

Lies and Biased Evaluation: A Real-Effort Experiment

Julie Rosaz, Marie Claire Villeval

▶ **To cite this version:**

Julie Rosaz, Marie Claire Villeval. Lies and Biased Evaluation: A Real-Effort Experiment. Journal of Economic Behavior and Organization, 2012, 84 (2), pp. 537-549. halshs-00617120

HAL Id: halshs-00617120
https://shs.hal.science/halshs-00617120

Submitted on 26 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GROUPE D'ANALYSE ET DE THÉORIE ÉCONOMIQUE LYON - ST ÉTIENNE

W P 1124

Lies and Biased Evaluation: A Real-Effort Experiment

Julie Rosaz, Marie-Claire Villeval

Août 2011

Documents de travail | Working Papers

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France
Tel. +33 (0)4 72 86 60 60
Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France
Tel. +33 (0)4 77 42 19 60
Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr
Téléchargement / Download : http://www.gate.cnrs.fr – Publications / Working Papers

Lies and Biased Evaluation: A Real-Effort Experiment

Julie Rosaz[1] and Marie Claire Villeval[2 #]

Abstract: This paper presents the results of a laboratory experiment in which workers perform a real-effort task and supervisors report the workers' performance to the experimenter. The report is non verifiable and determines the earnings of both the supervisor and the worker. We find that not all the supervisors, but at least one third of them bias their report. Both selfish black lies (increasing the supervisor's earnings while decreasing the worker's payoff) and Pareto white lies (increasing the earnings of both) according to Erat and Gneezy (2009)'s terminology are frequent. In contrast, spiteful black lies (decreasing the earnings of both) and altruistic white lies (increasing the earnings of workers but decreasing those of the supervisor) are almost non-existent. The supervisors' second-order beliefs and their decision to lie are highly correlated, suggesting that guilt aversion plays a role.

JEL Classification: C91, D82, M52.

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

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.

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

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

Corresponding author

1. INTRODUCTION

This paper investigates the supervisors' bias in performance appraisals when the payoffs of both the supervisors and the agents depend on these appraisals and when the supervisors' reports are not verifiable. The personnel economics literature has shown that the evaluation of employees can be biased at the principal's advantage when information about the employees' effort is imperfect (Gibbs, 1991; Prendergast and Topel, 1993; Prendergast, 1999; Breuer *et al.,* 2010). This is particularly the case when the evaluation is subjective, which is more and more frequent in companies. Supervisors may bias evaluation upwards because they are reluctant giving bad news to avoid employees' discouragement and to preserve their morale or because they are lenient and fear being pressured to change the ratings (Gibbs, 1995; Lazear and Gibbs, 2009). They may also exaggerate the reported level of performance to preserve the reputation of a production unit. A biased appraisal can also distort the agents' performance downwards. This is the case when a supervisor has once and for all judged an employee as bad and merely seeks evidence in support of this judgment (the "Matthew effect", Gabris and Mitchell, 1989). Finally, supervisors may bias feedback both upwards and downwards to reduce the wage differentials between the employees (Murphy, 1991). These examples suggest that supervisors may not only have a choice between reporting the employee's performance truthfully or not, but they may also choose the direction of the distortion. Of course, these hypotheses are difficult to test with field data since lies are difficult to detect beyond anecdotal evidence. A laboratory experiment allows us to better identify such biases in evaluation.

While psychologists have been interested in deception for a long time (see Hyman, 1989, for a review; DePaulo *et al*., 1996; Tyler *et al*., 2006), economists have become interested in lying behavior more recently. Gneezy (2005) studies deception in a sender-receiver game.

Analyzing the role of monetary incentives in the decision to lie, he shows that a fraction of people care about their opponent's payoff and refrain from lying although a lie would increase their own payoff. Mohnen and Pokorny (2006) study the lying behavior in an employer-employee relationship. In their game, agents choose an effort level without knowing their ability level while principals learn the actual ability of the agents and choose whether to send a honest feedback if any. It is found that principals are less likely to lie on the agent's ability when the agent has a low ability. Erat and Gneezy (2009) propose a taxonomy of lies. "Selfish black lies" increase the individual'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 at a cost for 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 harm or benefit it causes to others relative to oneself.

