UTS UNIVERSITY OF TECHNOLOGY SYDNEY

UTS Business School –
Department of Economics
Dr Chau Chak Wing Building
14 – 28 Ultimo Road, Ultimo NSW 2007

T: +61 2 9514 7713
economics@uts.edu.au

Department of Economics
Working Paper Series

'Strategically delusional'

Alice Soldà [1, 4]
Changxia Ke [1]
Lionel Page [1, 3]
William von Hippel [2]

[1] Queensland University of Technology
[2] University of Queensland
[3] University of Technology Sydney
University of Lyon

No. 5 / 2019 **March 2019**

Strategically delusional

Alice Soldà[1,4], Changxia Ke*[1], Lionel Page[1,3], and William von Hippel†[2]

[1]Queensland University of Technology
[2]University of Queensland
[3]University of Technology Sydney
[4]University of Lyon

Abstract

We aim to test the hypothesis that overconfidence arises as a strategy to influence others in social interactions. We design an experiment in which participants are incentivised either to form accurate beliefs about their performance at a test, or to convince a group of other participants that they performed well. We also vary participants' ability to gather information about their performance. Our results provide, the different empirical links of von Hippel and Trivers' (2011) theory of strategic overconfidence. First, we find that participants are more likely to overestimate their performance when they anticipate that they will try to persuade others. Second, when offered the possibility to gather information about their performance, they bias their information search in a manner conducive to receiving more positive feedback. Third, the increase in confidence generated by this motivated reasoning has a positive effect on their persuasiveness.

JEL-Codes: C91, D03, D83

Keywords: Overconfidence, motivated cognition, self-deception, persuasion, information sampling, experiment.

*Soldà: alice.solda@qut.edu.au; Ke: changxia.ke@qut.edu.au; Page: lionel.page@uts.edu.au; von Hippel: billvh@psy.uq.edu.au.

†We are grateful to Joël van der Weele, Marie Claire Villeval, Benno Torgler, researchers at GATE in Lyon and Monash University in Melbourne and participants at the 12th and 13th Annual Australia New Zealand Workshop on Experimental Economics and the QUT brown bag seminar series for valuable comments on this paper. We also address special thanks to G. Gogoi for his help in programming the experiment and V. Desvernois for valuable research assistance. This research is funded by a Australian Research Council Discovery Project grant (ARC-DP No. DP160100942).

1 Introduction

For a long time, most standard theories in economics (Von Neumann and Morgenstern, 1953) and psychology (Maslow, 1950; Körding and Wolpert, 2004) have assumed that people collect and process information in a way that gives them an accurate perception of reality. But empirical research has shown that most people are overconfident regarding their own abilities. They believe that they are more skilled, more attractive or in better health than others (Svenson, 1981; Gabriel et al., 1994; Weinstein, 1980; Epley and Whitchurch, 2008). This overconfidence occupies a particular place in the collection of behavioural biases. The widespread presence of inflated self-beliefs presents an instance where people are not just making random mistakes, it appears instead as a systematic tendency to venture in self-serving delusions.

Many psychological studies have suggested that overconfidence arises because it has a consumption value (Taylor and Brown, 1988): people enjoy basking in the belief that they are better than they actually are. However, having mis-calibrated beliefs has a cost. The perception of our own abilities/attributes influences how we make decisions. Thus, the outcome of the decisions we make depends - in part at least - on being able to accurately evaluate our own abilities (Dunning et al., 2004). If overconfidence can lead to costly mistakes, an adequate explanation of its prevalence and persistence most likely requires for it to provide some benefits as well. In the present paper, we investigate the idea that overconfidence emerges as a strategy to gain advantages in social interactions.

We design an experiment to investigate whether overconfidence is more likely to emerge in situations where people anticipate the need to convince others about their performance and whether this strategically motivated overconfidence helps them be more persuasive. In service of these goals, we use a 2x3 design in which we manipulate participants' anticipation of strategic interactions and their ability to gather information about their performance. Participants were first asked to complete a general-knowledge test. At the end of the test, half of the participants were initially incentivised to give an accurate estimate of their performance ("Accuracy task"), while the other half were initially incentivised to write a persuasive essay to convince a group of other participants that they performed well on the test ("Persuasion task"). Participants who completed the Accuracy Task were next instructed to complete the Persuasion Task, and vice versa. Participants were not given any information about the second task until they finished the first one.

This design allows us to compare beliefs of participants who had previously attempted to persuade others of their strong performance to the beliefs of those who had not attempted to persuade others. We cast a light on how these beliefs are formed by examining how participants engage in information sampling when they have the freedom to self-select information,

compared to when they are either given no information about their performance or when the information is exogenously selected by the experimenter. The goals of this design is to study whether overconfidence emerges in anticipation of the need to persuade, whether biased information gathering facilitates the emergence of overconfidence, and whether overconfidence in turn facilitates persuasion.

We first conducted this experiment on the online platform of Amazon Mechanical Turk (in Study 1) and then replicated a subset of conditions in the controlled laboratory environment (in Study 2) as a robustness check for our findings. Across the two experiments we find that participants show greater confidence in their performance when they had anticipates the need to persuade others, although their performances are very similar. Furthermore, participants in this situation tend to engage in biased information sampling in a manner that facilitates self-confidence. Finally, participants with more confidence about their performance are also more successful at convincing others that they performed well. Overall, these results provide support for the idea that overconfidence can arise because of the strategic benefits it provides in social interactions.

Our paper broadly relates to the research on motivated beliefs (Bénabou and Tirole, 2016). The idea that beliefs can be used strategically to achieve higher payoffs has been formalized in economics by Bénabou and Tirole (2002). In particular, motivated beliefs have been shown as a way to alleviate the under-investment problem associated with present bias (Hong et al., 2018). These predictions are supported by empirical evidence (Puri and Robinson, 2007; Vialle et al., 2011). These first economics models of motivated beliefs focused on the *intrapersonal* advantages of holding positively biased beliefs. Another strand of the literature proposed instead that motivated beliefs can provide an *interpersonal* advantage. Agents who are overconfident may be more effective at signalling their ability in strategic situations where agents' true types cannot be observed (Bénabou, 2015; Bénabou and Tirole, 2016). Our paper contributes to this literature and presents evidence that overconfidence provides interpersonal benefits because it advantageously influence others in social interactions (Heifetz et al., 2007; Johnson and Fowler, 2011; von Hippel and Trivers, 2011). Building on Trivers (1976)'s hypothesis that self-deception evolved to deceive others more effectively, they propose that overconfidence plays an interpersonal role by enhancing others' perception of one's positive qualities. While bluffing may be sufficient to deceive others, self-deception could provide additional benefits. First, self-deception may alleviate the cognitive costs of deception (e.g. holding in your mind two competing versions of the reality, the one you believe in and the one you want to impart to others). Second, if the cognitive costs of deception generate visible cues of deception (e.g. being slower when generating arguments), self-deception could be a way to avoid such cues.[1] The idea that an evolutionary process can lead agents to generate mis-calibrated beliefs when one agent's beliefs can influence others

[1]Self-deception itself can incur other costs, like most other strategies. The decision whether to engage in it depends on how benefits weigh against costs.

in social interactions has been formalised by Heifetz et al. (2007). They show that there is no reason to expect that an evolutionary process should lead to agents with well-calibrated beliefs, as soon as agents' beliefs can influence others' perceptions in social interactions.

Our paper fits in the emerging strand empirical research investigating the interpersonal advantages of overconfidence. In a series of experiments, Anderson et al. (2012) provide evidence that individuals can attain a higher status in social interactions when they are overconfident. A similar effect was found by Murphy et al. (2017), who tracked 894 high school boys across two school years, and documented that overconfidence in sporting ability predicted increased popularity over time. These social benefits of overconfidence were extended to an economic experiment in the lab, in which Charness et al. (2018) found that participants are more likely to overstate claims of their strengths when it is optimal to deter opponents' entry in a contest. They also find evidence that these overstatements may not be fully self-aware.

The closest study to ours in the literature is Schwardman and van der Weele (2017) who find two results in support of the strategic-overconfidence hypothesis. First, people who expect to convince other participants of their own strengths are more overconfident than those who do not have this expectation. Second, by exogenously varying confidence via a noisy signal, they find that overconfidence enables participants to be more persuasive. Our design differs from Schwardman and van der Weele (2017)'s in various ways[2], but our results corroborate with both of their findings, providing further complementary evidence on the utility of strategic overconfidence. Beyond these findings, we show that participants engage in biased information search in a manner that is conducive to receiving more positive information about their performance. We also use our experimental setup to provide causal identifications of the different empirical links upon which the theory of von Hippel and Trivers (2011) relies. Namely: incentives to be persuasive in strategic interactions lead people to engage in motivated reasoning and biased information gathering, which makes them more confident and as a consequence more effective players in the strategic interactions. The results that we present in the remainder of the paper point to the fact that agents are not necessarily naively delusional as it is often suggested. Instead, they may often be "strategically delusional", forming unrealistic beliefs that give them an advantage in social interactions.

The remaining sections are organised as follows. Section 2 details the experimental design and hypotheses. In Section 3, we describe the procedures and display our main results. Section 4 concludes.

[2]Instead of persuading others using a written essay, Schwardman and van der Weele (2017) allow participants to undertake a face-to-face interview to persuade interviewers that they were in the top 50% of their group. In their experiment, two beliefs -one prior belief and one posterior belief after seeing a noisy signal about their own performance- were elicited before the interview with only half of the participants anticipating the interview. In comparison, we elicited beliefs in "accuracy task" under three different feedback conditions: no information, given information, and self-chosen information.

2 Experimental design and hypotheses

2.1 General Design

We design an experiment to investigate whether the propensity to engage in self-deception responds to the potential advantage of possessing overconfident beliefs in a given strategic interaction. We first present participants with a timed general-knowledge test in the form of a multiple-choice questionnaire. The test is composed of 30 questions of moderate difficulty, as indicated by the percentage of participants who answered the questions correctly in a previous MTurk experiment (Murphy et al., 2015).[3] Participants have 15 seconds to answer each question and do not receive any information regarding the following parts of the experiment at this point.

We then vary participants' anticipation of strategic interactions by asking participants to undertake two incentivised tasks sequentially: an accuracy task and a persuasion task. In the accuracy task, participants are incentivised to give their best guess about their absolute and relative performance. In the persuasion task, participants are incentivised to convince others that they performed well in the general-knowledge test. We design the experiment such that half of the participants join the "*Accuracy-first*" treatment (i.e., are presented with the accuracy task first, which is then followed by the persuasion task), while the other half of the participants join the "*Persuasion-first*" treatment and do both tasks in the reverse order. Participants are only informed of the nature of each task as they undertake it. This manipulation of task order enables us to observe participants engaging in the same set of tasks, but varies whether they are initially incentivised to form the most accurate beliefs about their performance or to form beliefs in a manner that would help them to convince others of their strong performance. This design enables us to investigate how participants form their beliefs and how their ability to persuade others is affected by the anticipation of strategic interactions.[4]

Accuracy task: Participants are asked how well they think they did in the test. They are informed that they will be rewarded for being accurate. Participants are asked to give both (i) an estimate of the number of correct answers they achieved in total (i.e., absolute performance) and (ii) an estimate of how well they did compared to other participants (i.e., relative performance). In (i), we ask participants to guess how many questions they believe

[3] A 31st question served as an attention check, and was designed such that as long as they paid attention they should be able to answer it correctly. In total, 16 people incorrectly answered this question and hence were dropped from analyses.

[4] Given the symmetric design of our study, in both the *Accuracy-first* and the *Persuasion-first* treatments, participants have to do the same tasks. The only difference is the initial expectation of having or not to persuade others when having to think about their performance.

they answered correctly in the test on a scale from 0 to 31 (thereby including the final, attention-check question, which was not identified as separate from the rest of the test). In (ii), we ask participants to guess their relative position in the performance distribution on a scale from 0 to 100%. The number they are asked to choose is the percentage of participants whom they believe they outperformed.[5] In both (i) and (ii), participants are rewarded if they give more accurate estimates than other participants: If participants' estimates are among the top 10% of the most accurate estimates, they receive $2; If participants' estimates are among the top 50% (but below the top 10% of the most accurate estimates), they receive $1.

Persuasion task: Participants are asked to convince a group of reviewers about the strength of their performance by writing a short essay.[6] They are told that the reviewers are another group of participants who did not take the test. Participants are informed that reviewers will be reading their essays and will rate them on: (i) how many questions they believe the participant answered correctly and (ii) how convincing they think the essay is. Each essay is reviewed by five different reviewers independently and participants' rewards in the persuasion task are based on comparisons of the average ratings given by all five reviewers.[7] Participants were told that if the reviewers' average assessment of their own absolute performance is in the top 10% of the (average) ratings within their comparison group, they receive $2; If the reviewers' assessment was in the top 50% (but below top 10%), they receive $1. Similarly, if participants' essays are rated in the top 10% of the most convincing essays within the comparison group, they receive $2; if participants' essays are rated in the top 50% (but below top 10%) of the most convincing essays, they receive $1.[8]

Information Feedback: To further investigate how participants form beliefs about their performance, before proceeding to the accuracy or persuasion task, we also vary the information available to participants with which they can form a belief about their performance. We use three different information conditions. In the *No Information* condition, participants do not receive any external feedback about their performance after the test.[9] In the *Given*

[5]The following examples were provided in the experimental instructions to facilitate participants' understanding: "if you think you did better than 3/4 of your fellow participants but worse than 1/4 of them, you will choose 75%. Conversely, if you think that you did better than 1/4 of your fellow participants but worse than 3/4, you will choose 25%." We provide two sets of full experimental instructions in Appendix C.

[6]Examples and analyses of these essays are provided in Appendix B.6.

