19- Lies and Biased Evaluation in a Real-Effort Experiment[1]

Julie Rosaz[2]

Marie Claire Villeval[3]

April 2011

Preliminary version

Abstract

This paper reports on the results of a laboratory experiment in which workers perform a real-effort task and supervisors report the workers' performance. The report is non verifiable and determines the earnings of both the supervisor and the worker. We find that the majority of supervisors are willing to bias their report to earn more. While selfish black lies and altruistic white lies (according to Erat and Gneezy's terminology) are almost nonexistent, both selfish black lies and Pareto white lies are frequent. In most situations, making the second order beliefs more salient affects neither the propensity of lying nor the nature of lies. There is a strong correlation between the second-order beliefs and the decision to lie or not to lie, suggesting that guilt aversion plays a role.

JEL Classification: C91, D82, M52.

Keywords: Lies, deception, self-image, guilt aversion, evaluation, experiments.

1 The authors are grateful to participants at the world meeting of the Economic Science Association in Copenhagen and at the conference of the French Experimental Economics Association in Grenoble for comments. We thank A. Brut, K. Straznicka, and I. Vialle for research assistance in the laboratory. Financial support from the EMIR program of the French National Agency for Research (ANR BLAN07-3-185547) is gratefully acknowledged.

2 Université de Lyon, Lyon, F-69003, France ; Université Lyon 2, Lyon, F-69007, France ; CNRS, GATE Lyon St Etienne, Ecully, F-69130, France. E-mail: rosaz@gate.cnrs.fr

3 Université de Lyon, Lyon, F-69003, France ; Université Lyon 2, Lyon, F-69007, France ; CNRS, GATE Lyon St Etienne, Ecully, F-69130, France. E-mail: villeval@gate.cnrs.fr

I Introduction

In this paper, we study the issues related to the honesty of the appraisal of agents by their supervisors when the payoffs of the supervisor and the agent depend on the appraisal of the agent's performance and when performance is not verifiable. Indeed, several studies in personnel economics have shown that the evaluation of employees can be biased when information about the employees' effort and ability is imperfect (see for example Gibbs (1991); Prendergast and Topel (1993); Prendergast (1999); Breuer, Nieken, and Sliwka (2010)). This is particularly the case when the evaluation is subjective, which is more and more frequently the case in companies. There are many reasons for which an appraisal can be voluntarily biased. For example, supervisors may not be willing to differentiate between their subordinates. They may hesitate when announcing a poor performance to avoid employees' discouragement. They may also report a level of performance higher than the actual level to artificially improve the reputation of the unit. Conversely, supervisory bias in an appraisal can also distort the agents' performance downwards. This is for example the case when a supervisor has once and for all judged an employee as bad and merely seeks evidence in support of this judgment (this corresponds to the "Matthew effect", Gabris and Michell (1989)). These examples suggest that in real settings, supervisors may have a choice not only between an honest and a biased appraisal, but also between various types of biases.

While psychologists have been interested in deception for a long time (see Hyman (1989) for a review; DePaulo et al. (1996); Tyler, Feldman, and Reichert (2006)), economists have become interested in lying behavior more recently. Gneezy (2005) studies deception in a sender-receiver game. Analyzing the role of incentives in the decision to lie, he shows that a fraction of people care about their opponent's payoff in deciding whether or not to lie. Erat and Gneezy (2009) propose a taxonomy of lies distinguishing between "black" and "white lies". "Selfish black lies" increase the player's payoff but decrease the other side's payoff. "Spite black lies" decrease both sides' payoffs. "Pareto white lies" increase both sides' payoffs, while "altruistic white lies" increase the other side's payoff but are costly to the decider. Erat and Gneezy (2009) experimentally test truth-telling versus each type of lie and conclude that the propensity to lie depends on the nature of the lie and the harm or benefit it causes to others relative to oneself.

Mohnen and Pokorny (2006) study the lying behavior in an employer-employee relationship. A principal and agent are matched for a two round game. The agent's ability is determined by a random draw and can be either low or high. Agents choose an effort level without knowing their ability level. Principals observe the actual productivity and learn the actual ability of their agents

and may send feedback. Indeed, they choose whether to send a feedback and whether to send a true feedback. They find that principals are less likely to lie when their agent has a low ability.

In this paper, we have designed an original experiment in which workers have to perform a real-effort task (counting the occurrence of letters in paragraphs) and supervisors have to report to the experimenter their worker's performance without any risk of verification. The structure of payoffs has been chosen such that the supervisors may have to choose between reporting the truth and telling a lie but they may also have the opportunity to tell different types of lies by increasing or decreasing their worker's performance. In contrast with Erat and Gneezy (2009), we can observe choices between different subsets of selfish black lies, spite black lies, Pareto white lies and altruistic white lies. Most of the existing experiments on deceptive behaviour involving two players use cheap talk such that the first mover sends a message on the state of nature or on his own ability (Gneezy (2005); Erat and Gneezy (2009); Lundquist et al. (2009); Sutter (2009)) or a promise (Charness and Dufwenberg (2006); Charness and Dufwenberg (2008); Ellingsen and Johanneson (2004); Vanberg (2008)) that is expected to influence the second mover's decision. In our experiment there is no cheap talk and it is the second mover who has an opportunity to cheat on the first mover's real effort, which better represents biased supervisory appraisals. Another approach of deceptive behaviour is based on individual decision-making where subjects can only cheat the experimenter (Fischbacher and Heusi (2008), Mazar, Amir, and Ariely (2008), and the literature on tax evasion)[4]. In our experiment, the supervisors can also cheat on the experimenter (for example by telling Pareto white lies) but with (positive or negative) known consequences on their workers' payoffs.

Standard economic theory predicts that lying derives from a comparison between the marginal costs and benefits of an action, with no consideration for the other's payoff. In our game, this means that supervisors should always report a medium performance, which would maximize their own payoff. However, several recent experiments on deceptive behaviour have revealed that a significant proportion of individuals do not lie as much as they should to maximize their earnings[5].

