured on a 10-item scale. DeYoung et al. (2007) show that these aspects correlate with other questionnaire measures, as well with genetic

argue that it captures the aspect of extroversion that is related to dominance and personal agency. Column 6 of Table 5 introduces our measure of dominance and the interaction with the treatment. We elicited dominance from contestants in both our original and our replication experiment, so we combine the data to achieve maximum power.[21] Dominant individuals have much higher confidence: going from the bottom rank to the top rank of the scale increases confidence by 17 ppt. We thus replicate the result in Burks et al. (2013). The negative interaction term shows that in contrast to our hypothesis, more dominant individuals react much *less* strongly to the treatment. When taking this into account, the treatment effect for the least dominant person rises to 12.5 ppt.

One way to understand these results is that confidence is a character trait of dominant individuals, who are simply confident in any situation, competitive or not. By contrast, the confidence of less dominant individuals is more amenable to the strategic situation.

**Summary 4** *We find that participants who believe that persuasion is important for interview success exhibit more strategic self-deception than participants who do not hold this belief. A moral motive, wishful thinking, and preferences for consistency do not explain the treatment effect. Dominant individuals are more confident and their confidence reacts less to information about the interview.*

## 7.4  "Self-deception" as a metaphor

In the introduction, we defined "self-deception" as a change in confidence that is motivated by the consequences of holding the belief, not by actual performance or ability. This definition does not satisfy the criterion for self-deception posed in Gur and Sackeim (1979) and Pinker (2011), as we do not show that subjects hold two contradicting beliefs at the same time. Following Von Hippel and Trivers (2011), we posit that self-deception can occur by simply failing to encode negative information about the self, without necessarily constructing a realistic image alongside the distorted one. The terminology of "self-deception" has been criticized by Kurzban (2011), among others. Kurzban argues that in a modular conception of the brain there is little space for a "self" that is being deceived. We think these and other conceptual criticisms of self-deception are effectively countered in Von Hippel and Trivers (2011, see the discussion section).

We do not take the term "self-deception" to imply a literal description of how the brain works. Rather, we take it to be a useful metaphor for our purposes, as it captures both strategic and intentional shift in confidence (the "deception") and the fact that people make these beliefs their own and are willing to wager money on them (the "self"). While this duality may be too simple for some purposes, it enlightens the conflict between self-enhancement and realism

---

factors. "Assertiveness" is one of the aspects of the Extroversion domain. Items are e.g. "I know how to captivate people", and "I can talk others into doing things". We added the scores on all these items, and normalized the scale to fall between 0 (least dominant) and 1 (most dominant).

[21]In one session, our software malfunctioned during the questionnaire, resulting in lost data on this measure from 16 individuals. The aggregate results reported here are qualitatively robust in each individual experiment.

that is the focus of our experiment. In this sense, it is not different from other multiple-self metaphors in economic psychology, like the "system 1" and "system 2" distinction that highlights a fundamental tradeoff between the quality and speed of decision making (Kahneman, 2011), and planner-doer models that elucidate specific aspects of intertemporal choice (Thaler and Shefrin, 1981).

# 8    Conclusion

The results of our experiment demonstrate that the strategic environment matters for cognition about the self. We find that information about deception opportunities increases confidence, and that this effect is driven by a (possibly subconscious) belief on behalf of our subjects that higher confidence makes one more persuasive. We also find that increased confidence pays off: controlling for ability, more confident participants are more convincing in face-to-face interactions. Participants in our experiment successfully employ overconfidence for social gain, which may help explain the prevalence of overconfidence in the field.

Although our experiment takes place in the lab and features a highly stylized interaction, there are reasons to believe that our results will generalize to other settings. Weiss and Feldman (2006) present evidence from a field experiment that people routinely tell falsehoods in job interviews to enhance the impression they make on the employer. Barrick et al. (2009) conduct a meta-study on "impression management" in job interviews, including both laboratory ("mock interviews") and field studies ("real hiring decisions"). The results show that the impact of impression management on interview ratings is similar in the lab and in the field.