[7]In Study 1, reviewers receive a fixed wage. In study 2, we use an incentive-compatible mechanism to reward reviewers' guess of the participants' score. More details about the reviewers' incentives are provided at the beginning of section 3.1 for Study 1 and 3.2 for Study 2.

[8]Our reward scheme has purposely an identical payoff structure in both *Accuracy-first* and *Persuasion-first* treatments to ensure consistent incentives across treatments. A different incentives scheme could affect how much attention participants give to the feedback and their selection of information in the *SCI* treatment. To ensure an identical payoff structure, we adopted a reward scheme based on relative performance in both the *Accuracy-first* and *Persuasion-first* treatments.

[9]For the sake of clarity, we refer to *Accuracy-first* (*Acc.1st*) and *Persuasion-first* (*Per.1st*) as "treatments" and *No Information* (*NI*), *Given Information* (*GI*), and *Self-Chosen Information* (*SCI*) as information "conditions".

Information condition, participants are shown 10 pre-selected questions and whether they answered them correctly or not. The 10 questions are selected to reflect the general level of difficulty of the 30 proper general knowledge questions.[10] These questions are the same for all participants. By virtue of our question sampling, the percentage of correct answers shown to participants in this feedback should predict the overall percentage of correct answers they are likely to receive for the entire test. In the *Self-Chosen Information* condition, participants are presented with the list of all the questions they faced during the test (excluded the last item that was used as an attention check). The questions appear in a random order and participants are told they are to select 10 questions of their choice to check whether they answered them correctly.[11] The feedback is displayed in exactly the same way in both *Given Information* and *Self-Chosen Information* conditions. The only difference between the two conditions lies in whether the questions were selected by the participants or the experimenter.

Factorial design: We cross the information conditions and the treatments in a 2x3 design represented in Table 1. Figure 1 provides the structure of this design in a flow chart.

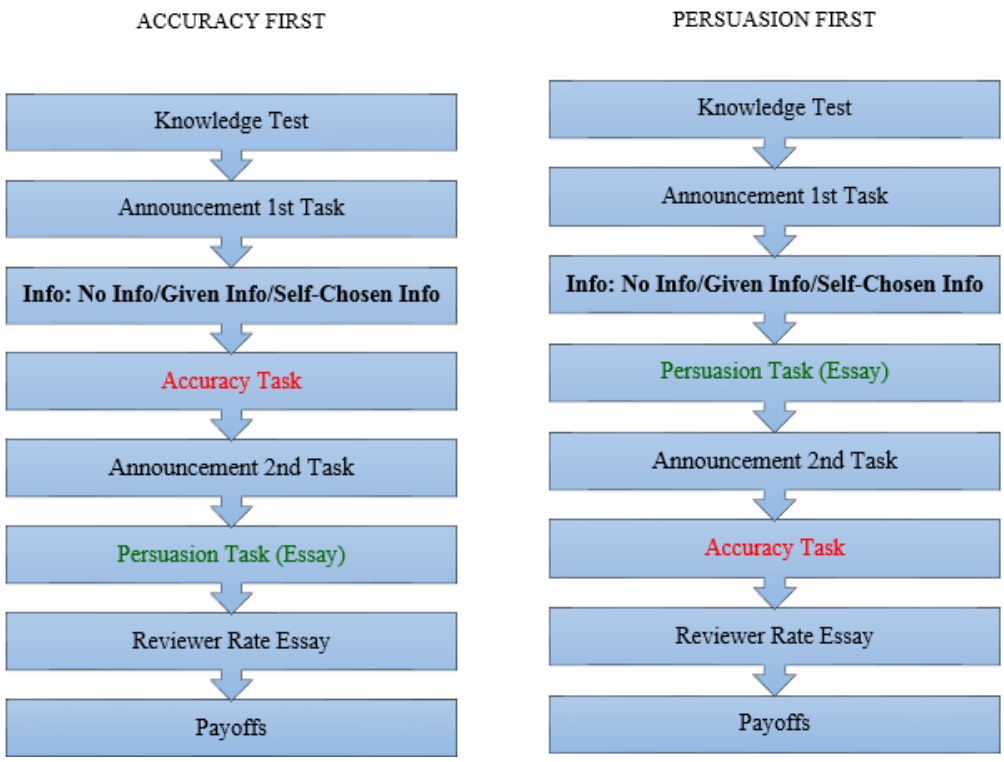

Figure 1: A timeline of the two different treatments.

At the end of the experiment, we recorded a range of individual characteristics to use as control variables in our analyses: participants' sex and age, as well as their dispositional overconfidence using the Over-Claiming Questionnaire (OCQ). [12]

[10]These 10 questions were chosen according to the accuracy rate of each question in an experiment run by Murphy et al. (2017) using the same knowledge test and a sample from the same population.
[11]Participants are not told what the correct answers are if their answers are wrong.
[12]Dispositional overconfidence is the idiosyncratic trait level of overconfidence, as opposed to situational

Table 1: A 2X3 experimental design.

	Information conditions		
Treatments	*No Information*	*Given Information*	*Self-Chosen Information*
Accuracy-first	*NI* x *Acc.1st*	*GI* x *Acc.1st*	*SCI* x *Acc.1st*
Persuasion-first	*NI* x *Per.1st*	*GI* x *Per.1st*	SCI x *Per.1st*

Notes: Table 1 displays the six cells of our 2X3 factorial design. NI stands for the *No Information*, GI for the *Given Information* condition, SCI for the *Self-Chosen Information* condition. *Acc.1st* refers to the Accuracy-first treatment and *Pers.1st* refers to the Persuasion-first treatment.

2.2 Hypotheses

If overconfidence provides a strategic advantage in social interactions, we may expect that its emergence is influenced by the existence of possible gains from being more confident. The two treatments we designed induce different incentives for participants to engage in motivated reasoning. In the *Accuracy-first* treatment, participants are initially incentivised to be accurate while in the *Persuasion-first* treatment participants are initially incentivised to be persuasive. If overconfidence facilitates persuasion about one's positive qualities/attributes, there are gains from forming more confident beliefs in the *Persuasion-first* treatment. Schwardman and van der Weele (2017) already suggests that this hypothesis might be true. Hence, we expect to observe that participants in the *Persuasion-first* treatments are more likely to be overconfident than participants in the Accuracy-first treatment about their own performance and their relative position in the group. Importantly, if their overconfidence is self-deceptive rather than just bluffing, then it should be carried forward to their judgments made on the following accuracy task, even though at that point accuracy is incentivised. In contrast, people in the *Accuracy-first* treatments should show minimal beliefs distortion, due to the benefits of evaluating their performance dispassionately in the initial accuracy task. This leads to our first hypothesis.

Hypothesis 1 (Strategic confidence) *Participants in the Persuasion-first treatment form more confident beliefs about their absolute and relative performance than participants in the Accuracy-first treatment.*

We also conjecture that it is easier for participants to distort their perception of their own performance when they have greater freedom in gathering information to form their beliefs. This hypothesis follows the insights from the literature on mental wiggle room in games where self-signalling can play a role (Grossman and Van Der Weele, 2017). In the *No Information* (*NI*) condition, participants have no outside information on which to base their self-assessment and hence can only rely on their subjective impression to form a belief about their performance. We conjecture that this condition gives them the least mental wiggle

overconfidence (here, the knowledge test). We use the 25-item versions of the test proposed by Bing and Davidson (2012). See Appendix B.1 for further details.

room to engage in motivated reasoning. In the *Given Information* (*GI*) condition, even though participants are given objective information that is representative of their overall performance, individuals may overweight good news (and underweight bad news) compared to Bayes rules when receiving feedback about their own performance/abilities.[13] Hence, we conjecture that GI condition could give them slightly more freedom to engage in motivated reasoning.

Finally, in the *Self-Chosen Information* (*SCI*) condition, participants are given the most freedom to engage in motivated reasoning and form the most favourable views about their performance. Specifically, we expect participants to gather information in a biased way when given the opportunity to choose freely and when there are strategic incentives for being more confident. We therefore expect participants in the *Persuasion-first* treatment to select more questions they believe they have answered correctly compared to participants in the *Accuracy-first* treatment, because doing so will give them more positive feedback, helping them to form more confident beliefs and be more persuasive.

In summary, we hypothesise that participants will become more confident when they are incentivised to persuade versus when they are incentivised to be accurate and the discrepancy between the two treatments should increase when participants have the opportunity to shape the feedback they receive from the test. This leads to the following two hypotheses.

Hypothesis 2 (Selective/biased information search) *Participants in the Self-Chosen Information condition (SCI) engage in selective/biased information search in a manner that is conducive to form more confident beliefs (i.e., by sampling easier questions) when in Persuasion-first treatment.*

Hypothesis 3 (Mental wiggle room & strategic confidence) *The difference between Persuasion-first treatment and Accuracy-first treatment (in beliefs about absolute and relative performances) will increase from No Information (NI) condition to Given Information (GI) condition, and increase further in Self-Chosen Information (SCI) condition.*

Furthermore, holding more confident beliefs will provide an advantage to participants in their effort to convince reviewers that they did well on the test. We thus propose the following hypothesis.

Hypothesis 4 (Effectiveness of strategic confidence) *More confident beliefs generated through motivated reasoning will help participants be more successful at persuading reviewers to rate them favourably.*

[13] See for instance Eil and Rao (2011); Mobius et al. (2014). More recent studies have however sometimes failed to replicate this result or found the opposite Buser et al. (2018)

We pre-registered this design and hypotheses on Open Science Framework.[14] Figure 2 below summarises how our main hypotheses fall within von Hippel and Trivers (2011)'s theory.

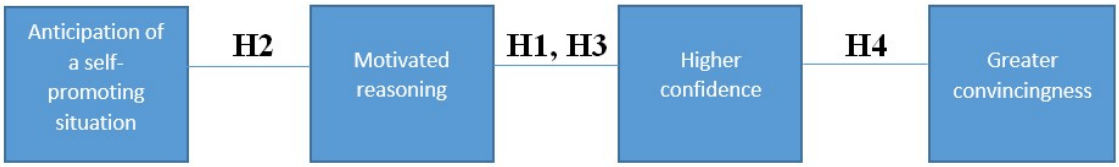

Figure 2: A summary of von Hippel and Trivers (2011)'s theory.

3 Data and Results

3.1 Study 1: MTurk Experiment

We first implemented our 2x3 factorial design online via Amazon MTurk, where 600 individuals participated in the main part of the experiment and 3000 others participated as reviewers. The main participants were randomly allocated to one of the six treatment-conditions. Most of the participants finished the tasks within 35 minutes, and on average they earned \$3.25 (s.e. $= 0.84$) plus a fixed payment of \$2. Reviewers in this study only received one essay each and were paid a fixed amount of \$0.25 for an average of 5 minutes spent on it.[15] Because our experiment is based on a knowledge test validated on US participants, only native English speakers from the USA were invited to join our study. The experiment was programmed using Qualtrics.

In this study, we told participants that they would be compared against their fellow MTurk workers who participated in the experiment (when payments were calculated at the end of the whole experiment).[16] They were also told that reviewers (after reading the essay) would be asked to estimate the total number of questions they author answered correctly and rate how convincing they think their essay is.[17]

3.1.1 Results

Figure 3 displays the summary statistics when all information conditions are pooled together. Each bar represents the mean of each main variable of interest (with confidence intervals).

[14]The hypotheses' statements were improved for exposition purposes. The pre-registration can be found at the following link: `https://osf.io/z5266/?view_only=e26aeeef9d794b9c8a91887d57323c53`

[15]All payments are in USD. Participants were allowed to keep the experiment window open on their internet browser for a few hours without disconnection, so several participants took much longer.

[16]The data collection process on MTurk was continuous and not split into small "sessions".

[17]Participants were not given information about how the reviewers' payment would be calculated.

The top panels display participants beliefs about their absolute and relative performance, and the reviewers' guess of their performance. The bottom panels show the corresponding biases in these beliefs. We use the difference between participants' beliefs about their absolute performance and their actual performance as a measure of "*Overconfidence*", and the difference between participants' beliefs about their relative performance (i.e, the percentage of people they have outperformed) and their actual relative performances to measure "*Overplacement*".[18] We also use the difference between reviewers' guesses of participants' performance and their actual performance to measure the biases in reviewers' estimates. The blue/left bar (in each pair-wise comparison) represents the value for the *Accuracy-first* treatment and the red/right bar represents the value for the *Persuasion-first* treatment. On top of the bars, mean values, the mean treatment differences, and the p-values from two-sided Mann-Whitney rank-sum tests are also provided in Figure 3.[19]

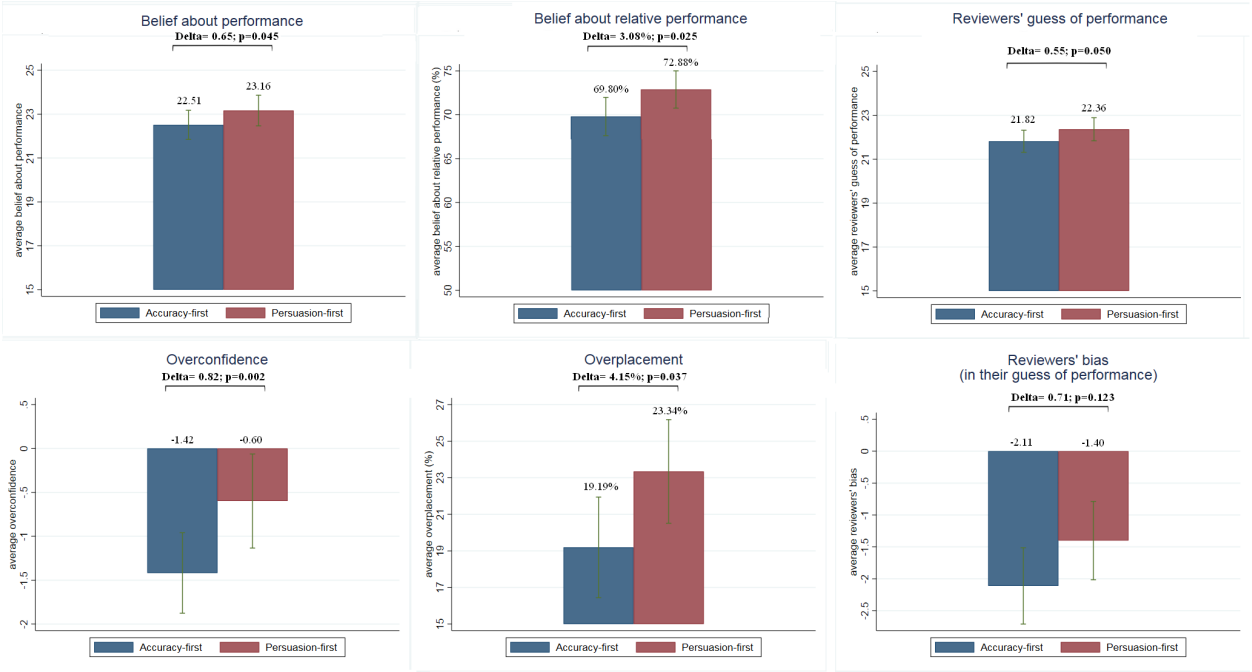

Figure 3: A summary of our main variables of interest comparing *Accuracy-first* treatment (in blue bars) and *Persuasion-first* treatment (in red bars) with mean values, treatment differences and p-values (from two-sided Mann-Whitney rank-sum tests) indicated on top. All information conditions in Study 1 are pooled together.