Our paper differs from the previous literature on several points. First, in our experiment the workers have to perform a real-effort task (counting the occurrence of letters in paragraphs). The supervisors have to report to the experimenter their worker's performance. This report determines the payoffs of both the worker and the supervisor. Lying on a real effort may be perceived differently than lying on a state of nature as the worker has beard a cost of effort. The second novelty is that the structure of payoffs has been chosen such that the supervisors may have the opportunity to choose between different types of lies by increasing or decreasing their worker's reported performance. In contrast with Erat and Gneezy (2009) who test each type of lie separately, we can observe choices between different types of lies. Indeed, for some workers' performance levels, a supervisor may be able to report the truth, to tell a Pareto white lie or to tell a selfish black lie. In a work environment, it is not clear that individuals have the same willingness to lie when lying implies increasing or decreasing the performance of another individual. Third, in our experiment there is no cheap talk and it is the

second player who has an opportunity to cheat on the first player's real effort, which better represents supervisory appraisals. In contrast, most of the existing two-player experiments on deceptive behavior use cheap talk such that the first mover sends either 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 that is expected to influence the second player's decision (Charness and Dufwenberg, 2006; Charness and Dufwenberg, 2008; Ellingsen and Johanneson, 2004; Vanberg, 2008). Another approach of deceptive behavior is based on individual decision-making where individuals can only cheat on the experimenter (Fischbacher and Heusi, 2008; Mazar *et al*., 2008b; and most of the experimental literature on tax evasion – see also Pruckner and Sausgruber (2009) for a field experiment on newspaper purchasing on the streets). In our experiment, the supervisors can cheat on both the other player and the experimenter (for example by telling Pareto white lies that increase the payoffs of both players) but with (positive or negative) known consequences on the worker's payoff.

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. However, all the above-mentioned experiments on deceptive behavior have revealed that a significant proportion of individuals do not lie as much as they should to maximize their earnings. Guilt-aversion and lie-aversion are evoked to explain this behavior. Charness and Dufwenberg (2006 and 2008) have shown the importance of guilt aversion through the correlation between the decision to tell the truth and the players' beliefs about others' beliefs regarding their action. In contrast, lie-aversion assumes that the likelihood of telling the truth is uncorrelated to second order beliefs. This should also hold if truth-telling is motivated by the maintenance of self-image (Mazar *et al*., 2008a, 2008b). In our experiment, we manipulate the saliency of the second order beliefs to study how this affects the supervisors' decision whether or not reporting truthfully their worker's performance. At the end of our Baseline treatment, we

elicit the workers' beliefs about both the number of correct answers they expect their supervisor to report and their actual performance. In contrast, in our Second Order Belief treatment, we also elicit the supervisors' beliefs about their worker's answer to the question on the number of correct answers he expects his supervisor to report. Supervisors report their belief before they report their worker's performance to an assistant in the payment room. This is the only difference between the two treatments. The Second Order Belief treatment allows us to analyze to which extent the second order beliefs differ according to whether the supervisor is lying or not, for a given performance level.

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 (77 out of 221). If one excludes the cases in which players maximize their payoffs by reporting the truth, this percentage is 56.62% (77 out of 136). A majority of supervisors (64.71%) report a performance that maximizes their payoff. Selfish black lies represent 37.66% of the lies while the Pareto white lies represent 53.25% of the lies. Spite black lies and altruistic white lies are rare (3.9% and 6.49%, respectively). As regards beliefs, the workers overestimate their performance but they correctly predict the proportion of supervisors' lies except in the medium performance category. The saliency of second order beliefs does not affect the reporting behavior. The supervisors anticipate a higher proportion of lies predicted by their workers than what is actually the case. This may be due to either self-justification or prediction errors. Finally, the decision to lie is less likely when the supervisors believe that their worker expects them 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 2 presents the literature on the behavioral determinants of honesty. Section 3 describes the experimental design and presents

the predictions. Section 4 presents the results of the experiment. Section 5 concludes.

2. RELATED LITERATURE ON THE DETERMINANTS OF HONESTY

Contrary to the predictions of economics-of-crime models based on standard 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 behavior (see Sip *et al*., 2008). For example, in Fischbacher and Heusi (2008) participants have to roll a dice and they receive the payoff corresponding to the amount they report to have rolled, with no monitoring by the experimenter. While cheating is riskless, 39% of the participants are honest, 22% of the participants lie completely and, even more interestingly, a large minority of individuals lies but incompletely. In Mazar *et al*. (2008b), individuals 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 individuals report their performance. The number of correct answers reported is only 10% higher in the condition treatment than in the control treatment. Erat and Gneezy (2009) show that while many people are reluctant to tell even a Pareto white lie (39% of the participants), a significant proportion of individuals (30%) are willing to tell altruistic lies at their own expense, questioning the interactions between social preferences and lie aversion. While deceptive behavior has been shown to depend on the size of lies and the strength of promises (Lundquist *et al*., 2009) or on the mode of appointment of deciders (see Corazzini *et al*., 2009 about lying in electoral campaigns), 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.[3] Indeed, 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 to the player regardless of its impact. Therefore, it predicts that the likelihood of telling the truth is uncorrelated with the player's beliefs about 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 *et al*., 2009).