Guilt-aversion and lie-aversion are evoked to explain this behaviour (Charness and Dufwenberg (2006); Charness and Dufwenberg (2008)). While guilt-aversion is based on a correlation between the decision to tell the truth and the players' beliefs about others' beliefs, lie-aversion assumes that

4 See also Pruckner and Sausgruber (2006) for a natural field experiment on newspaper purchasing in the streets.

5 In Mazar, Amir, and Ariely (2008), subjects have to fill out a questionnaire and they are paid according to their number of correct answers. In the control treatment, the experimenter controls the performance whereas in the condition treatment the subjects report their performance. The number of correct answers reported is only 10% higher in the condition treatment. In Fischbacher and Heusi (2008) the participants roll a dice privately and report the result to the experimenter. The payoff is equal to the figure reported except for figure 6 that gives a zero payoff. Only 22% of the subjects maximize their payoff. Most people do not lie or lie but not enough to maximize their gains.

the likelihood of telling the truth is uncorrelated to second order beliefs. The maintenance of self-image could also explain the extent to which people are willing to lie (Mazar, Amir, and Ariely (2009)). In our experiment, we manipulate the saliency of second order beliefs to study how this can affect the supervisors' decision whether or not to truthfully report their workers' performance. In the baseline treatment, we ask the workers to indicate how many correct answers they expect their supervisor to report. In the second order belief treatment, we also elicit the supervisors' belief about their worker's answer to the previous question. This is the only difference between the two treatments. We can test whether the second order beliefs differ according to whether the supervisor is lying or not, for a given performance level. We can also test whether focusing the players' attention on second order beliefs changes lying decisions.

If we define a lie as a report that gives both players a payoff different from what they should have received by reporting the truth, we find that 34.84% of the supervisors lie. If one only considers the cases in which the workers did not actually perform at the medium level (i.e. when the supervisor can increase his own payoff by lying), these percentages are 54.62%. A majority of supervisors (64.71%) report medium performances. Spite black lies and altruistic white lies are rare. Selfish black lies represent 37.66% of the lies while the Pareto white lies represent 53.25% of the lies. The saliency of second order beliefs does not affect the reporting behaviour much. The workers anticipate a higher performance than what they actually do. The supervisors anticipate a higher proportion of lies predicted by their workers than what is actually the case. The decision to lie decreases when the supervisor believes that the worker expects him to tell the truth. This last result suggests that guilt-aversion has an important role in the decision to lie.

The remainder of this paper is organized as follows. Section II presents briefly the literature on the determinant of honesty and lying behaviour. Section III describes the experimental design and procedures and presents the theoretical predictions. Section IV presents the results of the experiment. In Section V, we discuss our results and conclude.

II Related literature on the determinants of honesty

Contrary to predictions of economics-of-crime models based on standards assumptions, the experimental literature on deception shows that a significant fraction of people behave honestly or do not lie as much as they should in order to maximize their earnings. This complements evidence from psychology and neuroscience that honesty is an important guideline of human behaviour (see Sip et al. (2008)). For example, Erat and Gneezy (2009) show that while many people are reluctant

to tell even a Pareto white lie (39% of their participants), a significant proportion of people (30%) are willing to tell altruistic lies at their own expense, questioning the interactions between social preferences and lie aversion. While deceptive behaviour has been shown to depend on the size of lies and the strength of promises (Lundquist et al. (2009)), an interesting discussion is about the fundamental determinants of honesty.

Studying promises and cooperation, Charness and Dufwenberg (2008) investigate the role of guilt-aversion and lie-aversion in the players' decisions to keep their promises. Guilt-aversion assumes that there is a positive correlation between the player's second order beliefs and the decision to tell the truth. The feeling of guilt depends on the subject's belief about the other's belief. On the contrary, lie-aversion assumes that lying brings disutility (??) regardless of the impact of the lie on beliefs, and therefore predicts that the likelihood of telling the truth is uncorrelated with the player's beliefs about the others' beliefs. In their experiment, Charness and Dufwenberg (2008) find more evidence of guilt aversion (like in Charness and Dufwenberg (2006)) without rejecting the presence of some lie-aversion. This result is in line with the literature studying actions and beliefs (for example Ellingsen, Johanneson, and Lilja (2009)).

However, other recent experiments have minimised the effect of guilt aversion and conclude that people have an intrinsic preference for promise keeping (Ellingsen et al. (2010); Vanberg (2008)). Using a dictator game, Ellingsen et al. (2010) inform the dictator about the recipient's beliefs before he plays. Their results reject a correlation between beliefs and actions. Vanberg (2008) rejects an expectation-based explanation for promise keeping and concludes that people have a preference for promise keeping *per se*.

For their part, Mazar, Amir, and Ariely (2008) suggest that people do not lie as much as they could because an internal reward system exerts control over their behaviour. People are influenced by the way they view and perceive themselves. Attaching a great importance to their self-image, they may make dishonest decisions but they may not fully exploit their possibilities of lying because this would force them to change their self-concept.

III Experimental Design

III.1 The game

This real-effort experiment consists of two different treatments and is based on a between-subjects design. The experiment was carried out with pen and paper to make the effort exerted at the task by the workers more visible to the supervisors. Indeed, the Baseline treatment involves workers and

supervisors, in equal proportions. The workers and the supervisors are located in two different rooms without any possibility of communicating either visually or verbally. Each participant receives the instructions for both roles. Each worker is matched randomly with an anonymous supervisor. The workers have to perform a task for 10 minutes. They receive sheets of paper displaying several paragraphs. These paragraphs include words that are randomly combined. The task consists of counting the occurrence of four random letters in each paragraph (like in Hogarth and Villeval (2010)). The four letters may differ from one paragraph to another. There is a maximum of 50 letters to be counted in total. All the participants receive paragraphs with the same words though these words are arranged in a different order. The workers do not receive any feedback on their performance while performing the task. Before performing the task, they are informed that their answers are to be checked by their supervisor and that their earnings will depend on the number of correct answers reported by their supervisor. The relationship between the workers' and the supervisors' payoffs and the reported performance is also made common information from the very beginning of the session.