Figure 3 shows the following regularities on participants' and reviewers' beliefs and biases in these beliefs. First, we find participants in the *Persuasion-first* treatment overall hold more confident beliefs about their performances - their beliefs are around 0.65 higher (p=0.045), even though their actual performances are very similar across treatments (23.93

[18] Because we only asked reviewers to guess participants' absolute performance and not their relative performance), and participants were informed of this fact, it is possible that "*Overconfidence*" is a more salient measure (than "*Overplacement*") to participants.

[19] A full summary table with mean values and standard errors for all the variables is also provided in Table A.1 in Appendix.

vs. 23.77, p=0.763). If we further examine biases in beliefs, we find that overall the "*Overconfidence*" measure is 0.82 higher (p=0.002) in the *Persuasion-first* treatment, compared to the *Accuracy-first* treatment. Second, beliefs about their relative performances show similar patterns. Participants in the *Persuasion-first* treatment believe that they have on average outperformed 72.88% of the other participants, while those in the *Accuracy-first* treatment believe they have on average outperfomed 69.80% of the people. The treatment difference (3.08%) is significant at 5% (p=0.025). When the "*Overplacement*" measure is assessed, the average value is 4.15 percentage points higher (23.43% vs. 19.19%, p=0.037).[20] Finally, reviewers' guess of performances are 0.55 higher in the *Persuasion-first* than in the *Accuracy-first* treatment (22.36 vs. 21.82, p=0.005), however, when we look at the difference between reviewers' guess of performances and participants' actual performance, the differential bias in reviewers' guess is no longer significant (0.71, p=0.123). These results together support for Hypothesis 1 that participants will show strategic overconfidence when motivated to persuade, but they do not address the possible causal role of participants' overconfidence on persuasiveness.

Result 1 (Strategic confidence) *Participants in the Persuasion-first treatment form more favourable beliefs about their absolute and relative performance than participants in the Accuracy-first treatment, even though their actual performances are similar.*

If the expectation of having to convince others leads to strategic self-deception, we would expect it to be reflected not only in participants' final beliefs but also indirectly in how participants choose to process information in order to form favourable beliefs. There are two conditions where participants observe information about their performance in our experiment, the *Given Information* (*GI*) condition where participants do not choose what information they receive and the the *Self-Chosen Information* (*SCI*) condition where participants choose the questions for which they want to receive feedback. We expect that participants in the *Persuasion-first* treatment will selectively choose questions they are more likely to have answered correctly (compared to those in the *Accuracy-first* treatment) in order to facilitate positive feedback. This approach may allow them to sustain a more positive belief about their performance. Our measure of "Feedback" (i.e., the proportion of correct answers contained in the 10 pre-selected questions) presented in Figure 4 (left panel in the last row) is consistent with this prediction. We observe that participants in the *Self-Chosen Information* condition chose a set of questions with on average 12% more correct answers (79.69% in the *Persuasion-first* treatment *vs.* 67.4% in the *Accuracy-first* treatment, $p < 0.001$). In contrast, by virtue of the experimental design, there should be no difference in performance on this measure in the *Given Information* condition between the *Persuasion-first* and *Accuracy-first* treatments, as only random variation would cause performance to vary on the same 10

[20] Overall, our participants slightly underestimate their absolute performance but substantially overestimate their relative performance. These results are consistent with previous studies that find underconfidence and overplacement emerges jointly depending on task difficulty (Moore and Healy, 2008; Larrick and Soll, 2007). Hence, we will only focus on the treatment differences in the following analysis.

questions across the two treatments.

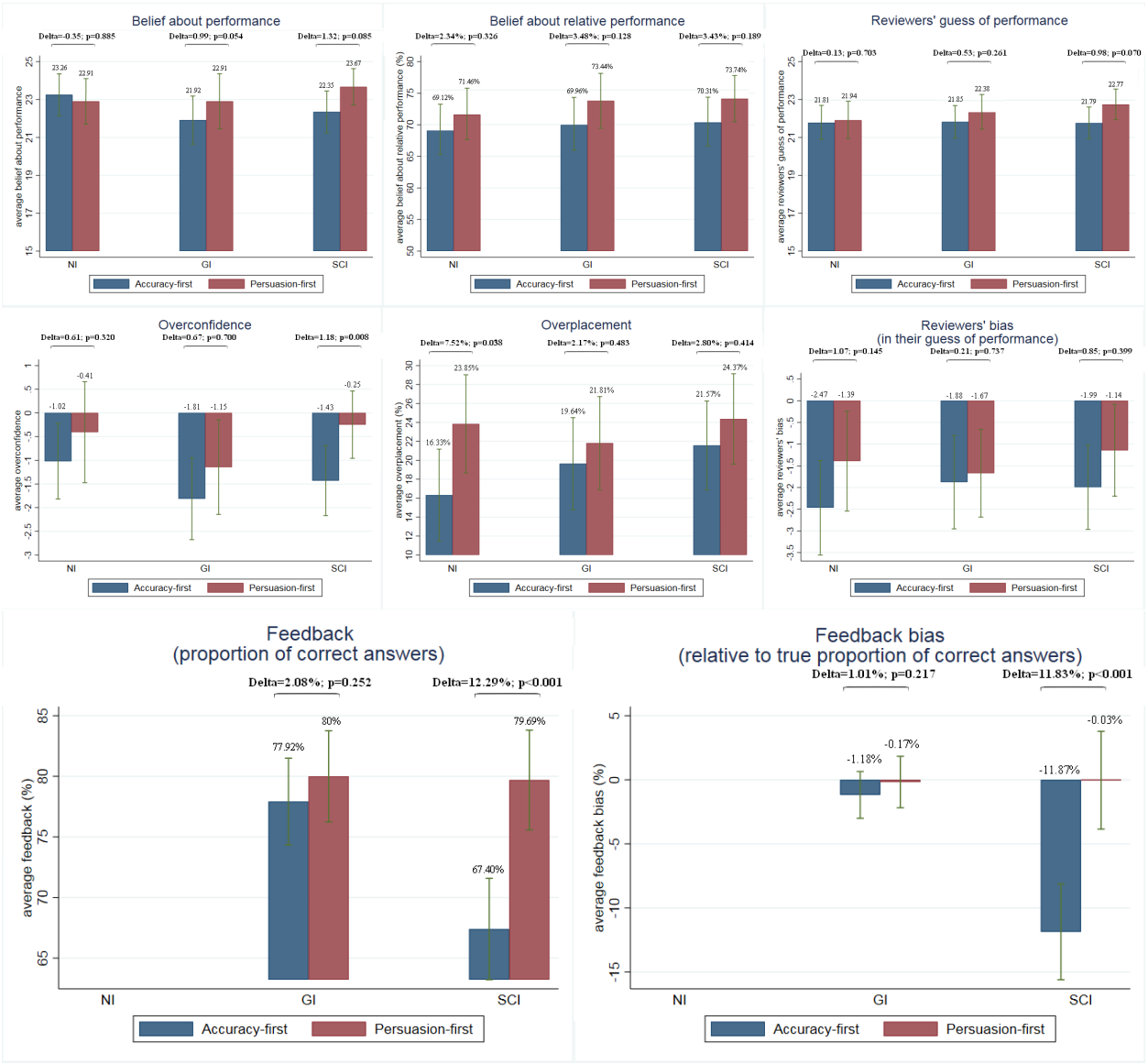

Figure 4: A summary of our main variables of interest comparing *Accuracy-first* treatment (in blue bars) and *Persuasion-first* treatment (in red bars) across information conditions in Study 1 (with mean values, treatment differences and p-values (from two-sided Mann-Whitney rank-sum tests) indicated on top.

We also calculate a "*Feedback bias*" as the difference between the percentage of correct answers observed in the feedback questions and the percentage of correct answers received for the whole test. A positive feedback bias means that participants received feedback with a higher percentage of correct answers than the actual proportion of correct answers they achieved overall. If the sampling is unbiased, the expected proportion of correct answers revealed through the feedback should be equal to the actual proportion of correct answers. In that case, the "*Feedback bias*" will be equal to 0. As shown in bottom right panel in Figure 4, we find no significant difference in feedback bias in *Given Information* (*GI*), as expected ($p = 0.217$). In contrast, we find a significant positive difference in the *Self-Chosen Information*

(*SCI*) condition ($p < 0.001$). This difference indicates that on average, participants in the *Persuasion-first* treatment chose to sample questions where they were more likely to have been right than those in the *Accuracy-first*, as suggested by Hypothesis 2.[21]

Result 2 (Biased information search) *Overall, participants in the Persuasion-first treatment sample questions that are 12 percentage points easier than participants in the Accuracy-first treatment.*

In Hypothesis 3 we propose that in the situation with more mental wiggle room (i.e. *Self-Chosen Information*), participants may be able to form more confident beliefs. To test this hypothesis, we examine the treatment effect within each condition. Figure 4 presents the treatment comparisons on the same main variables of interest as in Figure 3 in the first two rows, but within each condition separately. We find that the participants' beliefs on performance and relative performance are almost always higher in the *Persuasion-first* treatment than in the *Accuracy-first* treatment (except beliefs about performance in *NI* condition). However, most differences are not significant. Results on beliefs about their relative performances and the measure of overplacement also do not provide evidence for the effect of the treatments across different information conditions. When we use simple OLS regressions and pair-wise tests on the estimated treatment effects to directly test Hypothesis 3, we did not find any significant results either.

Result 3 (Mental wiggle room & strategic confidence) *Inconsistent with Hypothesis 3, we find no clear evidence that the difference in participants' beliefs about their absolute and relative performances between Persuasion-first treatment and Accuracy-first treatment becomes significantly higher when they are given more freedom to select their feedback.*

Overall, Study 1 provides evidence for strategic use of overconfidence in social interactions (Hypothesis 1) and for biased information sampling (Hypothesis 2). Although Study 1 doesn't provide evidence for Hypothesis 3 and 4, it remains possible that the effect of the anticipation of strategic interactions may be stronger in the *SCI* condition, when participants can actively engage in selective information search.[22] Hence, it is possible that for self-deceptive overconfidence to emerge, sufficient mental wiggle room may be necessary. To assess this possibility, and to examine the robustness of our results, we report a replication of the SCI condition in the controlled environment of the laboratory in the next subsection. We then use our experimental results to identify the causal effect of information sampling on confidence and the causal effect of endogenously affected confidence on participants' persuasiveness, to assess the empirical links in von Hippel and Trivers (2011)'s theory.

[21] An interesting finding in this study is that participants in *Accuracy-first* treatment (who are motivated to be as accurate as possible) sample 11.84% more difficult questions than those in the *Persuasion-first* treatment.

[22] This is shown in the treatment comparisons on our primary measure of confidence and overconfidence in SCI condition (see the first two left panels in Figure 4).

3.2 Study 2: replication in the laboratory

Online experiments using MTurk appear to be reliable (Arechar et al., 2018), but the MTurk environment is not as controlled as in the lab. Moreover, the incentives on MTurk are quite low, which may impair the motivation of the participants. Although we paid our participants on average more than twice the standard hourly rate typically available on MTurk, the reviewers were not incentivised to be accurate in their guess of the participants' score in Study 1. For these reasons, to ensure the reliability of the first study's results, we reproduce the *Self-Chosen Information* condition in a controlled laboratory environment in Study 2 at Queensland University of Technology (QUT). We recruited 100 QUT students for the main part of the experiment (50 in each treatment) and another 100 QUT students to participate as reviewers. At the end of the experiment, both the participants and the reviewers were paid in cash. The experiment was programmed using o-Tree (Chen *et al.*, 2016). The experiment took on average 45 minutes and the average payoff was \$11.20 (s.e. = 3.58) for the main participants and \$8.70 (s.e. = 4.08) for the reviewers.