However, other recent experiments have minimized 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 *et al*. (2008b) suggest that people do not lie as much as they could because an internal reward system exerts control over their behavior. 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. Fischbacher and Heusi (2008) also interpret the incomplete lying behavior they observe in their experiment as due to the desire to maintain a favorable concept of the self.

3. EXPERIMENTAL DESIGN

3.1. The game

[3] See Battigalli and Dufwenberg (2007) for a theory of guilt-aversion.

Our real-effort experiment consists of two different treatments, the Baseline and the Second Order Beliefs treatment. It is based on a between-subjects design. The Baseline treatment involves workers ("participants A") and supervisors ("participants B"), in equal proportions. The workers and the supervisors are located in two different rooms and they know that they will never meet during or after the experiment. Each worker is matched randomly with an anonymous supervisor. Each participant receives the instructions of the game for both roles. 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 workers have to perform a task for 10 minutes. They receive a form 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. 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 verbally by their supervisor in the payment room. The correspondence between the reported performance and the workers' and the supervisors' payoffs is also made common information.

Once the 10 minutes have elapsed, the workers have to answer the following question: «How many correct answers do you think the participant B who is matched with you will report in the payment room?». This question is incentivized: a subject earns one more Euro if his prediction is correct (plus or minus one). Once the answers to the previous question have been collected, the workers have to answer a second question: « How many correct answers do you think you have provided when counting the letters?». This question is not incentivized because checking during the session the correspondence between the actual number and the

reported number of correct answers seemed inappropriate. Assuming that players have brought the same care to both answers, the comparison between these answers tells us whether or not the workers expect to be deceived, and if so, which type of lie they anticipate.

After collecting the exercise forms in the workers' room, an assistant distributes them to the supervisors in the other room. Each pair is only identified by a number that is indicated on the forms. Meanwhile, the supervisors are given the list of correct answers for each paragraph in order to minimize involuntary correction errors and to save time. They are asked to count the number of correct answers in their worker's form. They are told that they will have to verbally report their worker's performance in the payment room to an assistant. The supervisors' instructions and the experimentalist mention the basket in which they will have to leave all the documents before proceeding to the payment room. This strengthens the notion that there will be no control of verbal reports. Once all the supervisors have finished, they are called one by one to proceed to the payment room, so that they cannot communicate with any other participant. In the payment room, they verbally report their pair identification number and their worker's performance. Then, they get paid according to their verbal report. 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 report.

The Second Order Belief treatment is identical to the Baseline treatment except that once all the supervisors have checked their workers' answers, we elicit the supervisors' second order beliefs about their workers' expectations on their verbal report. Precisely, they have to answer the following question: *«What do you think your worker answered to the following question: 'How many correct answers do you think the participant B who is matched with you will report in the payment room?'?»*. 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 the propensity to lie and the nature of lies are modified when the supervisors are invited to think about their worker's expectation. If behavior is different than in the Baseline treatment, this will suggest that they have updated their self-concept. A correlation between second order beliefs and behavior will suggest the presence of guilt-aversion; the absence of correlation will suggest that lie-aversion or self-concept maintenance is determinant.

In both treatments, the structure of payoffs is made common information. There are five payoff levels for the workers and the supervisors, corresponding to five categories of reported performance. Let us call the various performance categories "very low", "low", "medium", "high" and "very high".[4] Table 1 displays the payoffs for each category of performance.

[Table 1 about here]

Reaching an immediately superior category of performance increases the workers' payoff by two points and the supervisor's payoff by five points. 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. We acknowledge that this structure of payoffs is arbitrary but it allows us to observe all types of lies we are interested in, as explained in the next sub-section.

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 these preferences are correlated with lying behavior. The Holt and Laury (2002) 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 (the safer lottery) and option B.[5] To approach

[4] The instructions mention the number intervals but do not refer to low, medium or high performance categories.