Once the 10 minutes have elapsed, the workers have to answer the following question: *«How many correct answers do you believe your supervisor will report?»*. This question is incentivized: a subject earns one more Euro if his prediction is correct (plus or minus one). Then, the workers have to answer a second question: *«How many correct answers do you believe you have made?»*. The second question is not incentivized because we do not want the workers to learn with certainty when being paid that their actual performance has been actually under or over-estimated by their supervisor. The comparison between these two answers tells us whether or not the workers expect to be deceived , and if so, which type of lie they anticipate. Then an assistant collects the sheets in the workers' room and distributes them to the supervisors in the other room. Meanwhile, the supervisors are given the list of correct answers for each paragraph in order to minimize involuntary correction errors. They are asked to count the number of correct answers in their worker's sheet. The supervisors are told that they have to verbally report their worker's performance in the payment room to an assistant who is not aware of the content of the experiment. It is also made common information that the exercise sheets remain in the laboratory together with the instruction sheets. Once all the supervisors have finished, they are called one by one to the payment room where they report their worker's performance and get paid according to their verbal report. When leaving the supervisors' room, they put all the sheets in a box, which strengthens the notion that there can be no control of verbal reports. Once all the supervisors have been paid and have left the institute, the workers are called one by one to the payment room where they are paid according to their supervisor's verbal report and the correctness of their prediction.

The *Second Order Belief treatment* is identical to the baseline treatment except that once all the supervisors have finished checking their workers' answers, we elicit the supervisors' second order beliefs about their workers' expectations on their verbal report. Precisely, we ask them to answer the following question: *«What do you think your worker answered to the following question: 'How many correct answers do you believe your supervisor will report'?»*. The supervisors receive one more Euro if their prediction is correct (plus or minus one). The comparison between the two treatments aims at investigating whether both the propensity to lie and the nature of lies are modified when the supervisors are forced to think about their image in their workers' eyes. If supervisors lie more or less in this treatment than in the baseline, this suggests that they have updated their self-concept (Mazar, Amir, and Ariely (2008)).

In both treatments, the structure of payoffs is made common information. There are five payoff levels for the workers and the supervisors depending on the category of the reported performance. Table 1 displays the workers' and supervisors' payoffs for each category of performance.

[Table 1 about here]

These performance categories correspond respectively to a very low, low, medium, high and very high level[6]. Reaching an immediately superior category of performance increases the workers' payoff by two points and the supervisor's payoff by five points. Moreover, once the high performance level has been reached, a transfer is made from the supervisor to the worker that can be thought of as a bonus. This transfer increases the workers' payoff by 5 points and decreases the supervisors' payoff by 14 points when moving from the medium to the high performance level. This structure of payoffs is clearly arbitrary but it allows us to observe all types of lies we are interested in, as explained in the next sub-section.

In addition, the supervisors' degrees of risk attitude and inequity aversion are elicited in both treatments while the workers perform the task. This aims at investigating whether the attitudes towards risk and advantageous inequity are correlated with lie aversion. The Holt and Laury (2002)'s procedure has been used to test for risk attitudes. The supervisors completed a ten-decision questionnaire. Each decision consists of a choice between two lotteries, option A and option B. The payoffs for option A (the safer lottery) are either €2 or €1.60, whereas option B pays either €3.85 or €0.10. In the first decision, the probability of the high payoff for both options is one tenth. In the second decision, the probability increases to two tenths, and so on. The high-payoff probability in

6 Note that in the instructions, we only use the number categories and we do not refer to low, medium or high categories.

each decision increases as the number of the decision increases, up to the tenth decision were payoffs are certain. When the probability of the higher payoff is large enough (1/2), subjects should switch from option A to option B. Risk neutrality corresponds to a cross at the fifth decision, while risk loving subjects are expected to switch earlier and risk averse subjects at the sixth decision or after.

To approach attitudes towards inequality, we use a modified dictator game. All the supervisors made 21 decisions in the role of the dictator, knowing that their actual role (either the dictator -"player X"- or the receiver -"player Y") would be determined by tossing a coin in the payment room at the end of the session. Each decision consists of a choice between two payoff distributions between player X and player Y (option A and option B). Option A always consists of an equal payoff for the two players (5 points each). In the first decision, option B pays 20 points to player X and 0 to player Y. In the second decision, option B pays 19 points to player X and 1 point to player Y, and so on. In option B, player X's payoff decreases while player Y's payoff increases as the number of the decision increases. An inequality-neutral subject should choose option B for the 16 first decisions and then switch to option A for the remaining decisions. In the first ten decisions, player X who chooses option B always earns more than player Y and more than if choosing option A. Therefore, choosing option A in some of the first ten decisions indicates a strong degree of advantageous inequity aversion and gives us an indication of guilt. In decisions 12 to 16, player X who chooses option B always earns less than player Y but more than choosing option A. Therefore, choosing option A between decisions 12 and 16 indicates disadvantageous inequity aversion. Finally, in decisions 17 to 21, player X who chooses option B always earns less than player Y and less than choosing option A but allows player Y to earn much more than by choosing option A. Choosing option B at least once in the last five decisions suggests altruistic preferences. Note that we are not interested in determining a precise measure of attitudes towards inequality but this parsimonious test gives us very broad indications about these attitudes.

At the end of the session, a random draw determines whether it is the test of risk aversion or the test of inequality aversion that gives rise to payment. There is no feedback on these tests before the end of the session.

III.2 Predictions

Considering the payoffs displayed in Table 1, a supervisor who aims at maximizing his own monetary payoff should always report a number of correct answers comprised in between 22 and 28 regardless of the worker's actual performance. This prediction holds for both treatments since the

second order belief question is not related to the payoff matrix of the reporting decision.

This standard prediction is compatible with two categories of lies, the selfish black lies and the Pareto white lies (according to the terminology of Erat and Gneezy (2009)). A supervisor tells a selfish back lie that increases his own payoff but reduces his worker's payoff if he reports a medium performance (between 22 and 28) whereas the actual performance is either high (between 29 and 35) or very high (above 35). A supervisor tells a Pareto white lie that improves both players' payoffs if he reports a medium performance (between 22 and 28) whereas the actual performance is either very low (below 15) or low (between 15 and 21).