To implement the experiment in the laboratory we made some minor changes. First, we adapted 4 questions from the general knowledge test from Study 1 to make them more suitable for non-Americans. Second, participants and reviewers were both invited to the lab at the same time. Upon arrival, the participants and the reviewers were separated into two different rooms on different floors. Each reviewer received four to six essays to ensure five independent assessments for each essay. Third, we incentivised the reviewers in the accuracy of their guesses about participants' score.[23] We did not provide incentives for the second question on convincingness given that it is purely subjective.[24] Fourth, we asked participants to complete the 25-item version of the OCQ at the end of the experiment to avoid any impact OCQ might have on the main experiment.[25] Finally, we made clear in the instructions that to calculate participants earnings (both for the accuracy and persuasion tasks), we would compare them to other participants in the same experimental session.

[23] One essay was randomly drawn for each reviewer's payment according to the following rule: they would receive \$10 if their guess of the participant's score is equal to the participant's score or deviates from that score by only one item. They would receive \$8 if their guess deviates by two items. They would receive \$4 if their guess deviates by three items. They would receive \$2 if their guess deviates by four or five items and they would receive nothing if their guess deviates by more than five items.

[24] Reviewers' ratings on convincingness is an auxiliary measure we use to ensure participants not only think about persuading reviewers that their performance is strong but also try to be as convincing as possible. Summary statistics provided in Table A.1 in the appendix show that there is no significant difference on reviewers' ratings of convincingess across treatments.

[25] Comparing Study 1 and 2, we find that it does not make a difference whether we place it at the beginning or the end of the experiment.

3.2.1 Results

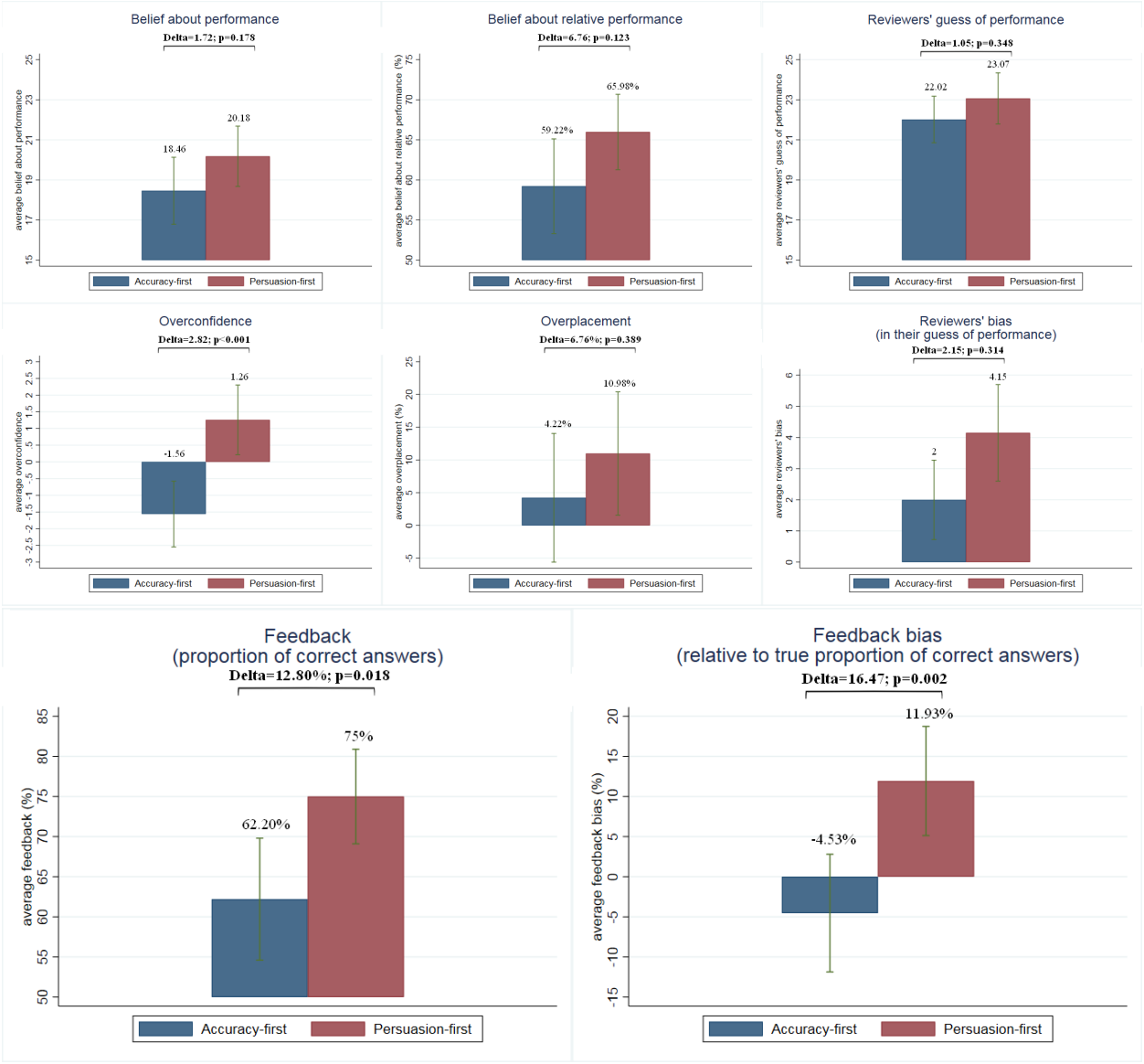

Figure 5: A summary of the mean values of our main variables of interest comparing Accuracy-first (blue bars) and Persuasion-first (red bar) in Study 2 (with mean differences between treatments and p-values indicated).

Figure 5 displays the summary statistics for Study 2. The measures we use are exactly the same as in Figure 4.[26] We find that the results from Study 2 are generally very similar to those in Study 1. We observe no statistically significant difference in performances across treatments (20.02 *vs.* 18.92, $p = 0.200$), but the overall performance is slightly lower than in Study 1. In the *Persuasion-first* treatment, participants on average hold more favourable beliefs about their absolute performance (1.72 higher) and relative performance (6.76 percentage points higher). However, these differences are not significant (two-sided MW t-tests: p =0.178 and p = 0.123, respectively). When we examine biases in these beliefs, we find

[26] A similar summary of the mean values and standard errors for all the variables is displayed in Table A.2 in Appendix as well.

that the "*overconfidence*" measure is 2.82 higher in the *Persuasion-first* treatment (two-sided MW rank-sum tests: $p < 0.001$) but the "*overplacement*" measure remains nonsignificant ($p = 0.389$).

We also find that participants on average sample relatively easier questions in the *Persuasion-first* treatment and the proportion of correct answers (75%) is 12.8 percentage points higher than that in the *Accuracy-first* treatment ($p = 0.018$). The treatment comparison on "*Feedback bias*" measure is also significant at 1% (11.93% vs. -4.53%). In contrast to the SCI condition in study 1 in which "Feedback bias" is mainly driven by participants sampling more difficult questions (-11.87%) in *Acc.1st* treatment, the discrepancy in study 2 is driven by deviations to opposite directions from both treatments. Namely, not only participants in *Acc.1st* treatment sample more difficulty questions (-4.53%) to be accurate, but also participants in *Per.1st* treatment sample easier questions (11.93%) to get more positive feedback.

Finally, we find that the impact of the *Persuasion-first* treatment on the reviewers' estimates of participants' scores is also positive (1 item more), but not significant. The reviewers' bias in their guess of performances is also higher in *Persuasion-first* treatment (4.15 vs. 2), but the difference is again not significant ($p = 0.314$).

In summary, the additional results from Study 2 are largely consistent with the results of Study 1 and when taken together, these two studies provide evidence of both the existence of strategic overconfidence (Hypothesis 1) and selective information sampling (Hypothesis 2).

3.3 Causal identification: information sampling on confidence

As stated in Hypothesis 2, the observed differences in beliefs between participants in the *Accuracy-first* and *Persuasion-first* treatments under *SCI* condition is likely to be facilitated by the tendency for people to bias their collection of information. Having established that participants in the *Persuasion-first* treatment indeed sampled information in a self-serving way, in this section, we further investigate how the bias in sampled information affects participants' beliefs.

If the hypothesis made by von Hippel and Trivers (2011)–that participants sample information in a self-serving way to inflate their perceptions of their own performance–holds, we should observe that a higher proportion of correct answers in the feedback has a positive effect on participants' beliefs. In order to test this hypothesis, we could regress participants' beliefs about their performance on the feedback received (see Tables A.4 and A.5 in the appendix). However, since participants are allowed to sample information at their own discretion, their feedback is likely to be endogenous. To address this issue, we instrument "*Feedback*" variable by a dummy variable that equals 1 if the participant was in the

Persuasion-first treatment and 0 if the participant was in the *Accuracy-first* treatment.[27] Table 2 reports the 2SLS regressions of participants' beliefs on "*Feedback*" (instrumented by the treatment dummy), controlled for actual performance. We use beliefs on absolute performance as the dependent variables in models (1), (3) and (5) and beliefs on relative performance in models (2), (4) and (6).[28]

Table 2: Causal identification of the effect of information sampling on beliefs about performance and relative performance.

Dep. Var:	*SCI* (MTurk)		*SCI* (lab)		*SCI* (MTurk +lab)	
Beliefs about	perf.	relative perf.	perf.	relative perf.	perf.	relative perf.
	(1)	(2)	(3)	(4)	(5)	(6)
Feedback	0.100**	0.259	0.644**	0.195**	0.132***	0.400***
	(0.047)	0.179)	(0.294)	(0.081)	(0.039)	(0.149)
Performance	0.672***	1.824***	0.703***	1.347**	0.670***	1.572***
	(0.128)	(0.491)	(0.158)	(0.570)	(0.087)	(0.329)
Constant	-0.356	9.484	-7.733	-7.789	-2.676	4.911
	(1.834)	(2.061)	(4.593)	(5.366)	(1.790)	(1.978)
First-stage F-stat	37.21	14.25	6.94	12.78	39.45	33.18
Observations	197	197	100	100	297	297

Notes: Column (1) to (6) report 2SLS regressions with standard errors in parentheses. *Feedback* is instrumented by the treatment dummy and is the *proportion* of correct answers contained in the sampled questions. Columns (1) to (2) shows the results from Study 1. Columns (3) and (4) shows the results from Study 2. Columns (5) and (6) shows the results for pooled observations from both studies. *** p <0.01, ** p <0.05.

Models (1) to (6) show that an increase in the proportion of correct answers in the feedback increases participants' beliefs about their absolute and relative performance and the effect is significant at the 5% (1%) level for each individual study (pooled studies), with the exception of model (2).[29]

3.4 Causal identification: the effect of confidence on persuasiveness

The key hypothesis motivating our study is that overconfidence can arise strategically as people attempt to be more persuasive in social interactions. The above results provide evidence that when people anticipate a need to be persuasive, they form more favourable

[27]This instrumentation requires the assumption that the *Persuasion-first* treatment affects participants beliefs only through the feedback received as a result of active biased sampling. It is a restrictive assumption and our results need to be interpreted in that light.

[28]Table A.7 in the Appendix also shows the same regressions with more control variables (sex, age, OCQ).

[29]Since the proportion of correct answers in feedback is only indirectly linked to participants' percentiles, it is expected that the effect of this feedback on beliefs about a participant's relative performance would be less significant than on beliefs about her actual performance.

beliefs through biased information search. In this subsection, we use the randomness of the treatment assignment to instrument participants beliefs to further explore how participants' beliefs affect their ability to persuade others.[30] Table 3 reports the 2SLS regressions of participants' persuasiveness on participants' beliefs about their performance (instrumented by a treatment dummy). Persuasiveness is measured by reviewers' average guessed scores and we control for participants' performance. We use beliefs on absolute performance in models (1), (3) and (5) and beliefs on relative performance in models (2), (4) and (6).[31]

Table 3: Causal identification of the effect of participants' beliefs about performance and relative performance on persuasiveness.

Dep. Var:	*SCI* (MTurk)		*SCI* (lab)		*SCI* (MTurk + lab)	
Persuasiveness	(1)	(2)	(3)	(4)	(5)	(6)
Beliefs-Perf.	0.794	–	0.514**	–	0.629***	–
	(0.534)		(0.257)		(0.193)	
Beliefs-Rel.Perf.	–	0.306	–	0.155**	–	0.207***
		(0.252)		(0.061)		(0.069)
Performance	-0.467	-0.492	-0.153	-0.001	-0.352	-0.256
	(0.369)	(0.620)	(0.223)	(0.157)	(0.179)	(0.182)
Constant	15.14	11.95***	15.60	12.83***	16.55***	13.85***
	(2.020)	(2.437)	(1.727)	(2.305)	(0.942)	(1.158)
First-stage F-stat	107.64	48.71	73.80	20.69	213.27	80.21
Observations	197	197	100	100	297	297

Notes: Table 3 reports 2SLS regressions for *SCI* conditions only with standard errors in parentheses. Participants' beliefs are instrumented by the treatment dummy. Columns (1) and (2) shows the results for observations from Study 1. Columns (3) and (4) shows the results for observations from Study 2. Columns (5) and (6) shows results for observations from pooled data. *** p <0.01, ** p <0.05.

Results from Table 3 show that participants' beliefs about their performance have a positive effect on the reviewers' average guess of participants' scores. These effects are not significant for Study 1 in both models, but they are significant at 5% in Study 2 for both models. When pooling the data of *SCI* condition from both studies, the effect is significant at the 1% level in both models. Table A.8 in the Appendix displays similar results when controlling for more variables. These results provide evidence that an increase in participants' confidence makes them more successful at persuading others that they performed well.

Result 4 (Effectiveness of strategic confidence) *Participants holding higher beliefs (generated via motivated reasoning) are more successful at convincing others that they did well*

[30] Our identification assumption is that the *Persuasion-first* treatment only has an effect on persuasiveness via its effect on confidence. It is a restrictive assumption and our results need to be read in that light. We provide supporting evidence in Appendix B.6, by looking at the content of the participants' essays.