[5] The payoffs for option A 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, it increases to

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 only. Each decision consists of a choice between two payoff vectors between player X and player Y (option A and option B). Option A always pays an equal payoff to the two players (5 points each) while option B represents an unequal division of 20 points.[6] In option B, player X's payoff decreases while player Y's payoff increases as the number of the decision increases. 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. An inequality-neutral subject should choose option B for the first 16 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. Thus, choosing option A between decisions 12 and 16 indicates some 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

two tenths, and so on. The high-payoff probability in each decision increases as the number of the decision increases, up to the tenth decision where payoffs are certain. Risk neutrality corresponds to a cross at the fifth decision, while risk loving participants are expected to switch from option A to option B earlier and risk averse participants at the sixth decision or after.

[6] Although inspired by them, this test differs from Blanco, Engelmann and Normann (2010) in that in their game the first option pays always £20 to the dictator and £0 to the recipient while the second option offers equal outcomes ranging from (£0,£0), (£1,£1) up to (£20,£20).

indications about these attitudes.

At the end of the session in the payment room, random draws determine which of the two tests gives rise to actual payment and which decision is paid.

3.2. Predictions

Table 2 summarizes the various possible types of lies depending on the comparison between the true and the reported categories of correct answers, based on the Erat and Gneezy (2009)'s terminology.

[Table 2 about here]

In Table 2, the light grey cells along the diagonal correspond to truth-telling and the dark grey cells to the standard equilibrium of the game. Indeed, considering the payoffs displayed in Table 1, a supervisor who aims at maximizing his own monetary payoff should always report a medium performance (a number of correct answers comprised in between 22 and 28) regardless of the worker's actual performance. This prediction holds for both treatments.

Reporting a medium performance may reflect the truth but it may also result from two types of lies, the selfish black lies and the Pareto white lies. A supervisor tells a selfish back lie that increases his own payoff but reduces his worker's payoff if he reports a medium performance whereas the actual performance is either high (between 29 and 35) or very high (above 35). He tells a Pareto white lie that improves both players' payoffs if he reports a medium performance whereas the actual performance is either very low (below 15) or low (between 15 and 21).

From a behavioral point of view, however, behavior may deviate from these predictions. First, if they are lie-averse or guilt-averse or if they want to maintain their self-concept, supervisors may speak the truth although they could earn more by lying. If they are guilt-averse, they may

lie only if the lie improves the worker's payoff.[7] Second, altruistic supervisors may tell altruistic white lies by reporting a number of correct answers that increases the workers' payoffs but reduces their own payoff. This is the case if they report a high or a very high performance whereas the actual performance is medium. Finally, although unlikely in the context of this one-shot experiment, some supervisors may tell 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 a very low performance when they observe at least a medium performance, or if they report a high performance when they observe a very high one.

Behavior should be similar in the Second Order Belief treatment except if the treatment manipulation strengthens the correlation between second order beliefs and actions due to a higher salience of others' beliefs. In this case, a stronger guilt aversion (if any) may lead more supervisors to speak the truth. If lie-aversion or self-concept maintenance is the main determinant of honesty, we should observe that there is no correlation between second order beliefs and actions.

3.3. Procedures

23 sessions have been 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 non-adjacent rooms were used, one for the workers and the other one for the supervisors. Upon arrival, the participants randomly drew a tag from a bag assigning them to one room and to a

[7] We assume that the players are indifferent to the harm a lie may possibly cause to the experimentalist.

seat in this room. A unique number identified each pair. This procedure ensured that both the role assignment and the pairing of subjects were random.

All the instructions (see Appendices 1 and 2) were distributed and read aloud in each room. A neutral wording was used. Workers were named "participants A" and supervisors " participants B". In the worker's room, after reading the instructions and answering to questions in private, an experimenter distributed the paragraph sheets. Then, the workers 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 and next, another sheet asking for the workers' beliefs about their actual performance. Next, the paragraph forms were brought to the supervisors' room. Meanwhile, the workers completed a demographic questionnaire and were instructed to remain silent until the supervisors were paid and had left the institute.

Upon entering their room the supervisors completed the same demographic questionnaire. Attitudes toward risk and inequality were elicited. Next, the instructions for the game were read aloud. Once all questions were answered privately, each participant received the paragraph form filled out by his paired worker for evaluation. In the Second Order Belief treatment, once all the supervisors checked the forms, their belief was elicited. Finally, the supervisors were called upon one by one and put all their documents together (instructions, paragraph sheets) in a basket before proceeding to the payment room.

In the payment room, each supervisor reported his worker's number of correct answers to an assistant. Then, each participant tossed a coin to determine which of the test of risk aversion or the test of inequality aversion would be paid.[8] Payment was made accordingly in cash,

[8] If the risk elicitation task was randomly drawn, the participant rolled a ten-sided dice to determine which decision would give rise to payment and played the corresponding lottery. If the inequity aversion elicitation task was drawn, he 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

including the payment for accurate prediction. Once all the supervisors had left, the workers were called one by one to get paid. They also received a €2 show-up fee and the payment for accurate prediction. A session lasted 50 minutes on average. Workers earned €8 and supervisors earned €9.5 on average.