From a behavioural point of view, however, behaviour may deviate from these predictions. Consider the Baseline treatment. First, if they are lie-averse or guilt-averse, supervisors may speak the truth although they could earn more by lying. Or they may lie but not as much as what they should to maximize their payoff. Second, altruistic supervisors may tell altruistic white lies by reporting a number of correct answers that increase the workers' payoffs but reduce their own payoff. This is the case if they report a high or a very high performance whereas the actual performance is medium. Finally, some supervisors may also make spite black lies by reporting a number that decreases both their own and their workers' payoffs. This is the case if they report a low or very performance (below 21) when they observe a medium performance between 22 and 28 and 50. The motivation behind this behaviour is however less clear inasmuch as there is no direct interaction between the worker and the supervisor.

Behaviour could differ in the Second Order Belief treatment compared with the baseline if the treatment manipulation strengthens the correlation between beliefs and actions. Guilt aversion may lead more supervisors to speak the truth in this treatment. But it may be also the case that if they believe that others believe in their generosity, honest supervisors may be willing to tell altruistic white lies. Therefore, in this treatment not only the likelihood of a lie but also its nature may differ.

Table 2 summarizes the various types of lies depending on the true number of correct answers and the supervisors' report, according to the Erat and Gneezy (2009)'s terminology. The light grey cells along the diagonal correspond to truth-telling and the dark grey cells to the standard equilibrium of the game.

[Table 2 about here]

III.3 Procedures

The experiment consists of 23 sessions conducted at the laboratory of the GATE (Groupe d'Analyse et de Théorie Economique) research institute in Lyon, France. Between 12 and 30

individuals took part in each session, for a total of 442 participants invited via the ORSEE software (Greiner (2004)). The Baseline treatment was implemented in 11 sessions with 224 participants (48.66% were females) and the Second Order Belief treatment in 12 sessions with 218 participants (47.25% were females). No individual participated in more than one session. Two laboratories were used, one for the workers and the other one for the supervisors. The two rooms were located in the same corridor. Upon arrival, the subjects randomly drew a tag from a bag assigning them to one room and to a seat in this room. This procedure ensured that both the role assignment and the pairing of subjects were random.

All the instructions were distributed and read aloud in each room. A neutral wording was used in the instructions (see Appendix VIII). Workers were named "players A" and supervisors "players B". In the worker's room, after reading the instructions and answering to questions in private, the experimenter distributed the paragraph sheets. Then, the subjects performed the task for ten minutes. After the ten minutes elapsed, the experimenter distributed one sheet asking for the workers' beliefs about their supervisors' report. Then the experimenter took back this decision sheet and distributed a new sheet asking for the workers' beliefs about their actual performance. After this question was answered, the experimenter took back all the documents and the paragraph sheets were brought to the supervisors' room. Meanwhile, the workers answered a final demographic questionnaire. They were instructed to remain seated and silent until the supervisors were paid and had left the institute. They were allowed to read books or magazines and were also provided with sudoku puzzles.

In the supervisors' room, they completed a demographic questionnaire. Then, we elicited risk attitudes with the Holt and Laury (2002) procedure and we then elicited attitudes toward inequality. Next, the instructions for the main game were read aloud and questions were answered privately. Then, each participant received the paragraph sheet of his paired worker for evaluation. In the Second Order Belief treatment, once all the supervisors have completed the correction, they answered the prediction question. Last, the supervisors were called upon one by one and sent to the adjacent payment room. Before leaving the laboratory they left all the documents together (instructions, paragraph sheets, ...) in a basket.

In the payment room, a person who was not aware of the content of the experiment (this was made common information in the instructions) paid each participant in cash. Each supervisor reported his worker's number of correct answers and payment was made accordingly. In addition, each participant tossed a coin to determine whether the test of risk aversion or the test of inequality aversion would give rise to payment. If the risk elicitation task was randomly drawn, the subject rolled a ten-sided die to determine which decision would give rise to payment and played the

corresponding lottery. If the inequity aversion elicitation task was drawn, the participant tossed a coin to determine his role and selected one of the 21 decisions by extracting a ticket from a bag. If player X was drawn, the option chosen by the participant was implemented; otherwise, the decision of his paired player was implemented.

Once all the supervisors left the institute after payment, the participants in the role of workers were called one by one to get paid according to their supervisor's report. Moreover they received a €2 show-up fee and the payment for accurate prediction.

A session lasted less than one hour on average. Workers earned €8 and supervisors earned €9.5 on average.

IV Results

We first examined the supervisors willingness to lie depending on the actual performance of their workers, before characterizing the nature of lies and their distribution. Then, we analyzed the workers' expectations and the relationship between the supervisors' second order beliefs and their decision to lie or to report truthfully. Finally, we report an econometric analysis of the individual decision to lie.

Note that in this section, the data analysis is conducted at the level of the category of performance (payoff). The reported number of answers that do not correspond to the category of the worker's actual number of correct answers are considered as lies. We disregard lies that occur within a performance category (i.e. that does not affect the players' payoffs)[7]. These behaviours can be considered as errors.

IV.1 The willingness to lie and the distribution of lies

Table 3 displays the distribution of the workers' actual performance and the corresponding supervisors' reports in each treatment. The presence of lies can be identified directly.

[Table 3 about here]

If we consider the case where the true number of correct answers is included between 29 and 35[8],

7 37/221 (16.74%) subjects declare a number of correct answers that differs from the actual number of correct answers without affecting the player's payoffs. 17/37 (45.95%) decrease the actual number of correct answers.

8 When the true number of correct answers is included between 29 and 35, the supervisors have the choice between the truth, a selfish black lie and a Pareto white lie.

we observe that 40% (22/55) of the supervisors report the truth. 44% of the supervisors declare a medium performance in order to increase their payoff at the expense of their workers and 16% of the supervisors increase both players' payoffs by reporting a very good performance.

The maximization of one's own payoff should lead all supervisors to report a medium performance (between 22 and 28 correct answers). We observe in Table 3 that 64.71% (143/221) of the supervisors report a medium performance, while 41.18% of the workers actually perform in this category. If one only considers the Baseline treatment, these percentages are 63.39% and 38.39%, respectively. In the Second Order Belief treatment, they are 66.06% and 44.04%. Regarding both measures, we observe no statistically significant difference between the Baseline and the Second Order Belief treatments[9].