We find support for 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 evolved to construct a version of reality that aids our survival. A similar argument can be applied to the social world, where our cognition has evolved to generate representations that improve our payoffs. Self-confidence appears to be a form of cognition where inaccurate representations can sometimes provide higher benefits than accurate ones.

Our results have several implications for the link between economic incentives and psychology. For instance, they suggest when and where we should expect overconfidence to occur. It may arise in employer-employee relationships because of its strategic benefits in job interviews and wage negotiations. Arguably, confidence may be even more valuable amongst the self-employed, whose economic survival often depends more immediately on persuading investors and customers.[22] Generally, overconfidence is likely to be more pronounced in settings where its

---

[22]While several studies provide evidence for the selection of overconfident individuals into entrepreneurship (see e.g. Dawson et al., 2014, for an overview), our results suggest a reverse causality by which self-employment may encourage overconfidence. Consistent with this argument, Dawson et al. (2014) find that, in addition to the selection effect of overconfidence, becoming self-employed is followed by a boost in confidence.

strategic value is highest, i.e. where measures of true ability are noisy, job competition is high and persuasion is an important part of success. Accordingly, we would expect overconfidence to be rife amongst high level professionals in business, finance, politics and law.

The results also suggest that asymmetric information changes with the economic environment. If people deceive themselves to better deceive others, mechanisms devised by a principal trying to detect an agent's private information may themselves trigger cognitive investments in biased beliefs, perhaps sparking an "arms race" between detection and deception efforts. Endogenizing asymmetric information in this way would be an interesting theoretical exercise, complementing current models of screening potentially honest agents (Alger and Renault, 2006), and agents who engage in wishful thinking (Bridet and Schwardmann, 2017).

A crucial challenge for researchers on overconfidence is to figure out whether policy makers should attempt to debias overconfident individuals. If overconfidence mainly serves to reduce anxiety or to provide motivation, it may well be socially beneficial. On the other hand, hiring or funding the confident instead of the able leads to a misallocation of talent and resources. Therefore, if overconfidence is mainly the result of strategic self-deception and does not possess psychological benefits, it is more likely to be socially wasteful. 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 these welfare implications constitutes an important research agenda.

# References

**Alger, Ingela and Régis Renault**, "Screening Ethics When Honest Agents Care About Fairness," *International Economic Review*, 2006, *47* (1), 59–85.

**Anderson, Cameron, Sebastien Brion, Don A. Moore, and Jessica A. Kennedy**, "A status-enhancement account of overconfidence.," *Journal of Personality and Social Psychology*, 2012, *103* (4), 718–35.

**Arkes, Hal R.**, "Overconfidence in judgmental forecasting," *Principles of forecasting*, 2001, (30), 495–515.

**Aronson, Elliot and David R. Mettee**, "Dishonest behavior as a function of differential levels of induced self-esteem," *Journal of Personality and Social Psychology*, 1968, *9* (2), 121–127.

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

**Barrick, Murray R., Jonathan A. Shaffer, and Sandra W. DeGrassi**, "What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance," *The Journal of applied psychology*, 2009, *94* (6), 1394–1411.

**Baumeister, Roy F.**, "The Self," in David Gilbert, Susan Fiske, and Gardner Lindzsey, eds., *The Handbook of Social Psychology*, McGraw-Hill, 1998.

_ , **Dianne M. Tice, and Debra G. Hutton**, "Self-Presentational Motivations and Personality Differences in Self-Esteem," *Journal of Personality*, 1989, *57* (3), 547–579.

**Belot, Michele and Jeroen Van de Ven**, "How private is private information? The ability to spot deception in an economic game," *Experimental Economics*, 2017, *20* (1), 19–43.

**Bem, Daryl J.**, "Self-perception: An alternative interpretation of cognitive dissonance phenomena," *Psychological review*, 1967, *74* (3), 183–200.

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

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

_ **and** _ , "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–164.