[31] Table A.8 in the Appendix also shows the same regressions with more control variables (sex, age, OCQ).

in the test.

4 General Discussion and Conclusion

In the current research we tested the hypothesis that overconfidence emerges as a strategy to gain an advantage in social interactions. In service of this goal, we conducted two studies in which we manipulate participants' anticipation of strategic interactions and also the type of feedback they receive.

Our findings from both studies support the idea that self-beliefs respond to variations in the incentives for overconfidence. In our experiments, participants were put in situations where they could receive higher payoffs from persuading other players that they performed well in a knowledge test. We observe that their confidence in their performance increased in such situations.

Consistent with the interpretation that overconfidence is induced by strategic motivated reasoning, we observed that when given the freedom to choose their feedback, participants chose to receive feedback they could reasonably expect to be more positive. This choice, in turn, helped them form more confident beliefs about their performance. Finally, we find evidence that exogenous variations in confidence generated by the experiment led to variations in persuasiveness, with more confident players perceived as better performers.

These results support the hypothesis that overconfidence can grant interpersonal advantage. This hypothesis may help explain why overconfidence is so prevalent despite the obvious costs of having miscalibrated beliefs. Future research should investigate whether the type of interpersonal advantage observed in the context of this experiment can also be observed in different strategic contexts (e.g. negotiation, competition).

5 References

Anderson, C., Brion, S., Moore, D. A., Kennedy, J. A., 2012. A status-enhancement account of overconfidence. J Pers Soc Psychol 103 (4), 718–35.

Arechar, A. A., Gächter, S., Molleman, L., 2018. Conducting interactive experiments online. Experimental economics 21 (1), 99–131.

Bénabou, R., Tirole, J., 2016. Mindful economics: The production, consumption, and value of beliefs. Journal of Economic Perspectives 30 (3), 141–64.

Bing, M. N., Davidson, H., 2012. Measuring faking using the overclaiming instrument. River Cities Industrial and Organizational Psychology Conference, Chattanooga, TN.

Buser, T., Gerhards, L., van der Weele, J., 2018. Responsiveness to feedback as a personal trait. Journal of Risk and Uncertainty 56 (2), 165–192.

Bénabou, R., 2015. The Economics of Motivated Beliefs. Revue d'économie politique 125 (5), 665–685.

Bénabou, R., Tirole, J., 2002. Self-confidence and personal motivation. The Quarterly Journal of Economics 117 (3), 871–915.

Bénabou, R., Tirole, J., 2016. Mindful Economics: The Production, Consumption, and Value of Beliefs. Journal of Economic Perspectives 30, 141–164.

Charness, G., Rustichini, A., Van de Ven, J., 2018. Self-confidence and strategic behavior. Experimental Economics 21 (1), 72–98.

Dunning, D., Heath, C., Suls, J. M., 2004. Flawed self-assessment: Implications for health, education, and the workplace. Psychological science in the public interest 5 (3), 69–106.

Eil, D., Rao, J. M., 2011. The good news-bad news effect: asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics 3 (2), 114–38.

Epley, N., Whitchurch, E., 2008. Mirror, mirror on the wall: Enhancement in self-recognition. Personality and Social Psychology Bulletin 34 (9), 1159–1170.

Gabriel, M., Critelli, J., Ee, J., 1994. Narcissistic illusions in self-evaluations of intelligence and attractiveness. Journal of Personality 62 (1), 143–155.

Grossman, Z., Van Der Weele, J. J., 2017. Self-image and willful ignorance in social decisions. Journal of the European Economic Association 15 (1), 173–217.

Heifetz, A., Shannon, C., Spiegel, Y., 2007. What to maximize if you must. Journal of Economic Theory 133 (1), 31–57.

Hong, F., Huang, W., Zhao, X., 2018. Sunk Cost as a Self-Management Device. Management Science.

Johnson, D. D., Fowler, J. H., 2011. The evolution of overconfidence. Nature 477 (7364), 317.

Körding, K. P., Wolpert, D. M., 2004. Bayesian integration in sensorimotor learning. Nature 427 (6971), 244–247.

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

Likert, R., 1932. A technique for the measurement of attitudes. Archives of psychology.

Maslow, A. H., 1950. Self-actualizing people: a study of psychological health. Personality.

Mobius, M. M., Niederle, M., Niehaus, P., Rosenblat, T. S., 2014. Managing self-confidence. Working paper.

Moore, D. A., Healy, P. J., 2008. The trouble with overconfidence. Psychological Review 115 (2), 502.

Murphy, S. C., Barlow, F. K., von Hippel, W., 2017. A longitudinal test of three theories of overconfidence. Social Psychological and Personality Science, 1948–5506.

Murphy, S. C., von Hippel, W., Dubbs, S. L., Angilletta, M. J. J., Wilson, R. S., Trivers, R., Barlow, F. K., 2015. The role of overconfidence in romantic desirability and competition. Personality and Social Psychology Bulletin 41, 1036–1052.

Paulhus, D. L., Harms, P. D., Bruce, M. N., Lysy, D. C., 2003. The over-claiming technique: measuring self-enhancement independent of ability. Journal of personality and social psychology 84 (4), 890.

Puri, M., Robinson, D., 2007. Optimism and economic choice. Journal of Financial Economics 86 (1), 71–99.

Schwardman, P., van der Weele, J., 2017. Deception and self-deception. CRC TRR 190 Discussion Paper 25.

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

Taylor, S. E., Brown, J. D., 1988. Illusion and well-being: a social psychological perspective on mental health. Psychological bulletin 103 (2), 193.

Trivers, R., 1976. Foreword in: The Selfish Gene, R. Dawkins. Oxford University Press.

Vialle, I., Santos-Pinto, L. P., Rulliere, J.-L., 2011. Self-Confidence and Teamwork: An Experimental Test. SSRN Electronic Journal.

von Hippel, W., Trivers, R., 2011. Reflections on self-deception. Behavioral and Brain Sciences 34 (1), 41–56.

Von Neumann, J., Morgenstern, O., 1953. Theory of games and economic behavior. Princeton University Press Princeton, NJ.

Weinstein, N. D., 1980. Unrealistic optimism about future life events. Journal of personality and social psychology 39 (5), 806.

Appendix A Summary Statistics

Table A.1 presents the mean values (with standard errors in parentheses) of all the main variables of interest across treatments and conditions, as well as the difference between *Persuasion-first* and *Accuracy-first* within an information condition. Table A.2 displays the mean values for the same variables for Study 2, as well as for the pooled data for *SCI* condition from both studies.

Table A.1: A summary of results in Study 1.

	NI			*GI*			SCI			Pooled (*GI*+*NI*+*SCI*)		
	Acc.1st	*Per.1st*	Δ	*Acc.1st*	*Per.1st*	Δ	*Acc.1st*	*Per.1st*	Δ	*Acc.1st*	*Per.1st*	Δ
Performance	24.28	23.33	**−0.95**	23.73	24.05	**0.32**	23.78	23.92	**0.14**	23.93	23.77	**-0.16**
	(0.423)	(0.511)	(0.662)	(0.492)	(0.468)	(0.679)	(0.411)	(0.425)	(0.591)	(0.255)	(0.270)	(0.372)
Belief about perf.	23.26	22.91	**−0.35**	21.92	22.91	**0.99***	22.35	23.67	**1.32***	22.51	23.16	**0.65****
	(0.556)	(0.602)	(0.819)	(0.639)	(0.736)	(0.975)	(0.557)	(0.485)	(0.741)	(0.338)	(0.355)	(0.490)
Overconfidence	-1.02	-0.41	**0.61**	-1.81	-1.15	**0.67**	-1.43	-0.25	**1.18*****	-1.42	-0.60	**0.82*****
	(0.403)	(0.537)	(0.669)	(0.435)	(0.503)	(0.665)	(0.373)	(0.358)	(0.517)	(0.233)	(0.272)	(0.358)
Belief about relative	69.12%	71.46%	**2.34**%	69.96%	73.44%	**3.48**%	70.31%	73.74%	**3.43**%	69.80%	72.88%	**3.08**%**
perf.	(1.927)	(1.917)	(2.719)	(2.005)	(2.004)	(2.835)	(1.841)	(1.686)	(2.500)	(1.107)	(1.079)	(1.546)
Overplacement	16.33%	23.85%	**7.52****%	19.64%	21.81%	**2.17**%	21.57%	24.37%	**2.80**%	19.19%	23.34%	**4.15**%**
	(2.444)	(2.611)	(3.574)	(2.450)	(2.481)	(3.487)	(2.367)	(2.406)	(3.375)	(1.398)	(1.440)	(2.006)
Feedback	–	–	–	77.92%	80%	**2.08**%	67.4%	79.69%	**12.29**%***	–	–	–
				(1.801)	(1.891)	(2.611)	(2.106)	(2.075)	(2.958)			
Feedback bias	–	–	–	-1.18%	-0.17%	**1.01**%	-11.87%	-0.03%	**11.84**%***	–	–	–
				(0.921)	(1.011)	(1.367)	(1.885)	(1.923)	(2.692)			
Av. guessed score	21.81	21.94	**0.13**	21.85	22.38	**0.53**	21.79	22.77	**0.98***	21.82	22.36	**0.55****
	(0.459)	(0.504)	(0.681)	(0.446)	(0.475)	(0.652)	(0.426)	(0.408)	(0.590)	(0.255)	(0.268)	(0.370)
Reviewers' bias	-2.47	-1.39	**1.07**	-1.88	-1.67	**0.21**	-1.99	-1.14	**0.85**	-2.11	-1.40	**0.71**
	(0.549)	(0.579)	(0.798)	(0.543)	(0.510)	(0.745)	(0.490)	(0.533)	(0.723)	(0.304)	(0.312)	(0.435)
Av. convincingness	3.09	2.93	**−0.16**	3.21	3.26	**0.05**	3.10	3.29	**0.19***	3.13	3.16	**0.03**
	(0.081)	(0.103)	(0.131)	(0.073)	(0.077)	(0.106)	(0.081)	(0.074)	(0.110)	(0.045)	(0.050)	(0.068)
Obs.	98	97	195	96	96	192	100	97	197	294	290	583

Note: Table A.1 reports mean values of all measures (with standard errors in parentheses), and two-sided Mann-Whitney tests between *Accuracy-first* and *Persuasion-first* treatments within each information condition, and pooled across all information conditions (*NI*, *GI*, *SCI*). *** p <0.01, ** p <0.05, * p <0.10.

Table A.2: A summary of results in Study 2 with pooled data on *SCI* condition.

	SCI (lab)			*SCI* (MTurk + lab)		
	Acc.1st	Per.1st	Δ	Acc.1st	Per.1st	Δ
Performance	20.02	18.92	**−1.1**	22.53	22.22	**−0.31**
	(0.682)	(0.585)	(0.899)	(0.383)	(0.395)	(0.550)
Belief about perf.	18.46	20.18	**1.72**	21.05	22.48	**1.43****
	(0.833)	(0.748)	(1.119)	(0.486)	(0.430)	(0.650)
Overconfidence	-1.56	1.26	**2.82*****	-1.47	0.27	**1.74*****
	(0.491)	(0.519)	(0.714)	(0.297)	(0.300)	(0.422)
Belief about relative perf.	59.22%	65.98%	**6.76**%	66.61%	71.10%	**4.49**%*
	(2.937)	(2.341)	(3.756)	(1.622)	(1.397)	(2.144)
Overplacement	4.22%	10.98%	**6.76**%	15.79%	19.81%	**4.03**%
	(4.889)	(4.686)	(6.772)	(2.356)	(2.301)	(3.294)
Feedback	62.2%	75%	**12.8**%**	65.67%	78.10%	**12.43**%***
	(3.783)	(2.931)	(4.786)	(1.891)	(1.698)	(2.544)
Feedback bias	-4.53%	11.93%	**16.47**%***	-9.42%	4.04%	**13.46**%***
	(3.649)	(3.386)	(4.978)	(1.764)	(1.770)	(2.499)
Av. guessed score	22.02	23.07	**1.05**	21.87	22.87	**1.00****
	(0.578)	(0.635)	(0.858)	(0.342)	(0.344)	(0.485)
Reviewers' bias	2.00	4.15	**2.15**	-0.661	0.656	**1.32****
	(0.634)	(0.770)	(0.997)	(0.417)	(0.484)	(0.638)
Av. convincingness	3.62	3.43	**−0.18**	3.27	3.34	**0.07**
	(0.088)	(0.123)	(0.151)	(0.064)	(0.064)	(0.091)
N	50	50	100	150	147	297

Note: Table A.2 reports mean values of all measures (with standard errors in parentheses), and two-sided Mann-Whitney tests between *Accuracy-first* and *Persuasion-first* treatments for all measures related to performance, beliefs, and feedback. We also report the estimated average guessed score, reviewers' bias in their guess, and their ratings of convincingness using OLS regressions with robust standard errors clustered by session. (For all measures from the reviewers', MW tests do not apply anymore since in Study 2 reviewers rated multiple essays and one essay was reviewed by 5 different reviewers. Hence, we do not have fully independent observations.) *** p <0.01, ** p <0.05, * p <0.10.