The fact that not all supervisors report the medium performance indicates that some of them refrain from lying or do not lie as much as they could (some supervisors lie although they do not report the medium performance). This finding is consistent with previous studies showing that people do not fully exploit their lying opportunities (Fischbacher and Heusi (2008); Mazar, Amir, and Ariely (2008)). In total in the Baseline treatment, 32.14% (36/112) of the supervisors lie and report a number that does not belong to the same category of performance than the actual number of correct answers. In the Second Order Belief treatment, the percentage is 37.61% (41/109). A test of proportion indicates that the difference is not significant ($p = 0.3934$). If one adopts a more strict definition of lies by only considering the situations in which workers have not given between 22 and 28 correct answers, the percentage of liars is 49.28% (34/69) in the Baseline and 60.66% (37/61) in the Second Order Belief treatment. This difference is not significant either ($p = 0.1934$).

A specificity of our design is that depending on the worker's actual performance, supervisors can tell different types of lies and can sometimes choose between various types of lies. Table 4 displays the distribution of decisions for each treatment for an actual performance different from 22-28[10], by distinguishing between spiteful black lies, selfish black lies, Pareto white lies and altruistic white lies, according to the terminology of Erat and Gneezy (2009).

[Table 4 about here]

9 A Kolmogorov-Smirnov test indicates that there is no significant difference in the distribution of workers' actual categories of numbers of correct answers between the two treatments ($p = 0.793$, two-tailed). The same test also concludes that the distribution of the supervisors' reported categories of numbers of correct answers is not different between treatments ($p = 0.857$).

10 The distribution of decisions for an actual performance between 22-28 shows that most of the supervisors report the true category. As we can see in Table 3, only 6/91 (6.59%) supervisors choose to lie. Two supervisors tell spite black lies, whereas the others tell altruistic white lies.

Not surprisingly, as their rationality is unclear, spite black lies that diminish the payoffs of both the supervisor and the worker are virtually non-existent(3 cases out of 77 lies in total, 3.90%). Altruistic white lies are also not more frequent (4 observations, 5.19%). In contrast, selfish black lies represent 37.66% of the lies (29/77) and Pareto white lies represent 53.25% of the lies (41/77). These data from both treatments are pooled since the differences per treatment are not significant (proportion tests, $p = 0.8351$ for the selfish black lies and $p = 0.9384$ for the pareto white lies). The saliency of second order beliefs does not greatly affect the reporting behaviour.

IV.2 Workers' beliefs and supervisors' second order beliefs

In the second order belief treatment, supervisors have to predict their workers' answers to the following question: *«How many correct answers do you expect your supervisor to report?»*. This question is incentivized and thus they should report truthfully. Through his answer, the supervisor indicates his belief that his worker anticipates a lie from him. We can relate this belief to the supervisor's actual report. We find that 85.37% (35/41) of the supervisors who tell a lie believe that their worker likewise expects them to tell a lie. This is the case of only 30.88% (21/68) of the supervisors who tell the truth. If we consider only the case where the actual performance differs from the medium performance category, i.e. the cases where the principals should lie, 89.19% (33/37) of the supervisors who lie and 45.83% (11/24) of the supervisors who tell the truth, expect that their worker anticipate a lie.

Table 5 and Figure 1 give for each performance category, in the Second order Belief treatment, the workers' actual performance, the workers' beliefs about their performance, the workers' beliefs about their supervisor's report, the supervisors' reports and their second order beliefs.

[Table 5 about here]

[Figure 1 about here]

We find that the workers tend to overestimate their performance (see panels a and b in Figure 1). A Wilcoxon signed-rank test rejects the equality of distribution between the actual number of correct answers and the workers' belief about their number of correct answers ($p < 0.001$). Although the workers expect some lies from their supervisors, i.e. the distribution of the workers' beliefs about their performance is statistically different from the distribution of their beliefs about the supervisor's report, $p < 0.001$ (see panels b and c in Figure 1), they predict a higher reported

number of correct answers than the supervisors report ($p < 0.001$). Moreover, the supervisors do not anticipate the workers' beliefs about their report (see panels c and d in Figure 1). A Wilcoxon signed-rank test rejects the equality of distribution between the supervisors' belief and the workers' belief about their report of the number of correct answers ($p = 0.003$).

However, they do not base their belief on the same performance level. The workers base their prediction on their belief of their performance[11] whereas the supervisors observe the actual performance. In order to control their base rate, we construct dummy variables indicating whether the worker predicts that the supervisor will lie and whether the supervisor believes that the worker predicts a lie. Thus, a lie predicted by the worker corresponds to a case where the worker's belief about his performance is different from his belief about the supervisor's report. The supervisor's belief that the worker predicts a lie is the situation where the supervisor's belief about the worker's belief about the supervisor report is different from the actual performance. A test of proportion confirms that supervisors do not correctly anticipate the workers' beliefs. The proportion of supervisors who believe that their workers predict a lie is higher than the actual proportion of lies predicted by the workers ($p = 0.0145$). Moreover, the actual proportion of lies and the proportion of lies predicted by the workers are not different ($p = 0.5557$), whereas the actual proportion of lies is smaller than the proportion of supervisors who believe that their workers predict a lie ($p = 0.0205$).

In conclusion, the workers overestimate their performance. However, they correctly predicted the proportion of lies whereas the supervisors overestimate the workers' expectations of lies. Thus, the supervisors lie less than what they predicted from their workers' expectations, possibly to maintain a certain positive self-image and to avoid guilt. Another explanation for this overestimation of the predicted lies could be that the supervisors try to justify their lying behaviour by reporting a higher proportion of lies predicted by the workers. However, the supervisors are incentivized for their predictions, so we cannot exclude the self-justification reason, but it cannot be the only explanation for their overestimated predictions.

IV.3 The determinants of the decision to lie

We estimate a probit model to explain the lying decision in the pooled treatment data. We introduce a dummy variable indicating the Second Order Belief treatment. We include a dummy variable that indicates if the actual performance is different from 22-28 to control for the possible lying behaviour. Moreover, we introduce individual characteristics such as gender, experienced subjects[12]. Risk attitude is given by a dummy variable indicating if the subject is a risk lover[13].