4. RESULTS

First, we examine the supervisors' lying decision, before characterizing the nature of lies. Then, we analyze 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. The data analysis is conducted at the level of the performance category. We do not consider as lies the differences between actual and reported number of correct answers within the same performance category as they do not affect payoffs.[9]

Result 1. A significant proportion of supervisors are lying. Not all supervisors are lying.

Support for result 1. Table 3 displays the distribution of the workers' actual performance and the corresponding supervisors' reports in each treatment.

[Table 3 about here]

A Kolmogorov-Smirnov test rejects in each treatment the equality of distribution of the actual performance categories and the reported performance categories ($p<.001$). All the Kolmogorov-Smirnov tests are two-tailed and exact. The maximization of one's own payoff should lead all supervisors to report a medium performance. We observe in Table 3 that in the Baseline treatment, 63.39% of the supervisors play the equilibrium by reporting a medium

decision of his paired player in the same room was implemented.

[9] 37 out of 221 (16.74%) participants declare a number of correct answers that differs from the actual number of correct answers without affecting the player's payoffs. 17 out of 37 (45.95%) decrease the actual number of correct answers while 20 out of 37 (54.05%) increase it. This distribution suggests that this is due to mistakes.

performance, while 38.39% of the workers actually perform in this category. In the Second Order Belief treatment, they are 66.06% and 44.04%, respectively. We observe no statistically significant difference between the two treatments.[10] This is clear evidence that supervisors lie, but not completely.

In total, in the Baseline treatment, 32.14% (36 out of 112) of the supervisors lie. In the Second Order Belief treatment, the percentage is 37.61% (41 out of 109). A test of proportion indicates that the difference between the two treatments is not significant (p=.393). If one excludes the cases in which the worker's performance actually belongs to the medium category, the percentage of liars is 49.28% (34 out of 69) in the Baseline and 60.66% (37 out of 61) in the Second Order Belief treatment. This difference is not significant either (p=.193).

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 to maximize their payoffs. This finding is consistent with previous studies showing that people do not fully exploit their lying opportunities (Fischbacher and Heusi, 2008; Mazar *et al*., 2008a and 2008b).

Result 2. Pareto white lies and selfish black lies are the most frequent types of lies.

Support for result 2. A novelty 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 non medium actual performance[11], by distinguishing between the various types of lies.

[Table 4 about here]

In total, if one includes the six lies in case of an actual medium performance, the Pareto white

[10] A Kolmogorov-Smirnov test indicates that there is no significant difference in the distribution of workers' actual performance catégories between the two treatments (p=.794). The same test concludes that the distribution of the supervisors' reported performance catégories does not differ between treatments (p=.848).

[11] The distribution of decisions for an actual performance between 22-28 shows that most of the supervisors report the true category. As indicated by Table 3, only 6 out of 91 (6.59%) supervisors lie. Two supervisors tell spite black lies, whereas the four others tell altruistic white lies.

lies represent 53.25% of the lies (41 out of 77) and selfish black lies represent 37.66% of the lies (29 out of 77). Not surprisingly, as their rationality is unclear, spite black lies are virtually non-existent (3 out of 77 lies in total, 3.90%). Altruistic white lies are not more frequent (4 observations, 5.19%). The data from both treatments are pooled together since there is no difference between treatments (proportion tests, p=.835 for the selfish black lies and p=.938 for the Pareto white lies). The saliency of second order beliefs does not affect behavior.

Let us consider an interesting case in which supervisors can choose between various lies. When the actual performance is high (between 29 and 35 correct answers), 40% (22 out of 55) of the supervisors report the truth; 44% declare a medium performance in order to increase their payoff at the expense of their worker (selfish black lie) and 16% increase both players' payoffs by reporting a very high performance (Pareto white lie).

Result 3. Workers anticipate biased appraisals.

Support for result 3. Figures 1-1 and 1-2 display the distribution of the workers' actual performance (a), their beliefs about their performance (b), their beliefs about their supervisor's report (c), the supervisors' second order beliefs (d) and their reports (e), for each treatment separately. Table 5 indicates the distribution of the workers' beliefs about their supervisor's report for each category of workers' belief on performance. Data from the two treatments are pooled together.