Appendix B Additional analyses

B.1 Individual characteristics

Table A.3: Individual characteristics, by treatment

	Number of individuals	Mean score	Females (Percentage)	Mean age	Mean OCQ
	(1)	(2)	(3)	(4)	(5)
NI x Acc. 1st	98	24.28	0.480	35.92	4.12
NI x Pers. 1st	97	23.33	0.406	34.05	5.79
GI x Acc. 1st	96	23.73	0.479	36.29	4.54
GI x Pers. 1st	96	24.05	0.442	35.41	4.05
SCI x Acc. 1st	100	23.78	0.745	35.96	4.30
SCI x Pers. 1st	97	23.92	0.515	35.86	4.20
SCI lab x Acc. 1st	50	20.02	0.495	22.7	6.12
SCI lab x Pers. 1st	50	18.92	0.495	21.2	6.36

Note: A one-way between-subject ANOVA shows that there are no significant difference in the distribution of participants in terms of performance between treatments in Study 1 ($F(5, 578) = 0.50$; $p = 0.779$). We also find no significant difference in participants' performance between treatments in Study 2 (two tailed t-test: $p = 0.224$; Kolmogorov-Smirnov test: $p = 0.711$). A Chi2 test showed that there is no significant difference in the proportion of male and female between treatments in Study 1 ($Chi2(5) = 2.360$; $p = 0.797$). We also find no significant difference in the distribution of gender Study 2 (two-sample test of proportion: $p = 0.317$). Another one-way between-subject ANOVA shows that there are no significant difference in the distribution of participants in terms of age between treatments in Study 1 ($F(5, 578) = 0.55$; $p = 0.740$) and between treatments in Study 2 (two tailed t-test: $p = 0.100$; Kolmogorov-Smirnov test: $p = 0.544$). Finally, a one-way between-subject ANOVA shows that there are no significant difference in the distribution of participants in terms of dispositional overconfidence measured by the OCQ between treatments in Study 1 ($F(5, 578) = 1.59$; $p = 0.161$) and between treatments in Study 2 (two tailed t-test: $p = 0.816$; Kolmogorov-Smirnov test: $p = 0.393$).

The OCQ scale measures the dispositional overconfidence. Participants are asked to rate how familiar they are with each item on a scale from 0 to 4 (0 being not familiar at all and 4 being very familiar), 8 of which are non-existent (see appendix D for the list of items). The sum of ratings of nonexistent items constitutes the over-claiming index that we use as a measure of dispositional overconfidence. The Over-Claiming Questionnaire originally proposed by Paulhus et al. (2003) is composed of 150 items classified in 10 categories. In each categories, 3 out of the 15 items are non-existent. In this version, participants are asked to indicate how familiar they are with each item of the series on a 6-point Likert scale (Likert, 1932). The OCQ is placed before the general knowledge test in Study 1, whereas other control measures are collected at the end of the experiment.

B.2 Effect of anticipation of strategic interactions on beliefs

Table A.4 presents the determinants of participants beliefs about their performance. Columns (1), (3), (6), (9) and (12) report the OLS regression of participants' beliefs about their performance on the treatment dummy "Persuasion", controlling for performance. Columns (4), (7), (10) and (13) report the OLS regression of participants' beliefs about their performance

on the treatment dummy and the proportion of correct answers contained in the feedback. In Columns (2), (5), (8), (11) and (14), we further control for participants individual characteristics (sex, age and OCQ). Table A.5 presents the same models with participants' beliefs about their relative performance as the dependent variable.

Columns (1) and (2) present the results for the *NI* condition. We do not observe any significant effect of being in the *Persuasion-first* treatment in this condition. Higher performance always leads to higher beliefs about their absolute performance. Model (2) also shows that women tend to be less confident than men about their performance in the absence of feedback. Columns (3) to (5) from table A.4 present the results of the estimations for the *GI* condition. These models show that being in the *Persuasion-first* treatment affects positively participants' beliefs about their score but the effect is not significant, after controlling for performance. These models also show that the feedback has a significant effect on beliefs at the 1% level. Finally, columns (6) to (14) present the results for the *SCI* condition across two studies. Controlling for performance in the baseline model, we observe that being in the *Persuasion-first* treatment leads to higher beliefs about their absolute performance and the effect is significant at the 1% level. However, when we add the Feedback as an explanatory variable, the treatment effect disappears and the feedback has a positive effect on beliefs and the effect is significant at the 1% level in Study 1 and when pooling Study 1 and 2, but is only significant at the 10% level in Study 2. The difference in the level of significance may be driven by very different number of observations we have in each study. Table A.5 displays similar (stronger) results when we use participants' beliefs about their rank as the dependent variable.

Table A.4: Determinants of participants' beliefs about performance.

	Study 1								Study 2			Pooled		
Dep. Var: Beliefs	*NI*		*GI*			*SCI*			*SCI* (lab)			*SCI* (MTurk +lab)		
about perf.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Feedback	–	–	–	0.285***	0.282***	–	0.094***	0.087***	–	0.041*	0.045*	–	0.080***	0.071***
				(0.018)	(0.018)	(0.017)	(0.016)		(0.023)	(0.024)		(0.014)	(0.014)	
Persuasion	0.376	0.201	0.650	0.395	0.401	1.196**	0.167	0.185	2.776***	1.192	0.821	1.709***	0.429	0.551
	(0.654)	(0.652)	(0.666)	(0.631)	(0.638)	(0.516)	(0.719)	(0.692)	(0.722)	(1.147)	(1.158)	(0.419)	(0.642)	(0.619)
Performance	0.752***	0.696***	1.052***	–	–	0.903***	–	–	0.960***	–	–	0.904***		
	(0.071)	(0.073)	(0.071)			(0.062)			(0.081)			(0.044)		
Female	–	-2.304***	–	–	-0.430	–	–	-1.639**	–	–	0.278	–	–	-1.089*
		(0.651)			(0.673)			(0.673)			(1.118)			(0.603)
Age	–	0.029	–	–	0.038	–	–	0.042	–	–	-0.254**	–	–	0.087***
		(0.030)			(0.030)			(0.033)			(0.125)			(0.028)
OCQ	–	-0.015	–	–	-0.002 –	–		-0.219***	–	–	-0.101	–	–	-0.203***
		(0.064)			(0.069)			(0.063)			(0.108)			(0.057)
Constant	4.988***	6.487***	-3.041*	-0.329	-1.228	0.886	16.027***	16.747***	-0.757	15.894***	21.915***	0.696	15.768***	15.270***
	(1.773)	(2.209)	(1.753)	(1.435)	(1.840)	(1.529)	(1.223)	(1.699)	(1.691)	(1.652)	(3.413)	(1.042	(1.025)	(1.377)
R-squared	0.409	0.526	0.539	0.587	0.591	0.526	0.154	0.238	0.603	0.054	0.100	0.592	0.114	0.193
Observations	194	193	192	192	191	197	197	196	100	100	100	297	297	296

Notes: Table A.4 reports OLS regressions with standard errors in parentheses. Columns (1) to (8) shows the results for the observations of participants in Study 1. Columns (9) and (11) shows the results for the observations of participants in Study 2. Columns (12) and (14) shows the results for the observations of participants in studies 1 and 2 pooled together (*SCI* treatment only). *** p <0.01, ** p <0.05, * p <0.10.

Table A.5: Determinants of participants' beliefs about relative performance.

	Study 1								Study 2			Pooled		
Dep. Var: Beliefs	*NI*			*GI*			*SCI*		*SCI* (lab)			*SCI* (MTurk +lab)		
about relative perf.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Feedback	–	–	–	0.665***	0.631***	–	0.233***	0.204***	–	0.152*	0.162**	–	0.221***	0.190***
				(0.062)	(0.065)		(0.058)	(0.058)		(0.078)	(0.081)		(0.047)	(0.048)
Persuasion	4.593**	3.779*	2.569	2.0.94	2.271	3.099	0.563	0.962	9.175**	4.814	4.474	5.194***	1.747	2.120
	(2.245)	(2.247)	(2.149)	(2.250)	(2.230)	(2.055)	(2.513)	(2.448)	(3.237)	(3.837)	(3.959)	(1.741)	(2.155)	(2.132)
Performance	2.344***	2.185***	2.735***	–	–	2.423***	–	–	2.195***	–	–	2.284***		
	(0.242)	(0.253)	(0.230)			(0.249)			(0.361)			(0.184)		
Female	–	-7.449***	–	–	-4.556**	–	–	-7.470***	–	–	1.541	–	–	-4.871**
		(2.289)			(2.277)			(2.383)			(3.822)			(2.076)
Age	–	-0.012	–	–	0.232**	–	–	-0.039	–	–	-0.237	–	–	0.155
		(0.103)			(0.104)			(0.118)			(0.427)			(0.096)
OCQ	–	-0.009	–	–	-0.180	–	–	-0.580**	–	–	0.070	–	–	-0.454**
		(0.219)			(0.241)			(0.223)			(0.371)			(0.195)
Constant	12.223**	20.109***	5.061	18.144***	15.393**	12.701**	54.576***	64.125**	15.271**	49.763***	53.276***	15.161***	52.129***	53.954***
	(6.086)	(7.609)	(5.654)	(5.117)	(6.433)	(6.095)	(4.279)	(6.014)	(7.579)	(5.524)	(11.670)	(4.326)	(3.439)	(4.738)
R-squared	0.332	0.365	0.433	0.380	0.409	0.334	0.085	0.163	0.299	0.068	0.073	0.353	0.082	0.124
Observations	194	193	192	192	191	197	197	196	100	100	100	297	297	296

Notes: Table A.5 reports OLS regressions with standard errors in parentheses. Columns (1) to (8) shows the results for the observations of participants in Study 1. Columns (9) and (11) shows the results for the observations of participants in Study 2. Columns (12) and (14) shows the results for the observations of participants in studies 1 and 2 pooled together (*SCI* treatment only). *** p <0.01, ** p <0.05, * p <0.10.

B.3 Determinants of information sampling and its impact on beliefs about performance and relative performance

Table A.6 presents the determinants of the proportion of correct answers revealed through the feedback in *GI* and *SCI* conditions. Columns (1), (3), (5), and (7) report the OLS regressions of the proportion of correct answers revealed through the feedback on treatment dummy and performance. We control for participants' characteristics (sex, age and OCQ) in columns (2), (4), (6) and (8).

As expected, models (1) and (2) shows no treatment effect on the feedback content since the feedback is exogenous in the *GI* condition. On the other hand, models (3) to (8) show a positive and strongly significant (1%) treatment effect on the proportion of correct answers revealed through the feedback.

Table A.6: Determinants of Information sampling.

	Study 1				Study 2		pooled	
Dep. Var:	*GI*		*SCI*		*SCI* (lab)		*SCI* (MTurk + lab)	
Feedback	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Persuasion	1.025	0.860	11.973***	12.319***	14.250***	14.499***	12.978***	13.073***
	(1.371)	(1.381)	(2.631)	(2.631)	(4.696)	(4.791)	(2.354)	(2.350)
Performance	3.277***	3.231***	2.312***	2.427***	1.318**	1.333**	1.778***	1.782***
	(0.146)	(0.154)	(0.319)	(0.350)	(0.524)	(0.530)	(0.249)	(0.278)
Female	–	0.226	–	-2.105	–	-6.860	–	-4.344*
		(1.412)		(2.719)		(4.682)		(2.389)
Age	–	-0.045	–	-0.086	–	0.587	–	-0.050
		(0.065)		(0.136)		(0.533)		(0.117)
OCQ	–	-0.252*	–	0.355	–	-0.160	–	0.109
		(0.148)		(0.266)		(0.460)		(0.232)
Constant	0.154	3.903	12.425	11.995	35.814***	26.880	25.609***	28.525***
	(3.608)	(4.294)	(7.802)	(9.441)	(10.993)	(18.342)	(5.848)	(6.883)
R-squared	0.727	0.732	0.277	0.291	0.125	0.157	0.212	0.224
Observations	192	191	197	196	100	100	297	296

Notes: Table A.6 reports OLS regressions with standard errors in parentheses. We consider the observations of participants in the *GI* and *SCI* treatment only. Columns (1) and (2) shows the results for the observations of participants in *GI*. Columns (3) and (4) shows the results for the observations of participants in *SCI* from Study 1. Columns (5) and (6) shows the results for the observations of participants in *SCI* from Study 2. Columns (7) and (8) shows the results for the observations of participants in *SCI* from both Study 1 and 2 pooled together. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A.7 reports the corresponding regressions reported in Table 2 with all the control variabled included. Results stay robust.

Table A.7: Effect of information sampling on beliefs about performance and relative performance.

Dep. Var:	*SCI* (MTurk)		*SCI* (lab)		*SCI* (MTurk +lab)	
Beliefs about	perf.	relative perf.	perf.	relative perf.	perf.	relative perf.
	(1)	(2)	(3)	(4)	(5)	(6)
Feedback	0.091**	0.251	0.184**	0.660**	0.125	0.384***
	(0.044)	(0.170)	(0.077)	(0.295)	(0.038)	(0.146)
Performance	0.677***	1.807***	0.701***	1.363**	0.702***	1.694***
	(0.130)	(0.503)	(0.570)	(0.149)	(0.088)	(0.340)
Female	-0.475	-4.134*	1.770	6.181	0.259	-1.189
	(0.561)	(2.175)	(1.265)	(4.833)	(0.532)	(2.047)
Age	-0.018	-0.203*	-0.209*	-0.239	-0.027	-0.133
	(0.028)	(0.109)	(0.124)	(0.475)	(0.025)	(0.095)
OCQ	-0.048	-0.116	-0.016	0.293	-0.035	-0.029
	(0.057)	(0.220)	(0.108)	(0.413)	(0.049)	(0.189)
Constant	1.271	20.369**	-3.152	-8.817	-1.985	8.328
	(2.061)	(7.997)	(5.366)	(20.504)	(1.978)	(7.615)
First-stage F-stat	15.61	6.56	3.51	5.86	16.70	14.43
Observations	196	196	100	100	296	296

Notes: Table A.7 reports 2SLS regressions with standard errors in parentheses. Feedback is instrumented by the treatment dummy. Columns (1) to (2) shows the results from Study 1. Columns (3) and (4) shows the results from Study 2. Columns (5) and (6) shows the results for both studies 1 and 2 pooled together. *** p <0.01, ** p <0.05, * p <0.10.