**Benoît, Jean-Pierre and Juan Dubra**, "Apparent Overconfidence," *Econometrica*, 2011, *79* (5), 1591–1625.

_ , _ , **and Don Moore**, "Does the Better-Than-Average Effect Show That People Are Overconfident?: Two Experiments," *Journal of the European Economic Association*, 2015, *13* (2), 293–329.

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

**Block, Jack and C. Randall Colvin**, "Positive illusions and wellbeing: Separating fiction from fact," *Psychological Bulletin*, 1994, *116* (1), 28.

**Bodner, Ronit and Drazen Prelec**, "Self-signaling and diagnostic utility in everyday decision making," in Isabelle Brocas and Juan D. Carillo, eds., *Collected Essays in Psychology and Economics*, Oxford University Press, 2002.

**Bridet, Luc and Peter Schwardmann**, "Selling Dreams: Endogenous Optimism and Collateral Use in Lending Contracts," *Mimeo, Ludwig Maximilians Universität München*, 2017.

**Brunnermeier, Markus K. and Jonathan A. Parker**, "Optimal expectations," *American Economic Review*, 2005, *95* (4), 1092–1118.

**Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini**, "Overconfidence and Social Signalling," *The Review of Economic Studies*, 2013, *80* (3), 949–983.

**Buser, Thomas, Leonie Gerhards, and Joël J. Van der Weele**, "Measuring Responsiveness to Feedback as a Personal Trait," *Tinbergen Institute Discussion Paper*, 2016.

**Camerer, Colin and Dan Lovallo**, "Overconfidence and excess entry: An experimental approach," *The American Economic Review*, 1999, *89* (1), 306–318.

**Cameron, A. Colin and Douglas L. Miller**, "A Practitioner ' s Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

**Charness, Gary, Aldo Rustichini, and Jeroen van de Ven**, "Self-Confidence and Strategic Behavior," *SSRN Electronic Journal*, 2013.

**Clark, Jeremy and Lana Friesen**, "Overconfidence in forecasting of own performance: an experimental study," *Economic Journal*, 2009, *119* (2004), 229–251.

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

**Coutts, Alexander**, "Good News and Bad News are Still News: Experimental Evidence on Belief Updating," *Mimeo, Nova School of Business and Economics*, 2016.

**Dawson, Christopher, David de Meza, Andrew Henley, and G. Reza Arabsheibani**, "Entrepreneurship: Cause and Consequence of Financial Optimism," *Journal of Economics and Management Strategy*, 2014, *23* (4), 717–742.

**_ , _ , _ , and _** , "The Power of (Non) Positive Thinking: Self-Employed Pessimists Earn More than Optimists," *IZA working paper*, 2015, (9242).

**de Araujo, Pedro and Stephen Lagos**, "Self-esteem, education, and wages revisited," *Journal of Economic Psychology*, 2013, *34*, 120–132.

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

**DeYoung, Colin G., Lena C. Quilty, and Jordan B. Peterson**, "Between facets and domains: 10 aspects of the Big Five," *Journal of personality and social psychology*, 2007, *93* (5), 880–896.

**Di Tella, Rafael, Ricardo Pérez-Truglia, Andres Babino, and Mariano Sigman**, "Conveniently Upset: Avoiding Altruism by Distorting Beliefs About Others," *American Economic Review*, 2015, *105* (11), 3416–3442.

**Drago, Francesco**, "Self-esteem and earnings," *Journal of Economic Psychology*, 2011, *32* (3), 480–488.

**Dunning, David, Judith A. Meyerowitz, and Amy D. Holzberg**, "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability," 1989, *57* (6), 1082–1090.

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

**Eisenbach, Thomas M. and Martin C. Schmalz**, "Anxiety, Overconfidence, and Excessive Risk Taking," *FRB of New York Staff Report*, 2015, *711* (December).

**Ekman, Paul and Maureen O'Sullivan**, "Who can catch a liar?," *American Psychologist*, 1991, *46* (9), 913.

    and **Wallace V. Friesen**, "Nonverbal leakage and clues to deception," *Psychiatry*, 1969, *32*, 88–106.