[Table 5 about here]
[Figures 1-1 and 1-2 about here]

Panels *a* and *b* in Figures 1-1 and 1-2 show that workers tend to overestimate their performance. A Kolmogorov-Smirnov test indicates that the distributions of actual and expected performance levels are significantly different in both treatments (p<.001).[12] The

[12] We acknowledge that the predicted performance may be all the more imprecise that this answer to this question was not incentivized.

comparison between their beliefs on performance and their beliefs about the supervisor's report shows that the workers expect some lies from their supervisors (see panels *b* and *c* in Figures 1-1 and 1-2). Indeed, the distributions of these two sets of beliefs are statistically different in both treatments (Kolmogorov-Smirnov tests, $p<.001$). If we pool the data from the two treatments, 77.45% of workers believe their performance is high or very high but only 53.92% of the workers expect that the supervisors will report these categories. 16.67% believe that their performance is medium but 37.25% of the workers predict that the supervisors will report a medium performance. More precisely, Table 5 indicates that 28.56% of the workers who believe their performance is very low expect that their supervisors will tell a lie; the corresponding values are 40%, 26.47%, 56.25% and 26.92% for those whose belief on performance is low, medium, high and very high, respectively. Among those who believe their performance is high, 43.25% anticipate that the reported performance will be downgraded while 10% expect that it will be upgraded.

The comparison between Table 5 and Table 3 indicate that the workers who believe that their performance is very low or low underestimate the supervisors' rate of lies (for these categories, the actual rates are 55.56% and 57.45%, respectively). The workers who believe that their performance is medium overestimate the rate of lies (the actual rate is 6.60%). The predictions of those whose belief on performance is high or very high only slightly underestimate the rate of lies (the actual rates are 60% and 31.58%). Overall, the proportion of actual lies and the proportion of lies predicted by the workers are not significantly different ($p=.556$). Similar proportion tests for each category of performance conclude that the difference is significant for the medium performance ($p=.002$), but not for the other categories ($p=.280$ for the very low category, $p=.455$ for the low category, $p=.665$ for the high category and $p=.685$ for the very high category). In summary, workers expect biased appraisals and make rather accurate predictions in each performance category, except when they believe they

perform at the medium level. Overall, they underestimate the actual proportion of medium performance reported by the supervisors (37.10% instead of 64.71%; see panels *c* and *e* in Figures 1-1 and 1-2).

Result 4. The supervisors' decision to bias their appraisal is positively related to their second order beliefs but supervisors lie less than what they predict from their workers' expectations.

Support for result 4. In the Second Order Belief treatment, we can relate the second order beliefs to the actual reports. 85.37% (35 out of 41) of the supervisors who lie believe that workers likewise expect them to report a number that differs from the actual performance category. This is the case of only 30.88% (21 out of 68) of the supervisors who tell the truth. If one excludes the cases in which the actual performance is medium, 89.19% (33 out of 37) of the supervisors who lie but only 45.83% (11 out of 24) of those who tell the truth expect that their worker anticipate a biased report. Proportion tests indicate that these differences are both significant ($p<.001$).

The supervisors do not predict accurately the workers' beliefs about their report (see panels *c* and *d* in Figure 1-2).[13] A Kolmogorov-Smirnov test rejects the equality of the distributions of the workers's beliefs about reports and the supervisors' second order beliefs ($p<.001$). The proportion of supervisors who believe that their worker predicts a lie is higher than the actual proportion of lies predicted by the workers (proportion test, $p=.015$) and the actual proportion of lies is smaller than the proportion of supervisors who believe that their workers predict a lie ($p=.021$). Thus, supervisors lie less than what they predict from their worker's expectations, possibly to maintain a positive self-image and to avoid guilt. Another explanation is that supervisors try to justify their behavior by reporting higher workers' expectations. However,

[13] It should be noted that they do not base their beliefs on the same performance appraisals. Workers base their prediction on their belief about their performance whereas the supervisors can observe the actual performance. A lie predicted by the worker corresponds to a case when the worker's belief about his performance differs from his belief about the supervisor's report. The supervisor's expectation that the worker predicts a lie is the situation where the supervisor's second order belief differs from the actual performance.

the supervisors' predictions were incentivized, so although it cannot be ruled out self-justification cannot be the sole explanation for overestimating predictions.