B.4 Causal identification: the effect of confidence on persuasion

Table A.8: Causal identification of the effect of participants' beliefs about performance and relative performance on persuasiveness.

Dep. Var:	*SCI* (MTurk)		*SCI* (lab)		*SCI* (MTurk + lab)	
Persuasiveness	(1)	(2)	(3)	(4)	(5)	(6)
Beliefs about perf.	0.871	–	0.429*	–	0.656***	–
	(0.577)		(0.257)		(0.206)	
Beliefs about	–	0.317	–	0.120**	–	0.214***
relative perf.		(0.253)		(0.060)		(0.073)
Performance	-0.558	-0.542	-0.089	0.049	-0.362	-0.264
	(0.528)	(0.623)	(0.234)	(0.145)	(0.211)	(0.220)
Female	0.611	1.510	-0.754	-0.735	0.100	0.524
	(0.762)	(1.406)	(0.829)	(0.819)	(0.463)	(0.616)
Age	0.016	0.065	-0.092	-0.153	-0.029	-0.018
	(0.037)	(0.070)	(0.119)	(0.137)	(0.026)	(0.041)
OCQ	-0.052	-0.058	0.013	-0.029	-0.016	-0.033
	(0.066)	(0.079)	(0.046)	(0.060)	(0.029)	(0.033)
Constant	14.839***	9.480	18.278***	17.982***	17.087***	14.006***
	(2.437)	(6.848)	(2.305)	(3.065)	(1.158)	(1.897)
First-stage F-stat	45.05	22.72	29.76	8.22	87.96	33.81
Observations	196	196	100	100	296	296

Notes: Table A.8 reports 2SLS regressions with standard errors in parentheses. Participants' beliefs are instrumented by the treatment dummy. Columns (1) and (2) shows the results from Study 1. Columns (3) and (4) shows the results from Study 2. Columns (5) to (6) shows the results from studies 1 and 2 pooled together. *** p <0.01, ** p <0.05, * p <0.10.

B.5 Causal identification: the effect of confidence on persuasion (*Given Information* condition)

In the *GI* condition, the feedback given to the participants has a random component. The questions selected, for all, to be used as feedback, may have been easier or harder for different participants, relative to the other questions they faced. If the question chosen for feedback happen to have been relatively hard in comparison to the other question he/she faced, this contestant's feedback may look more negative than his/her result over the whole test. We use this exogenous variation in positive *vs*. negative feedback bias to study the effect of the induced variations in confidence on persuasiveness.

Table A.9: Effect of participants' beliefs about performance and relative performance on persuasiveness.

	Given Information (*GI*)			
Dep Var:	Beliefs about performance		Beliefs about relative performance	
Persuasiveness	(1)	(2)	(3)	(4)
Beliefs	0.306*	0.306*	0.265	0.270
	(0.169)	(0.164)	(0.181)	(0.179)
Performance	0.038	0.068	-0.367	-0.321
	(0.190)	(0.182)	(0.504)	(0.476)
Female	–	1.117*	–	2.185**
		(0.614)		(1.089)
Age	–	-0.048*	–	-0.086
		(0.028)		(0.045)
OCQ	–	0.025	–	0.070
		(0.064)		(0.089)
Constant	14.350***	14.808**	11.875***	12.24***
	(1.670)	(5.217)	(2.268)	(2.571)
First-stage F-statistic	147.41	58.27	74.43	31.67
Observations	192	191	192	191

Notes: Table A.9 reports 2SLS regressions for *GI* condition only with standard errors in parentheses. Participants' beliefs are instrumented by the Feedback variable. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A.9 reports the corresponding 2SLS estimations of the effect of confidence on participants' persuasiveness, using the exogenous variation in feedback as an instrument. The dependent variable is participants' persuasiveness measured as the average reviewers' guess of the participants' score in the test. The independent variables includes participants' beliefs about their performance instrumented by the exogenous variation in feedback in models (1) and (2). We control for performance in model (1) and as well for participants' individual characteristics (sex, age and OCQ) in model (2). We use participants' beliefs about their relative performance (i.e., percentage of participants they have outperformed) as the dependent variable also instrumented by the exogenous variation in feedback in model (3) and (4).

Model (1) to (4) in Table A.9 shows that an increase in beliefs has a positive effect on participants' level of persuasiveness. However, the effect is only (marginally) significant in models (1) and (2). Going back to Hypothesis 1, it is likely that the lack of effect is due to the fact that participants do not have much room to inflate their beliefs in the *GI* condition. Indeed, one can think that it is harder to ignore an incorrect answer when it is displayed on the screen. If this Hypothesis is true, we should find a stronger effect of beliefs on persuasion in the *SCI* condition. We have shown in Section 3.4 in the main text that this is indeed the case.

B.6 Summary Statistics on the essays content

In this section, we investigate whether participants' essay content differs between information condition. Indeed, since some participants were given extra information about their performance compared to others it is possible that participants in the *No Information* (*NI*) condition compensated their lack of facts about their performance by providing more information about themselves. Here is an example of an essay from a participant who was assigned to the *Self-Chosen Information* x *Persuasion-first* condition: *That test was a snap! I had the answers just rolling out of my head on to the answer list, usually before I even read the answer choices. I did the check to see if I got them right, I even tested most of the ones I was less sure about and got 9 of the 10 right! The other was a silly misclick, but I definetly knew the rest! I Even realized i hit the wrong thing as I hit next. I doubt many others did half as well as I did.* (actual score: 21/31). In contrast, here is an example of an essay from a participant who was assigned to the *No Information* x *Persuasion-first* condition: *My friends have always joked with me that I would be enormously successful if I could find a way to profit from all the random information and bits of trivia that I know. I tend to have an excellent ability to remember seemingly insignificant details and almost never pass up an opportunity to learn something. Regardless of whether I'm reading a book, listening to a presentation, reading wikipedia or watching television, I'm always absorbing information. Perhaps this is my opportunity to finally earn some sort of return on my investment, I hope that you'll trust me when I say that I knew every question in the quiz that I just took.* (actual score: 28/31)

We asked MTurk workers that did not participate in Study 1 (either as main participants or reviewers) to code the content of each essay. Each essay was coded by 5 MTurk workers and each MTurk worker coded 18 essays. For each essay, the MTurk workers were asked to answer the following yes/no questions: (i) The participant mentions the number (or percentage) of questions he or she thinks he or she correctly answered during the test; (ii) The participant mentions his or her rank compared to the other MTurkers (ex: I think I did better than 80% of the other participants); (iii) The participant mentions his or her own qualities or characteristics (ex: I am good at history) and (iv) The participant mentions the feedback he or she received about his or her performance. MTurk workers were also told that the authors of those essays undertook a knowledge test and that some of them received a feedback about 10 questions of the test and were told how many questions out of these 10 questions they answered correctly. Table reports the average proportion of participants that mention the features summarised above in their essay.

Table A.10: Essays content

% participants mentioning:	number/percentage of correct answers	relative performance in the sample	own characteristics	feedback
NI x Acc. 1st	51.02%	17.35%	77.55%	–
NI x Per. 1st	50.52%	18.56%	77.32%	–
GI x Acc. 1st	65.63%	25.00%	64.58%	25.00%
GI x Per. 1st	65.00%	20.83%	80.21%**	46.88%**
SCI x Acc. 1st (MTurk)	56.00%	16.00%	77.00%	26.00%
SCI x Per. 1st (MTurk)	68.04%*	15.46%	70.10%	46.39%***
SCI x Acc. 1st (lab)	58.00%	18.00%	74.00%	42.00%
SCI x Per. 1st (lab)	68.00%	12.00%	78.00%	36.00%

Note: Table A.10 indicates the proportion of participants who mentioned a particular feature in their essay and tests of proportions between *Accuracy-first* and *Persuasion-first* treatments within each information condition. *** p <0.01, ** p <0.05, * p <0.10.

Table A.10 shows that most participants write about their own characteristics (such as being good at Trivia or what their major is) and their score at the test. Only about 20% mention their beliefs about their relative position in the distribution. Overall, around 35% of the participants talk about the feedback they received in the *Given Information* (*GI*) and *Self-Chosen Information* (*SCI*).

The identification assumption we make in Section 3.4 is that the treatment affects persuasiveness only through beliefs. However, participants in the *Persuasion-first* treatment under *SCI* conditions could perhaps simply reveal more often the feedback they received on their answers (in particular given that they have more positive feedback than in the *Accuracy-first* treatment). Such a difference could potentially have an impact on persuasiveness independently of the confidence of the participant.

We observe that the content of the essays vary across treatments, which is to be expected if confidence leads participants to write more persuasive essays. We observe that both in the *GI* condition and the *SCI* condition of Study 1, participants are more likely to mention their feedback in the *Persuasion-first* treatment than in the *Accuracy-first* treatment. However, these differences are not associated with a very significant effect of our 2SLS estimation (see the first two columns in Table 3 and Table A.9). On the contrary, we do not observe that participants in the *Persuasion-first* treatment mention their feedback more in Study 2, but we observe that our 2SLS estimation in Table 3 is positive and significant for Study 2.

Therefore, it does not look like the estimates of our 2SLS estimation are driven by systematic tendencies of participants to simply reveal their feedback more in the *Persuasion-first* treatment.

Appendix C Instructions

C.1 Instructions (MTurk) - Accuracy-first x Given Information

— Main participants —

PART 1

Please rate your familiarity with each item by selecting the appropriate number from 0 to 4. If you are very familiar with an item, choose the number 4. If you have never heard of an item, choose the number 0.

PART 2

The second part of this study involves a general knowledge test in the form of a Multiple Choice Questionnaire.

The test is composed of 31 questions. You will have 15 seconds per questions to make your decision. Once you have made your decision, press the '>>' button to start the next question. If you fail to select an answer before the end of the 15 seconds, the next question will start automatically.

WARNING: You have to press the '>>' button to enter your answer. if you don't click on the '>>' button, the computer will score it as if you did not answer this question.

Click the '>>' button to start the practice trial before the real test.

PART 3

Now that you completed the knowledge test, you will have the chance to earn up to 2 extra dollars.

For this next task we would now like you to estimate how well you did on the general knowledge test. The closer your estimate is to your true performance, the more money you will earn.

In order to help you make this judgement, we will show you 10 of the questions and indicate whether you answered them correctly or not.

Press the '>>' button to see the answers.

Elicitation 1
You now have the opportunity to earn additional money by estimating your performance on the test. You will be paid as a function of the accuracy of your estimate.

You will earn $0.50 on this task if your estimate is in the 50% of most accurate estimates. You will earn another $0.50 if your estimate is in the 10% of most accurate estimates.

Elicitation 2
Thinking about your performance again, how do you think you compare to the average MTurk worker on this task, in terms of number of items correct?

On the slider below, place yourself between 0% and 100%.

For example, if you think you did better than 3/4 of your fellow MTurk users but worse than 1/4 of them, you will choose 75%. Conversely, if you think that you did better than 1/4 of your fellow MTurk users but worse than 1/3, you will choose 25%.

You will earn $0.50 on this task if the estimate of your rank compared to your fellow MTurk users is in the 50% of most accurate estimates. You will earn an additional $0.50 if the estimate of your rank is in the 10% of most accurate estimates.

PART 4

You now have the opportunity to earn up to an additional $2 by writing a short essay to convince some of your fellow MTurk workers about how well you performed on the test.

Your payment for this task will depend both on (A) how well the reviewers think you performed on the test and (B) on how convincing your essay is.

(A) If your score is rated in the 50% best scores by the reviewers, you will be paid $0.50 and an additional $0.50 if your score is rated in the 10% best scores.

(B) If your essay is rated in the 50% of the most convincing essays by the reviewers, you will get $0.50 and if your essay is rated in the 10% of most convincing essays, you will earn another $0.50, independent of how high they think you scored.

We will not be providing your fellow MTurk workers with any information about how you actually did, so you need to put your best foot forward in an effort to be as convincing as

possible about your performance.

When your essay is done, please hit the '>>' button.

— Reviewers —

In a previous study, your fellow MTurk workers took a 31 questions knowledge test. At the end of the test, we ask them to write an essay about their performance.

Your task during this experiment will be to rate one of these short essays both on (A) how convincing you think the essay is and (B) how many questions you think the author answered correctly in the knowledge test.

When you are ready to start, please hit the '»' button.

C.2 Instructions (lab) - Persuasion-first x *Self-Chosen Information*

Note: All the instructions are displayed on the participants' screens.

— Main participants —

PART 1

The first part of this study involves a general knowledge test in the form of a Multiple Choice Questionnaire.

The test is composed of 31 questions. You will have 15 seconds per questions to make your decision. Once you have made your decision, press the 'next' button to start the next question. If you fail to select an answer before the end of the 15 seconds, the next question will start automatically.

Click the 'next' button to start the test.

PART 2

For this next task we would now like you to write a short essay to convince a group of reviewers about how well you performed on the test. The reviewer will be rating your essay for how convincing you are and for how many items they think you got right. The more convincing you are and the more they think you answered correctly, the more likely you will

be to earn additional money.

In order to help you write this essay, you will be allowed to choose 10 of the questions and we will indicate whether you answered them correctly or not.

Here are the first 30 questions you just answered. Please choose 10 items for which you'd like to see if you answered correctly. Click on the individual questions to select them. When you have selected 10 questions, press the 'next' button to see the answers.