    ,   , and **Maureen O'Sullivan**, "Smiles when lying," *Journal of Personality and Social Pscyhology*, 1988, *54* (3), 414.

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

**Festinger, Leon**, "A theory of social comparison processes," *Human Relations*, 1954, *7* (2), 117–140.

    and **James M. Carlsmith**, "Cognitive consequences of forced compliance," *Journal of abnormal psychology*, 1959, *58* (2), 203–210.

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

**Fischer, Mira and Dirk Sliwka**, "Confidence in Knowledge of Confidence in the Ability to Learn; An Expeirment on the Causal Effects of Beliefs on Motivation," *Mimeo, University of Cologne*, 2016, (December).

**Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner**, "Preferences for truthfulness: Heterogeneity among and within individuals," *American Economic Review*, 2013, *103* (1), 532–548.

**Gino, Francesca, Michael I. Norton, and Roberto A. Weber**, "Motivated Bayesians: Feeling Moral While Acting Egoistically," *Journal of Economic Perspectives*, 2016, *30* (3), 189–212.

**Glaser, Markus and Martin Weber**, "Overconfidence and trading volume," *GENEVA Risk and Insurance Review*, 2007, *32* (1), 1–36.

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

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

**Gur, Ruben C. and Harold A. Sackeim**, "Self-Deception: A Concept in Search of a Phenomenon," *Journal of Personality and Social Pscyhology*, 1979, *37* (2), 147–169.

**Heger, Stephanie A. and Nicholas W. Papageorge**, "We Should Totally Open a Restaurant : Performance Uncertainty and Optimistic Beliefs," *Mimeo, Available at SSRN 2348410*, 2013.

**Kahan, Dan M**, "The Expressive Rationality of Inaccurate Perceptions," *Behavioral and Brain Sciences*, 2016, *In press.*

**Kahneman, Daniel**, *Thinking, fast and slow*, MacMillan, 2011.

**Kaniel, Ron, Cade Massey, and David T. Robinson**, "The importance of being an optimist: Evidence from labor markets," *National Bureau of Economic Research Working Paper Series*, 2010, *No. 16328* (February 2016).

**Koellinger, Philipp, Maria Minniti, and Christian Schade**, "I think I can, I think I can": Overconfidence and Entrepreneurial Behavior," *Journal of Economic Psychology*, 2007, *28* (4), 502–527.

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

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

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

**Larrick, Richard P., Katherine A. Burson, and Jack B. Soll**, "Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not)," *Organizational Behavior and Human Decision Processes*, 2007, *102* (1), 76–94.

**Le Yaouanq, Yves**, "A model of ideological thinking," *Mimeo, Toulouse School of Economics*, 2016.

**Maier, Johannes**, "A Model of Reference-Dependent Belief Updating," *Mimeo, Ludwig Maximilians Universität München*, 2017.

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

**Mayraz, Guy**, "Wishful Thinking," *CEP Discussion Paper*, 2011, (1092).

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

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

**Merkle, Christoph and Martin Weber**, "True overconfidence: The inability of rational information processing to account for apparent overconfidence," *Organizational Behavior and Human Decision Processes*, 2011, *116* (2), 262–271.

**Möbius, Markus M. and Tanya S. Rosenblat**, "Why beauty matters," *The American Economic Review*, 2006, *96* (1), 222–235.

⎯ , **Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat**, "Managing Self-Confidence," *Mimeo, Stanford University*, 2014.

**Moore, Don A. and Paul J. Healy**, "The trouble with overconfidence," *Psychological Review*, 2008, *115* (2), 502–17.

**Neale, Margaret A. and Max H. Bazerman**, "The Effects of Framing and Negotiator Overconfidence on Bargaining Behaviors and Outcomes," *Academy of Management Journal*, 1985, *28* (1), 34–49.

**Niederle, Muriel and Lise Vesterlund**, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

**Pinker, Steven**, "Representations and decision rules in the theory of self-deception," *Behavioral and Brain Sciences*, 2011, *34* (01), 35–37.