To provide a stronger support to result 4, we have estimated two Probit models to explain the probability to lie. In model (1) the data from both treatments are pooled together, while model (2) only includes the data from the Second Order Belief treatment. In both models, the independent variables include a dummy variable that indicates whether the actual performance belongs to another performance category than the medium one (since if performance belongs to this category supervisors maximize their earnings by reporting the truth). They also include several individual characteristics such as gender, having already participated in an experiment during the last twelve months, and being risk seeking (although lying is riskless in our environment). The risk seeking variable is equal to 1 if the participant has chosen option A (the safer lotery) less than 5 times in the Holt and Laury (2002) task, and 0 otherwise. Three dummy variables capture inequality aversion: advantageous inequality averse, disadvantageous inequality averse and altruistic participants. A participant is averse to advantageous inequality if he chose the equal payoff in the modified dictator game between decisions 1 and 10 at least once. A participant is averse to disadvantageous inequality if he chose the equal payoff between decisions 12 and 16 at least once. Finally, a participant is altruistic if he chose option B at least once between decisions 17 to 21 when choosing this option means earning less than the other player and less than with the alternative option but allowing the other player to earn more. Finally, the independent variables in model (1) also include a dummy variable for the Second Order Belief treatment and in model (2) they include a variable indicating whether the supervisor believes his worker expects him to lie. Table 6 displays the marginal effects of these variables.

[Table 6 about here]

Model (1) confirms that the Second Order Belief treatment has no significant impact on the decision to lie. Unsurprisingly, supervisors are more likely to lie when the performance category is not medium. Having participated in previous experiments increases the likelihood of lying. Finally, we find that the participants who are averse to advantageous inequality lie less. Gender has no significant impact. Model (2) shows that believing that the worker expects a lie increases the likelihood of a lie by 42.67%. This result confirms the findings of the previous subsection: the supervisors who believe that their worker expects the truth are less likely to lie. This finding is in line with Charness and Dufwenberg (2008) and supports the theory of guilt-aversion, without rejecting the concerns of some individuals for maintaining self-image or lie-aversion. We also find that altruistic subjects are marginally less likely to lie.

5. CONCLUSION

In this experimental paper, workers perform a real-effort task and supervisors have to report the workers' performance. We study the honesty of appraisals when the payoffs of both the supervisor and the agent depend on these appraisals. In our Second Order Belief treatment, we elicit the supervisors' second order beliefs about their worker's expectation on their report before the supervisors report their appraisal. This experiment is novel in that we use a real effort task and supervisors may have to choose between different types of lies depending on the workers' actual performance.

We find that a significant proportion of supervisors do not report the truth but supervisors do not lie as much as they could to maximize their payoffs. This is consistent with previous studies showing that individuals do not fully exploit their lying opportunities (Fischbacher and Heusi, 2008; Mazar *et al*., 2008a and 2008b). Using Erat and Gneezy (2009)'s terminology, we also find that Pareto white lies and selfish black lies are the most frequent types of lies, while spite black lies and altruistic white lies are almost non-existent. Regarding

expectations, workers anticipate biased appraisals and correctly predict the proportion of lies except when they believe they perform at the medium level. We also find a strong correlation between the supervisors' second-order beliefs and their decision to lie or not to lie, suggesting that guilt aversion plays an important role. Therefore, our results support Charness and Dufwenberg (2008)'s analysis of the impact of guilt aversion on lying behavior. We cannot, however, reject lie-aversion or a concern for self-concept maintenance. Indeed, some individuals do not lie although by doing so they could increase both players' payoffs.

We need to be cautious about the external validity of our results, especially since our payoff matrix is ad hoc, lying is riskless and interactions are one-shot. But finding that many supervisors bias their performance appraisal to increase their payoff while workers have to perform a real effort task suggests that this may also happen in some natural settings, especially when appraisals are based on subjective evaluation. Many extensions of this paper could be thought of. First, a repeated game would allow us to test how long-term relationships affect the likelihood of biased appraisals. Indeed, in a repeated interaction, repeated biased appraisals would probably lead workers to reduce their effort in further periods and supervisors to make different choices to preserve the expectation of future interactions (see Toma *et al*., 2008). Second, while our results show that second order beliefs are correlated with lying behavior, another extension would be studying the effect of showing the workers' beliefs to the supervisors before they report their evaluation. We could then observe the impact of empathy on the decision to lie. Finally, we could compare the efficiency of various measures aiming at reducing the likelihood of biased appraisals, such as face-to-face appraisals or pre-commitments such as promises or oaths.

References

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

Blanco, M., D. Engelmann, and H.T. Normann. 2010. "A Within-Subject Analysis of Other-Regarding Preferences." Discussion Paper No.6, Düsseldorf Institute for Competition Economics.