11 Acknowledgment: predicted performance may be biased since it was not paid.

12 A subject has experience if he had participated at least in one experiment during the past year.

Finally, we control the inequality aversion through three dummy variables: advantageous inequality averse, disadvantageous inequality averse and altruistic subjects. A subject is considered averse to advantageous inequality if he chooses the equal payoff (option A) between decisions 1 and 10 at least once, in the modified dictator game. A subject is averse to disadvantageous inequality if he chooses the equal payoff between decisions 12 and 16 at least once. Finally, a subject is altruistic if he chooses option B at least once between decisions 17 to 21, i.e. he chooses option B where he earns less than the other player and less than choosing option A, but allows the other player to earn more. Table 6 displays the results of this regression.

[Table 6 about here]

Table 6 confirms that being in the Second Order Belief treatment has no impact on the lying decision. Unsurprisingly, supervisors who should lie, i.e. the actual performance is different from 22-28, lie more. We find that the more a subject has participated in previous experiments, the more likely he is to lie. Finally we find that the subjects who are averse to advantageous inequality lie less.

Next, we restrict the analysis to the Second Order Belief treatment in order to study the impact of the supervisors' predictions on their lying choice. We estimate a probit model to explain the lying decision in the Second Order Belief treatment. We add a dummy variable indicating if the supervisor believes that his worker expects him to lie. The results are reported in Table 7.

[Table 7 about here]

Table 7 indicates that the supervisors who believe that their worker expects them to lie, are more likely to lie than the others. This result confirms the findings of the previous subsection: the supervisors who believe that the workers expect the truth are less likely to lie. This result suggests that supervisors' beliefs plays an important role in their decision to lie. This finding is in line with Charness and Dufwenberg (2008)'s result, and gives more evidence for guilt-aversion. We find also that altruistic subjects are marginally less likely to lie.

V Conclusion

13 A subject is risk lover if he chooses less than 5 times, option A (safer lottery) in the Holt and Laury (2002)'s game.

This paper reports on the results of a laboratory experiment in which workers have to perform a real-effort task and supervisors have to report the workers' performance. We study some issues related to the honesty of the appraisal of agents by their supervisor when the payoffs of the supervisor and the agent depend on the appraisal of the agent's performance and when performance is not verifiable.

We designed two treatments. The Second Order Belief treatment is identical to the baseline treatment except that once all the supervisors have finished checking their worker's answers, we elicit the supervisors' second order beliefs about their workers' expectations on their verbal report.

We find that the majority of supervisors are willing to bias their report in order to earn more. While spite black lies and altruistic white lies (according to Erat and Gneezy (2009)'s terminology) are almost non-existent, both selfish black lies and Pareto white lies are frequent. In most situations, making the second order beliefs more salient affects neither the propensity of lying nor the nature of lies. There is a strong correlation between the second-order beliefs and the decision to lie or not to lie, suggesting that guilt aversion plays an important role.

This experiment is the first one in which workers have to perform a real-effort task and supervisors have to report their workers' performance in order to determine both players' payoffs. Moreover, the supervisors may choose between different types of lies depending on the workers' actual performance. These differences result in more moderate conclusions than previously found in the literature. Indeed, Erat and Gneezy (2009) found that around 30% of participants are willing to tell an altruistic white lie when their choice consists in telling the truth or an altruistic white lie. We do not find this result. In our experiment, when subjects may choose between telling the truth, an altruistic white lie or a spite black lie, 4.40% of the supervisors choose to tell an altruistic white lie, 2.20% tell a spite black lie and 93.40% tell the truth. Moreover, we indirectly support Charness and Dufwenberg (2008)'s evidence of the impact of guilt aversion on lying behaviour. However, we were unable to reject lie-aversion.

Indeed some subjects do not lie although by doing so they could increase both players' payoffs. Our results may have some implications for performance appraisals in firms. Indeed, supervisors do not lie as often as they should. When they lie, they use it in order to increase their payoffs. However, we do not take into account the long term relationship between the supervisor and the worker. To complete our work we should run the experiment with a repeated game in order to study the impact of lies on the next periods effort choices. Indeed, reputation may be strategic for lying behaviour in performance appraisals. A supervisor may make different choices when the lie has some implications on future interaction. A supervisor may be reluctant to reduce the performance of a

good worker in order to avoid deception and have more incentive to increase bad performance in order to increase the worker's motivation. Our results show that second order beliefs are correlated with lying behaviour, suggesting a role for guilt aversion. An extension would be studying the effect of explicit workers' beliefs on the supervisors' willingness to lie. An additional treatment would consist of giving the workers' beliefs to the supervisors before they report their evaluation. We could then observe more directly the impact of empathy on the decision to lie.

References

Battigalli, P., and M. Dufwenberg. 2007. "Guilt in Games." *American Economic Review* 97:170–176.

Breuer, K., P. Nieken, and D. Sliwka. 2010. "Social Ties and Subjective Performance Evaluations: An Empirical Investigation." Discussion Paper No. 4913, IZA.

Charness, G., and M. Dufwenberg. 2008. "Broken Promises: An Experiment." University of California at Santa Barbara, Economics Working Paper Series No. 10-08, Department of Economics, UC Santa Barbara.

—. 2006. "Promises and Partnership." *Econometrica* 74:1579–1601.

DePaulo, B.M., J.A. Epstein, D.A. Kashy, S.E. Kirkendol, and M.M. Wyer. 1996. "Lying in Everyday Life." *Journal of Personality and Social Psychology* 70:979–995.

Ellingsen, T., and M. Johanneson. 2004. "Promises, Threats and Fairness." *Economic Journal* 114 (495):397–420.

Ellingsen, T., M. Johanneson, and J. Lilja. 2009. "Trust and Truth." *Economic Journal* 119 (534):252–276.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik. 2010. "Testing guilt aversion." *Games and Economic Behavior* 68:95–107.

Erat, S., and U. Gneezy. 2009. "White Lies." Working paper, Rady School of Management, UC San Diego.

Fischbacher, U., and F. Heusi. 2008. "Lies in Disguise. An experimental study on cheating." TWI Research Paper Series No. 40, Thurgauer Wirtschaftsinstitut, Universität Konstanz.

Gabris, G.T., and K. Michell. 1989. "The impact of merit raise scores on employee attitudes; the Mattehew effect of performance appraisal." *Public Personnel Management* 17 (4).

Gibbs, M.J. 1991. "An Economic Approach to Process in Pay and Performance Appraisals." Discussion paper, Harvard Business School.