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

**Schwardmann, Peter**, "Motivated health risk denial and the market for preventative health care," *Mimeo, Ludwig Maximilians Universität München*, 2017.

_ **and Joël J. Van der Weele**, "Deception and Self-deception," *Tinbergen Institute Discussion paper*, 2016, *012/2016.*

**Sedikides, Constantine, Vera Hoorens, and Michael Dufner**, "Self-Enhancing Self-Presentation: Interpersonal, Relational, and Organizational Implications," in Frederic Guay, Herbert Marsh, Dennis M. McInerney, and Rhonda G. Craven, eds., *Self-concept, motivation and identity: Underpinning success with research and practice*, Information Age Publishing, 2015, pp. 29–55.

**Sigall, Harold and Robert Gould**, "The Effects of Self-Esteem and Evaluator Demandingness on Effort Expenditure," *Journal of Personality and Social Psychology*, 1977, *35* (1), 12–20.

**Svenson, Ola**, "Are we all less risky and more skillful than our fellow drivers?," *Acta Psychologica*, 1981, *47* (2), 143–148.

**Taylor, Shelley E. and Jonathan D. Brown**, "Illusion and well-being: a social psychological perspective on mental health," *Psychological bulletin*, 1988, *103* (2), 193–210.

**Thaler, Richard H. and Hersh Shefrin**, "An economic theory of self-contol," 1981.

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

**Trivers, Robert**, "Deceit and self-deception," in "Social Evolution," Benjamin/Cummings, 1985, pp. 395–420.

_ , *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*, Basic Books, 2011.

**Von Hippel, William and Robert Trivers**, "The evolution and psychology of self-deception," *The Behavioral and brain sciences*, 2011, *34* (1), 1–16; discussion 16–56.

**Vrij, Aldert**, *Detecting lies and deceit: Pitfalls and opportunities*, John Wiley and Sons, 2008.

**Wang, Joseph Tao Yi, Michael Spezio, and Colin F. Camerer**, "Pinocchio's Pupil: Using Eye-tracking and Pupil Dilation To Understand Truth-telling and Deception in Games," *American Economic Review*, 2010, *100* (3), 984–1007.

**Weiss, Brent and Robert S. Feldman**, "Looking good and lying to do it: Deception as an impression management strategy in job interviews," *Journal of Applied Social Psychology*, 2006, *36* (4), 1070–1086.

**Zuckerman, Miron, Bella M. DePaulo, and Robert Rosenthal**, "Verbal and nonverbal communication of deception," in L. Berkowitz, ed., *Advances in Experimental Social Psychology, Vol 14*, New York: Academic Press, 1981, pp. 1–59.

# APPENDIX

## A    On-screen interview announcement (translated from German)

*(For a complete set of instructions in German and English, see `https: // www. dropbox. com/ s/ ftd6319tm56x6z0/ Instructions. pdf? dl=0`)*

## Interview instructions

In the next part of the experiment, the 16 participants will be split into two groups of eight. There will be four "employers" and four "contestants" in each group. Each participant will conduct a short "job interview" with all members of their group in the opposite role. Therefore, each participant will have four interviews.

**Based on your experimental ID, you were assigned the role of a contestant.**

Employers and contestants will have job inteviews that take the form of a face-to-face interaction. Contestants will make a verbal statement about the probability that they scored the highest or second highest test score (hereafter, the best two contestants will be referred to as "Top2").

Based on these statements, employers have to evaluate how likely it is that a contestant is truly in the top 2. For each employer, the two contestants with the highest evaluation are automatically hired.

After the hiring decision has been made, one of your four interviews is randomly selected. Your payment is based on this interview. To assure anonymity, it will not be revealed which interview determines the payment.

The payment of the employer is determined by how well they can evaluate contestants. Contestants' payment depends employers' evaluations.

**If you are hired in the randomly selected interview (that is, if you are in the top 2, according to the employer's evaluation), you will earn 15 euros in addition to your other earnings.**

As a contestant, you will therefore make more money, if you are better than the other contestants at convincing the employers of your superior relative performance.