Select the items you want to check:

- ☐ What bird has the widest wingspan?
- ☐ The unit of electrical resistance was named after whom?
- ☐ Titan is a moon of which planet?
- ☐ How many pieces are on a chessboard at the start of a game?
- ☐ Which of these is the largest in area?
- ☐ In Greek mythology, who was the multi-headed dog, encircled by a serpent, that guarded the portal to the underworld?
- ☐ At the opening ceremony of every Olympic Games when the athletes parade into the stadium, what is traditionally the first nation to enter?
- ☐ Which of these types of music did not originate in the Caribbean?
- ☐ What is the name of the engraved stone, discovered in 1799, that provided a key to deciphering the languages of ancient Egypt?
- ☐ The Dalai Lama is a high lama in which religion?
- ☐ Which Scotsman took out a patent in 1876 that was the nucleus of the telephone?
- ☐ What is another name for a blood clot?
- ☐ Which of these countries is not landlocked?
- ☐ The tibia and fibula are found where in the human body?
- ☐ "Facebook" was launched in what year?
- ☐ Where is it believed that fireworks were invented?
- ☐ Which of these is found in the brain?
- ☐ Which of these is in North America?
- ☐ Which of Galileo's achievements brought him into conflict with the church, resulting in his being confined to his house for the last years of his life?
- ☐ What is the closest planet to the sun?
- ☐ On which continent are the native fauna called ostrich, lion, giraffe and okapi?
- ☐ What do anthropologists study?
- ☐ In the Alfred Hitchcock film "Psycho", where did the murder take place?
- ☐ Where would one find a hypotenuse?
- ☐ Who has won the most Olympic Gold medals?

☐ What does the chemical symbol Fe stand for?

☐ Michael J Fox played which character in the "Back to the Future" trilogy (1985-1990)?

☐ In medicine, what do the s BMI stand for?

☐ A single flame gas burner frequently used in student science laboratories is named after whom?

☐ A sabre is what type of weapon?

PART 3

You now have the opportunity to earn up to $8 by writing a short essay to convince a group of reviewers of how well you performed on the test.

The reviewers are another group of QUT students participating in this experiment, in another room, as we speak. However, the reviewers have no experience in the test you just completed and we will not be providing the reviewers with any information about how you actually did in that test.

Your payment for this task will depend both on (A) how well the reviewers think you performed on the test and (B) on how convincing your essay is.

(A) If your score is rated in the 50% best scores of today's session participants by the reviewers, you will be paid $2 and an additional $2 if your score is rated in the 10% best scores.

(B) If your essay is rated in the 50% of the most convincing essays by the reviewers, you will get $2 and if your essay is rated in the 10% of most convincing essays, you will earn another $2, independent of how high they think you scored.

When your essay is done, please hit the "next" button.

Essay rules:

The content of your essay is not restricted in any way.

You are only forbidden to make threats, to reveal your identity, seat number or anything that might uncover your anonymity.

If you violate these restrictions you will not receive any payment at the end of the experiment.

PART 4

You now have the opportunity to earn additional money by estimating how many questions out of the 31 on the test you correctly answered. You will be paid as a function of the accuracy of your estimate.

You will earn $2 on this task if your estimate is in the 50% of most accurate estimates. You will earn another $2 if your estimate is in the 10% of most accurate estimates.

PART 5

On the slider below, place yourself between 0% and 100% in terms of your performance relative to the other participants in this room.

For example, if you think you did better than 3/4 of your fellow participants but worse than 1/4 of them, you will choose 75%. Conversely, if you think that you did better than 1/4 of your fellow participants but worse than 3/4, you will choose 25%.

You will earn $2 on this task if the estimate of your rank compared to your fellow participants is in the 50% of most accurate estimates. You will earn an additional $2 if the estimate of your rank is in the 10% of most accurate estimates.

PART 6

Please rate your familiarity with each item by selecting the appropriate number from 0 to 4. If you are very familiar with an item, choose the number 4. If you have never heard of an item, choose the number 0.

— Reviewers —

Some QUT students are participating in an experiment in a different room, as we speak.

In this experiment, your fellow participants took a 31 question general knowledge test and we asked them to write an essay about their performance.

Before writing their essay, they were allowed to check 10 questions and were told whether they answered each of these questions correctly or not.

Your task during this experiment will be to rate some of these short essays both on (A) how convincing you think the essay is and (B) how many questions you think the author answered correctly in the knowledge test.

Payment:

Each guess of how many questions you think the author of the essay answered correctly in the knowledge test gives you an opportunity to earn extra money.

If your guess is exactly equal to the true performance of the essay writer or deviates from the true performance of the essay writer by 1 question (meaning that the true score if the essay writer was either 1 point above or below your guess), you will earn $10.

If your guess deviates from the true performance of the essay writer by 2 question (meaning that the true score if the essay writer was either 2 points above or below your guess), you will earn $8.

If your guess deviates from the true performance of the essay writer by 3 questions (meaning that the true score if the essay writer was either 3 points above or below your guess), you will earn $4.

If your guess deviates from the true performance of the essay writer by 4 or 5 questions (meaning that the true score if the essay writer was either 4 or 5 points above or below your guess), you will earn $2.

If your guess deviates from the true performance of the essay writer by more than 5 questions, you will not earn anything.

At the end of the experiment, the program will randomly select one guess and you will be paid according to that guess only. Therefore, you should treat every guess as if it is the guess that matters.

When you are ready to start, hit the 'next' button.

Appendix D 25-item version of the Over-Claiming Questionnaire

Table A.11: OCQ items

Item	Foil or Real
Houdini	Real
Charlotte Bronte	Real
meta-toxins	Foil
Antigone	Real
cholarine	Foil
alliteration	Real
Gail Brennan	Foil
myth	Real
Queen Shattuck	Foil
Lewis Carroll	Real
free will	Real
Dale Carnegie	Real
Murphy's Last Ride	Foil
sentence stigma	Foil
Bay of Pigs	Real
hyperbole	Real
The Aeneid	Real
euphemism	Real
double entendre	Real
consumer apparatus	Foil
blank verse	Real
shunt-word	Foil
art-deco	Real
Artemis	Real
a cappella	Real

Note: In our version of the OCQ, we added 'Australia' as an attention check.

Appendix E General Knowledge test items

The items that differ between the two studies are highlighted in bold. The correct answers are displayed in red.

Study 1

1. **Who was the first person to sign the American Declaration of Independence?**
   David Crockett / George Washington / John Hancock / Benjamin Franklin

2. The unit of electrical resistance was named after whom?
   Georg Simon Ohm / Benjamin Franklin / Guglielmo Marconi

3. Titan is a moon of which planet?
   Mars / Uranus / Saturn / Venus

4. How many pieces are on a chessboard at the start of a game?
   8 / 32 / 16 /64

5. Which of these is the largest in area?
   Spain / Texas / Tanzania / Afghanistan

6. **Who did George W. Bush beat for the US presidency in 2000?**
   Al Gore / John Kerry / John F. Kennedy / John McCain

7. At the opening ceremony of every Olympic Games when the athletes parade into the stadium, what is traditionally the first nation to enter?
   Australia / Simbabwe / Greece / Denmark

8. Which of these types of music did not originate in the Caribbean?
   Gregorian chant / Flamenco / Ska / Reggae

9. What is the name of the engraved stone, discovered in 1799, that provided a key to deciphering the languages of ancient Egypt?
   Babel Stone / Blarney Stone / Rosetta Stone / Talking Stone

10. **The portrait of which US statesman appears on the US $100 bill?**
    Abraham Lincoln / George Washington / Theodore Roosevelt / Benjamin Franklin

11. Which Scotsman took out a patent in 1876 that was the nucleus of the telephone?
    Alexander Fleming / Thomas Edison / George Stephenson / Alexander Bell

12. What is another name for a blood clot?
    Abrasion / Carcinoma / Bursitis / Thrombosis

13. Which of these countries is not landlocked? Paraguay / Bolivia / Andorra / Australia

14. The tibia and fibula are found where in the human body? Arm / Lower leg / Ribcage / Fingers

15. ”Facebook” was launched in what year?
1990 / 1994 / 2004 / 2009

16. Where is it believed that fireworks were invented?
China / Mexico / Egypt / Greece

17. Which of these is found in the brain?
Tibia / Thalamus / Vertebra / Humerus

18. Which of these is in North America?
The Orzaks / The Urals / The Himalayas / The Pyrenees

19. Which of Galileo’s achievements brought him into conflict with the church, resulting in his being confined to his house for the last years of his life?
He attempted to measure the speed of light / He invented the thermometer / He said that Copernican view of the universe was correct / He attempted to measure the weight of air

20. What is the closest planet to the sun?
Venus / Mercury / Saturn / Mars

21. On which continent are the native fauna called ostrich, lion, giraffe and okapi?
Africa / South America / Australia / Asia

22. What do anthropologists study?
Human Beings / Coal / Monkeys / Minerals

23. In the Alfred Hitchcock film ”Psycho”, where did the murder take place?
In the bedroom / In the kitchen / On the front porch / In the shower

24. Where would one find a hypotenuse?
Under the wing of a chicken / In the roof of a wooden building / As part of a right angled triangle / In a vehicle’s gearbox

25. Who has won the most Olympic Gold medals? Larrisa Latynina, URSS / Paavo Nurmi, Finland / Michael Phelps, USA / Mark Spitz, USA

26. What does the chemical symbol Fe stand for?
Iron / Gold / Silver / Cheese

27. Michael J Fox played which character in the ”Back to the Future” trilogy (1985-1990)?
Mickey McMouse / Marty McFly / Morris McAustin / Maurice McKee

28. In medicine, what do the s BMI stand for?
    Bionic Machine Implants / Biochemical Mortuary Investigators / British Medical Institute / Body Mass Index

29. A single flame gas burner frequently used in student science laboratories is named after whom?
    John Tilley / Michael Faraday / Sir Humphry Davy / Robert Bunsen

30. **A filibuster is typically found where?**
    A decision-making body / A hospital / A society party / A horse race

31. (*Attention Check*) What day comes after Tuesday?
    Monday / Wednesday / Thursday / Friday

Study 2

1. **What bird has the widest wingspan?**
   Albatros / Condor / Eagle / Vulture

2. The unit of electrical resistance was named after whom?
   Georg Simon Ohm / Benjamin Franklin / Guglielmo Marconi

3. Titan is a moon of which planet?
   Mars / Uranus / Saturn / Venus

4. How many pieces are on a chessboard at the start of a game?
   8 / 32 / 16 /64

5. Which of these is the largest in area?
   Spain / Texas / Tanzania / Afghanistan

6. **In Greek mythology, who was the multi-headed dog, encircled by a serpent, that guarded the portal to the underworld?**
   Minotaur / Rover / Cerebrus / Buccephalus

7. At the opening ceremony of every Olympic Games when the athletes parade into the stadium, what is traditionally the first nation to enter?
   Australia / Simbabwe / Greece / Denmark

8. Which of these types of music did not originate in the Caribbean?
   Gregorian chant / Flamenco / Ska / Reggae

9. What is the name of the engraved stone, discovered in 1799, that provided a key to deciphering the languages of ancient Egypt?
   Babel Stone / Blarney Stone / Rosetta Stone / Talking Stone

10. **The Dalai Lama is a high lama in which religion?**
    Buddhism / Toaism / Hinduism /Christianism

11. Which Scotsman took out a patent in 1876 that was the nucleus of the telephone?
    Alexander Fleming / Thomas Edison / George Stephenson / Alexander Bell

12. What is another name for a blood clot?
    Abrasion / Carcinoma / Bursitis / Thrombosis

13. Which of these countries is not landlocked? Paraguay / Bolivia / Andorra / Australia

14. The tibia and fibula are found where in the human body? Arm / Lower leg / Ribcage / Fingers

15. "Facebook" was launched in what year?
    1990 / 1994 / 2004 / 2009

16. Where is it believed that fireworks were invented?
    China / Mexico / Egypt / Greece

17. Which of these is found in the brain?
    Tibia / Thalamus / Vertebra / Humerus

18. Which of these is in North America?
    The Orzaks / The Urals / The Himalayas / The Pyrenees

19. Which of Galileo's achievements brought him into conflict with the church, resulting in his being confined to his house for the last years of his life?
    He attempted to measure the speed of light / He invented the thermometer / He said that Copernican view of the universe was correct / He attempted to measure the weight of air

20. What is the closest planet to the sun?
    Venus / Mercury / Saturn / Mars

21. On which continent are the native fauna called ostrich, lion, giraffe and okapi?
    Africa / South America / Australia / Asia

22. What do anthropologists study?
    Human Beings / Coal / Monkeys / Minerals

23. In the Alfred Hitchcock film "Psycho", where did the murder take place?
    In the bedroom / In the kitchen / On the front porch / In the shower

24. Where would one find a hypotenuse?
    Under the wing of a chicken / In the roof of a wooden building / As part of a right angled triangle / In a vehicle's gearbox

25. Who has won the most Olympic Gold medals? Larrisa Latynina, URSS / Paavo Nurmi, Finland / Michael Phelps, USA / Mark Spitz, USA

26. What does the chemical symbol Fe stand for?
Iron / Gold / Silver / Cheese

27. Michael J Fox played which character in the "Back to the Future" trilogy (1985-1990)?
Mickey McMouse / Marty McFly / Morris McAustin / Maurice McKee

28. In medicine, what do the s BMI stand for?
Bionic Machine Implants / Biochemical Mortuary Investigators / British Medical Institute / Body Mass Index

29. A single flame gas burner frequently used in student science laboratories is named after whom?
John Tilley / Michael Faraday / Sir Humphry Davy / Robert Bunsen

30. **A sabre is what type of weapon?**
Rifle / Spear / Crossbow / Sword

31. (*Attention Check*) What day comes after Tuesday?
Monday / Wednesday / Thursday / Friday