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

Greiner, B. 2004. "An Online Recruitment System for Economic Experiments." *Kurt Kremer, Voler*

Macho (eds). Forschung und wissenschaftliches Rechnen, GWDG Bericht 63, Gttingen: Ges. fr Wiss. Datenverarbeitung, pp. 79–93.

Hogarth, R.M., and M.C. Villeval. 2010. "Intermittent Reinforcement and the Persistence of Behavior: Experimental Evidence." IZA Discussion Papers No. 5103, Institute for the Study of Labor (IZA), Jul.

Holt, C., and S. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92 (5):1644–1655.

Hyman, R. 1989. "The psychology of deception." *Annuel review of Psychology* 40:133–154.

Lundquist, T., T. Ellingsen, E. Gribbe, and M. Johannesson. 2009. "The aversion to lying." *Journal of Economic Behavior & Organization* 70:81–92.

Mazar, M., O. Amir, and D. Ariely. 2009. "More Ways to Cheat -Expanding the Scope of Dishonesty." Working paper, Duke University.

Mazar, N., O. Amir, and D. Ariely. 2008. "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research* 45 (6):633–644.

Mohnen, A., and K. Pokorny. 2006. "Is Honesty the Best Policy? An Experimental Study on the Honesty of Feedback in Employer-employee Relationships." Working paper, SSRN.

Prendergast, C. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37:7–63.

Prendergast, C., and R. Topel. 1993. "Discretion and bias in performance evaluation." *European Economic Review* 37:355–365.

Pruckner, G., and R. Sausgruber. 2006. "Trust on the Streets: A Natural Field Experiment on Newspaper Purchasing." Discussion Papers No. 06-01, University of Copenhagen. Department of Economics, Feb.

Sip, K.E., A. Roepstorff, W. McGregor, and C.D. Frith. 2008. "Detecting deception: the scope and limits." *Trends in Cognitive Sciences* 12:48–53.

Sutter, M. 2009. "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams." *Economic Journal* 119:47–60.

Tyler, J.M., R.S. Feldman, and A. Reichert. 2006. "The price of deceptive behavior: Disliking and lying to people who lie to us." *Journal of Experimental Social Psychology* 42:69–77.

Vanberg, C. 2008. “Why do people keep their promises? An experimental test of two explanations.” *Econometrica* 76 (6):1467–1480.

VI Appendix: Tables

Reported number of correct answers		Worker's payoff	Evaluator's payoff
0-14	(very low)	12	13
15-21	(low)	14	18
22-28	(medium)	16	23
39-35	(high)	23	14
36-50	(very high)	25	19

Table 1: Workers' and supervisors' payoffs per performance category in points[12].

		Reported number of correct answers by the supervisor				
		0-14	15-21	22-28	29-35	36-50
True number of correct answers	0-14	Truth	Pareto White	Pareto White	Pareto White	Pareto White
	15-21	Spite black	Truth	Pareto White	Altruistic White	Pareto White
	22-28	Spite black	Spite black	Truth	Altruistic White	Altruistic White
	29-35	Spite black	Selfish black	Selfish black	Truth	Pareto White
	36-50	Spite black	Spite black	Selfish black	Spite black	Truth
Payoff worker, supervisor		12;13	14;18	16;23	23;14	25;19

Table 2: Distribution of the various categories of lies

All Baseline SOB		Reported number of correct answers by the supervisor 0-14 (12;13)	15-21 (14;18)	22-28 (16;23)	29-35 (23;14)	36-50 (25;19)	Total
True number of correct answers	0-14	**4** 4 0	**1** 0 1	**4** 1 3			**9** 5 4
	15-21		**20** 13 7	**25** 15 10		**2** 0 2	**47** 28 19
	22-28		**2** 1 1	**85** 41 44	**1** 1 0	**3** 0 3	**91** 43 48
	29-35			**24** 13 11	**22** 11 11	**9** 3 6	**55** 27 28
	36-50			**5** 1 4	**1** 1 0	**13** 7 6	**19** 9 10
Total		**4** 4 0	**23** 14 9	**143** 71 72	**24** 13 11	**27** 10 17	**221** 112 109

Table 3: Distribution of workers' actual performance and supervisors' reported performance

Baseline 69 observations	**Lie** 34/69 (49.28%)				**Truth** 35/69 (50.72%)
	Black 15/34 (44.12%)		**White** 19/34 (55.88%)		
	Spite 1/15 (6.67%)	**Selfish** 14/15 (93.33%)	**Pareto** 19/19 (100%)	**Altruistic** 0/19 (0%)	
Second Order Belief treatment 61 observations	**Lie** 37/61 (60.66%)				**Truth** 24/61 (39.34%)
	Black 15/37 (40.54%)		**White** 22/37 (59.46%)		
	Spite 0/15 (0%)	**Selfish** 15/15 (100%)	**Pareto** 22/22 (100%)	**Altruistic** 0/22 (0%)	

Table 4: Distribution of decisions by treatment for an actual performance different from 22-28

	Actual performance	Workers' beliefs about their performance	Workers' beliefs about the supervisor's report	Supervisors' beliefs about the workers' report	Reported performance
0-14	3.67%	4.59%	4.59%	2.75%	0%
15-21	17.43%	1.83%	3.67%	4.59%	8.26%
22-28	44.04%	13.76%	37.61%	62.39%	66.06%
29-35	25.69%	41.28%	21.10%	12.84%	10.09%
36-50	9.17%	38.53%	33.03%	17.43%	15.60%

Table 5: Distribution of actions and beliefs in the Second Order Belief treatment

Dependent variable: indicator for lying behavior		
SOB treatment	.1121	(.0681)
Actual performance different from 22-28	.5021***	(.0510)
Female	.0149	(.0678)
Experienced subject	.2032***	(.0779)
Risk lover	-.1272	(.0814)
Advantageous inequality averse	-.1723**	(.0670)
Disadvantageous inequality averse	-.0291	(.0744)
Altruistic	-.0281	(.0706)
Observations	221	
LR Chi^2	84.28	
Prob > Chi^2	.0000	
Pseudo R^2	.2950	

Probit regression model on pooled data of Baseline and Second Order Belief treatments. Marginal effects are reported. Standard errors in parentheses. Levels of significance: * 10%; ** 5%; *** 1%.