# B    Bayesian updating regressions

Möbius et al. (2014) study updating behavior in an ego-relevant task. They find that subjects update conservatively, by responding too little to new information, and asymmetrically, by giving higher weight to good feedback than bad feedback. We investigate whether these updating biases exist in our sample, and whether they differ between subjects in the control and treatment condition. To control for prior beliefs and the distribution of signals between the two groups, we follow Möbius et al. (2014) in running logit regressions, which are based on a linearized version of Bayes' formula. The model is given by

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

Here, $\mu_{i,prior}$ and $\mu_{i,post}$ represent the prior and the posterior belief of person $i$ respectively. $\lambda_H = -\lambda_L$ is the log of 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 toward a posterior of 50 percent.

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 to 3 reflect the entire sample, whereas columns 4 to 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.

Our main findings are the following. First, we reject $\beta_H = 1$ for all specifications, and reject $\beta_L = 1$ for the full sample only. Thus, subjects in both groups are conservative on average. Second, in the restricted sample, $\beta_H < \beta_L$ with $p = 0.067$, so if anything, subjects respond more to bad signals than good signals. This contrasts with Möbius et al. (2014) who find the opposite, although there are several non-replications of this result by Coutts (2016) and Buser et al. (2016). Finally, we find no significant differences between the treatment (contestants) and control group. Although the treatment group seems to be somewhat less responsive to information in general, this tendency is not statistically significant.

|  | Full sample | | | Restricted sample | | |
|  | All | Employers | Contestants | All | Employers | Contestants |
|---|---|---|---|---|---|---|
| $\delta$ | 0.740*** | 0.784*** | 0.698*** | 0.803*** | 0.843** | 0.750** |
|  | (0.0584) | (0.0560) | (0.0985) | (0.0648) | (0.0605) | (0.114) |
| $\beta_H$ | 0.601*** | 0.621*** | 0.592*** | 0.781*** | 0.727*** | 0.857* |
|  | (0.0535) | (0.0576) | (0.0936) | (0.0425) | (0.0484) | (0.0788) |
| $\beta_L$ | 0.604*** | 0.648*** | 0.551*** | 0.897* | 0.903 | 0.874 |
|  | (0.0566) | (0.0785) | (0.0833) | (0.0572) | (0.0808) | (0.0840) |
| Observations | 280 | 139 | 141 | 208 | 108 | 100 |
| $R^2$ | 0.663 | 0.713 | 0.621 | 0.798 | 0.803 | 0.797 |

Table B.1: Belief updating regressions. Columns 1 to 3 reflect the entire sample, whereas columns 4 to 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 $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

|  | (1) | (2) |
|  | Full Sample | Restricted |
|---|---|---|
| $\delta$ | 0.784*** | 0.843** |
|  | (0.0560) | (0.0605) |
| $\beta_H$ | 0.621*** | 0.727*** |
|  | (0.0576) | (0.0484) |
| $\beta_L$ | 0.648*** | 0.903 |
|  | (0.0785) | (0.0809) |
| $\delta$ * contestant | -0.0857 | -0.0933 |
|  | (0.113) | (0.129) |
| $\beta_H$ * contestant | -0.0288 | 0.130 |
|  | (0.110) | (0.0924) |
| $\beta_L$ * contestant | -0.0978 | -0.0292 |
|  | (0.114) | (0.117) |
| Observations | 280 | 208 |
| $R^2$ | 0.665 | 0.800 |

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 $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

# C First stage regressions

Table C.1 shows the first stage regressions of the second stages we report in sections 6.2 and 6.3. Recall that we use our feedback signal as an instrument. The dummy variable *black ball* takes a value of 1 if the subject received good news. Since we can perfectly control for the state of the world, i.e. whether a subject was in the top 2, the conditional effect of the variable *black ball* is exogenous.

Other things equal, seeing a good signal boosts privately held beliefs by between 17 to 31 percentage points. This effect is highly significant in all first stage regressions. High F-Statistics provide further evidence for the strength of our instrument. Merely in column 5 do we see a

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Corresp. 2nd stage:* | *Table 3, C(3)* | *Table 3, C(4)* | *Table 4, C(3)* | *Table 4, C(4)* | *Table 4, C(5)* |
| **Dependent var.:** | Posterior | Posterior | Posterior | Posterior | Posterior |
| Black ball (d) | 20.68*** | 20.72*** | 17.63*** | 16.73*** | 31.07*** |
| | (4.279) | (3.647) | (4.540) | (4.053) | (9.275) |
| Top 2 | 11.01** | 0.931 | 16.24*** | 14.45*** | -5.759 |
| | (4.189) | (4.428) | (4.807) | (4.210) | (6.627) |
| Lying Aversion | | 4.113 | | | |
| | | (5.621) | | | |
| Female (d) | | -1.210 | | | |
| | | (3.613) | | | |
| Score | | 1.096*** | | | |
| | | (0.330) | | | |
| Round | | | | -0.212* | -0.0689 |
| | | | | (0.103) | (0.0944) |
| Message | | | | 0.479*** | 0.189 |
| | | | | (0.109) | (0.210) |
| Constant | 43.28*** | 32.26*** | 42.21*** | 8.433 | 33.42** |
| | (2.675) | (3.624) | (3.311) | (6.884) | (12.71) |
| Observations | 576 | 576 | 384 | 384 | 192 |
| $R^2$ | 0.346 | 0.408 | 0.368 | 0.468 | 0.393 |
| F-Statistic | 45.00 | 30.72 | 34.38 | 52.27 | 6.647 |

Table C.1: First stage regressions; ordinary least squares; standard errors (in parentheses) are clustered at the subject's group of 4 contestants and 4 employers; $^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

F-Statistic lower than 10. This column provides the first stage regression for the specification of column 5 in table 4, which was run exclusively over data from the lie detection treatment. The F-Statistic shows that our instrument in this subsample might be rather weak. This adds to the arguments in footnote 15 and casts doubt over the merit of interpreting second stage results from the LD treatment

# D  The incidence of lying

We define a lie as a message that is higher than the contestant's posterior belief. We therefore do not count untruthful understatements of one's beliefs as lies, although all results presented here are robust to defining a lie as any message that deviates from the posterior belief. The probit regressions in Table D.1 mirror the regressions in Table 3 except that the dependent variable is now the incident of a lie. Column 1 indicates that a higher posterior is associated with a lower likelihood of lying. Using the exogenous variation provided by the feedback signal, columns 2 and 3 suggest that the negative effect of confidence on the decision to lie is causal. Column 4 tells us that this causal effect is robust to the inclusion of controls. More lying averse subjects lie less, although this coefficient is not significant. Interestingly, women lie more in our experiment.

|  | Probit regressions | | IV probit regressions | |
|---|---|---|---|---|
|  | (1) | (2) | (3) | (4) |
| Dependent variable: | Lie | Lie | Lie | Lie |
| Posterior | -0.0368*** | | -0.0403*** | -0.0401*** |
|  | (0.00618) | | (0.0119) | (0.0116) |
| Round | 0.0249 | 0.0198 | 0.0241 | 0.0236 |
|  | (0.0292) | (0.0230) | (0.0295) | (0.0295) |
| Black ball | | -0.697*** | | |
|  | | (0.240) | | |
| Top 2 | | -0.479** | -0.0940 | -0.167 |
|  | | (0.218) | (0.355) | (0.357) |
| Lying aversion | | | | -0.662 |
|  | | | | (0.450) |
| Female | | | | 0.350* |
|  | | | | (0.192) |
| Constant | 2.641*** | 0.966*** | 2.891*** | 2.974*** |
|  | (0.480) | (0.193) | (0.662) | (0.711) |
| Observations | 576 | 576 | 576 | 576 |

Table D.1: Determinants of a contestant's decision to lie; in the IV probit regressions posterior is instrumented by *black ball*; standard errors (in parentheses) are clustered at the subject's group of 4 contestants and 4 employers; $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

# E    Robustness of results on employer evaluations

In column 1 of Table E.1 we add the variables *score*, *female*, *age* and *dominant* to the specification in column 3 of Table 4 and confirm that our results are robust to the inclusion of these of controls. We see that female contestants obtain lower evaluations from employers than male contestants We think that our experiment provides a promising interaction environment in which to study gender discrimination. However, an in-depth discussion of gender effects in the current dataset lies beyond the score of this paper.