Table 6: Determinants of lying behavior

Dependent variable: indicator for lying behavior		
Supervisors believes that his worker expects him to tell a lie	.4267***	(.1011)
Actual performance different from 22-28	.4277***	(.0946)
Female	-.1572	(.1092)
Experienced subject	.2058	(.1354)
Risk lover	.0192	(.1514)
Advantageous inequality averse	-.2087*	(.1125)
Disadvantageous inequality averse	-.1664	(.1058)
Altruistic	-.2146*	(.1128)
Observations	109	
LR Chi^2	65.34	
Prob > Chi^2	.0000	
Pseudo R^2	.4527	

Probit regression model on the Second Order Belief treatment. Marginal effects are reported. Standard errors in parentheses. Levels of significance: * 10%; ** 5%; *** 1%.

Table 7: Determinants of lying behavior in the Second Order belief treatment

VII Appendix: Figure

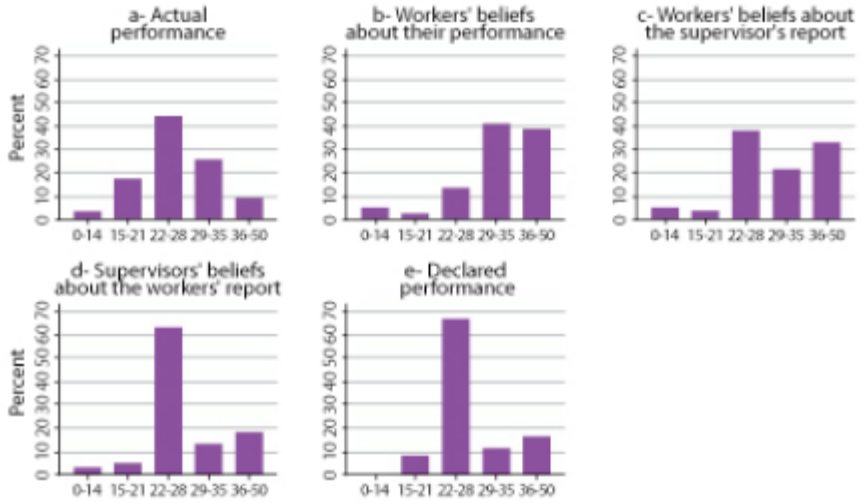

Figure 1: Actual and reported performance, beliefs and second-order beliefs in the Second Order Belief treatment

VIII Appendix: Instructions for workers (original in French)

The instructions for the evaluators are available from the authors upon request.

We thank you for participating in this experiment in economics.

This session will have two independent parts. We have given you the instructions for the first part. You will receive the instructions for the second part later.

At the end of the session, your earnings from both parts will be added. You will be paid in another room.

The earnings of each part will be added at the end of the session. It is forbidden to communicate with other subjects during the experiment.

Part 1 (instructions for participants A and B)

During this part, your earnings are expressed in points, with the following conversion rate: 3 points = €1.

There is two types of subjects (in equal number): participants A and participants B. Your type have been randomly assigned to you at the beginning of the experiment by the color of the ticket you draw.

All the participants in this room are participants A. All the participants B are in another room. Each participant A is paired with a participant B.

Participant A's role

The participants A have to perform a task during 10 minutes.

It consists on counting the occurrence of letters in a paragraph.

Each paragraph is composed by French words generated randomly and it has no meaning. For each paragraph, it is asked to count the occurrence of 4 letters.

The answer should be written in the case corresponding to the counted letter.

Here is an example of paragraph:

Problème		Nombre de			
		"a"	"d"	"n"	"t"
1	trichomie innovez calculaient commettiez tabla implantes tomettes accusateurs imprudente torpillions balancera balancier planifiions frissonnants cultiveraient				

In the previous example, the participant A should count the number of time the letter « a » appears in the paragraph. Then he should count the number of time the letter « d » appears. The participant A should then count the number of time the letter « n » appears. Finally, he should count the number of time the letter « t » appears in the paragraph.

There is a **maximum of 50 letters** to count. The letters may be different for the different paragraphs. The paragraphs are different for each participant A.

At the end of the 10 minutes, we pick up the answers sheets and give them out to the participants B in the other room.

Participant B's role

Each participant B receives the answers sheets of the participants A with whom he is paired. You will never be informed of the identity of the participant with whom you are paired.

The participant B's role is to count the number of correct answers in their participant A's answers sheets. In order to do it, we give the, the list of correct answers for the different problems. The participant B indicates « C» if the answers of their participant A are corrects and « I » if theyr are false, as indicated below.

Problème		Nombre de			
1	trichomie innovez calculaient commettiez tabla implantes tomettes accusateurs imprudente torpillions balancera balancier planifiions frissonnants cultiveraient	"a"	"d"	"n"	"t"
Vrai ou Faux					

When they finish, we call the participant B one by one. In another room, they report their detailed account to the person in charge of the payments and they receive their corresponding payoff. The participant A will receive their payoff corresponding to the report of their participant B when all the participants B will have been paid.

Earnings

The payoff of the participant A and the payoff of the participant B depend on the report of the participant B.

- If the participant B report a number between 0 and 14, the participant A receives 12 points and the participant B receives 13 points.
- If the participant B report a number between 15 and 21, the participant A receives 14 points and the participant B receives 18 points.
- If the participant B report a number between 22 and 28, the participant A receives 16 points and the participant B receives 23 points.
- If the participant B report a number between 29 and 35, the participant A receives 23 points and the participant B receives 14 points.
- If the participant B report a number between 36 and 50, the participant A receives 25 points and the participant B receives 19 points.

The following table reports the potential payoffs for the participant A and the participant B.

If you have any questions regarding these instructions, please raise your hand. We will answer you in private. Thank not to start before the experimentalist has given the start. Thank no to give any information on this session outside.

Table 8: Worker's and evaluator's payoffs

Reported number of correct answers	Worker's payoff	Evaluator's payoff
0		
1		
2		
3		
4		
5		
6		
7	12	13
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18	14	18
19		
20		
21		
22		
23		
24		
25	16	23
26		
27		
28		
29		
30		
31		
32	23	14
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43	25	19
44		
45		
46		
47		
48		
49		
50		