In column 2 we test the robustness of specification 4 in Table 4 to the inclusion of further controls. The effect of messages on evaluations is robust. In column 3 we run regressions that allow for a non-linear impact of messages by including separate dummy variables for messages in each decile (0-10 percent, 11-20 percent etc.). This does not influence the size or significance of the posterior's coefficient. For lack of space, Table E.1 does not include the individual dummy variables. However, we note that there seems to be no punishment by the employers for extreme messages higher than 90% (boasts). That is, evaluations are monotonically increasing in the message a contestant states.

In columns 4 and 5 we test whether our results in Table 4 are robust to the inclusion of variables pertaining to a contestant's three competitors. We report reduced form OLS regressions with the instrument (i.e. the *black ball* dummy) as an independent variable. We then add the number of black balls a contestant's competitors saw as a valid measure of the competitors'

| Dependent variable: | IV regressions | | | OLS regressions | |
|---|---|---|---|---|---|
| | (1) Evaluation | (2) Evaluation | (3) Evaluation | (4) Evaluation | (5) Evaluation |
| Posterior | 0.346*** (0.122) | 0.271** (0.116) | 0.245** (0.118) | | |
| Top 2 | 0.541 (4.870) | 3.485 (4.293) | 4.309 (3.854) | -0.989 (5.619) | 3.016 (4.434) |
| Round | -0.792 (1.245) | -0.949 (1.249) | -0.915 (1.251) | -0.792 (1.282) | -1.066 (1.298) |
| Female (d) | -7.336*** (2.240) | -6.336*** (2.025) | -7.415*** (1.873) | | |
| Score | -0.247 (0.240) | -0.583** (0.246) | -0.526** (0.223) | 0.335 (0.294) | -0.102 (0.256) |
| Age | -0.126 (0.274) | -0.240 (0.239) | -0.275 (0.217) | | |
| Dominant | -4.202 (7.664) | -1.263 (7.192) | -5.328 (6.171) | | |
| Message | | 0.356*** (0.0845) | | | 0.465*** (0.0652) |
| Black ball (d) | | | | 5.563** (2.631) | 3.073* (1.670) |
| Others' black balls | | | | -1.854** (0.877) | -1.024* (0.557) |
| Others' scores | | | | -0.112 (0.0981) | 0.0340 (0.0855) |
| Others' messages | | | | | -0.155*** (0.0217) |
| Message dummies | No | No | Yes | No | No |
| Constant | 44.91*** (6.974) | 26.83*** (9.121) | 57.40*** (6.879) | 52.47*** (4.714) | 51.16*** (4.349) |
| Observations | 384 | 384 | 384 | 384 | 384 |
| $R^2$ | 0.0838 | 0.161 | 0.203 | 0.0552 | 0.198 |

Table E.1: Determinants of employer evaluations of candidates in the baseline condition; in the IV regressions posterior is instrumented by *black ball*; standard errors (in parentheses) are clustered at the subject's group of 4 contestants and 4 employers; $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

average confidence.

Results in column 4 indicate that a contestant gets a higher evaluation if her competitors are less confident. Moreover, the variables score and others' scores indicate that relative ability matters: a higher own score and lower scores by one's competitors both lead to higher evaluations. In column 5, we add a contestant's own message and the sum of messages sent by her competitors to the set of regressors. Again, results suggest that the non-verbal benefits of confidence are robust to the inclusion of variables pertaining to a contestant's competitors.