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

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

Corazzini, L., S. Kube, M.A. Maréchal, A. Nicolo. 2009. Elections and Deceptions: Theory and Experimental Evidence. Working Paper No. 421, University of Zurich.

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. Mitchell. 1989. "The impact of merit raise scores on employee attitudes; the Mattew effect of performance appraisal." *Public Personnel Management* 17 (4): 369-387.

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

Gibbs, M.J. 1995. "Incentive Compensation in a Corporate Hierarchy." *Journal of Accounting and Economics / Journal of Labor Economics* joint issue 19 (2-3): 247-277.

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*, 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." *Annual Review of Psychology* 40: 133–154.

Lazear, E.P., and M. Gibbs, 2009. *Personnel Economics in Practice*, John Wiley & Sons.

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

Mazar, M., O. Amir, and D. Ariely. 2008a. "More Ways to Cheat -Expanding the Scope of Dishonesty." *Journal of Marketing Research*, 45(6): 651-653.

Mazar, N., O. Amir, and D. Ariely. 2008b. "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.

Murphy, K.J. 1991. "Performance measurement and appraisals," in W. Bruns (Ed.), *Performance measurement, evaluation and incentives*. Cambridge: Harvard Business School.

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. 2009. "Trust on the Streets: A Natural Field Experiment on Newspaper Purchasing." Discussion Papers No. 2009-24, Johannes Kepler University Linz.

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.

Toma, C.L., J.T. Hancock, and N.B. Ellison. 2008. Separating facts from fiction: an examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin* 34 (8): 1023-1036.

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.

Tables

Table 1. Workers' and supervisors' payoffs by performance category, in points

Reported number of correct answers		Worker's payoff	Supervisor's payoff
0-14	Very low	12	13
15-21	Low	14	18
22-28	Medium	16	23
29-35	High	23	14
36-50	Very high	25	19

Table 2. Distribution of the various types of lies

		Reported number of correct answers by the supervisor				
		0-14	15-21	22-28	29-35	36-50
Actual 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
Worker's payoff		12	14	16	23	25
Supervisor's payoff		13	18	23	14	19

Note: The dark grey cells represent the equilibrium and the light grey cells the truthful reports.

Table 3. Distribution of actual and sreported performance categories

		Reported number of correct answers by the supervisor					Total
		0-14	15-21	22-28	29-35	36-50	
Actual number of correct answers	0-14	**4 (44.44)** 4 0	**1 (11.12)** 0 1	**4 (44.44)** 1 3			**9 (100)** 5 4
	15-21		**20 (42.55)** 13 7	**25 (53.19)** 15 10		**2 (4.26)** 0 2	**47 (100)** 28 19
	22-28		**2 (2.20)** 1 1	**85 (93.40)** 41 44	**1 (1.10)** 1 0	**3 (3.30)** 0 3	**91 (100)** 43 48
	29-35			**24 (43.64)** 13 11	**22 (40)** 11 11	**9 (16.36)** 3 6	**55 (100)** 27 28
	36-50			**5 (26.32)** 1 4	**1 (5.26)** 1 0	**13 (68.42)** 7 6	**19 (100)** 9 10
Total		**4 (1.81)** 4 0	**23 (10.41)** 14 9	**143 (64.71)** 71 72	**24 (10.85)** 13 11	**27 (12.22)** 10 17	**221 (100)** 112 109

Note: In each cell, the first row in bold indicates the aggregate numbers from both treatments (and corresponding percentages in parentheses), the second row indicates the values for the Baseline treatment and the third row the values for the Second Order Belief treatment. The dark grey cells represent the equilibrium and the light grey cells the truthful reports.

Table 4. Distribution of decisions by treatment

Baseline treatment 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 69 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 5. Workers' beliefs on the supervisors' reports

Workers' beliefs on their performance	Workers' beliefs on supervisors' reports					
	0-14	15-21	22-28	29-35	36-50	Total
0-14	5 (71.44)	0	1 (14.28)	0	1 (14.28)	7 (100)
15-21	0	3 (60)	2 (40)	0	0	5 (100)
22-28	0	4 (11.76)	25 (73.53)	3 (8.82)	2 (5.88)	34 (100)
29-35	3 (3.75)	3 (3.75)	31 (38.75)	35 (43.75)	8 (10)	80 (100)
36-50	0	0	17 (21.79)	4 (5.13)	57 (73.08)	78 (100)

Note that in this table the total number of observations is 204 because in the first session of the Baseline treatment, we elicited the beliefs of the workers on their supervisor's report but not their beliefs on their actual performance.

Table 6. Determinants of lying behavior (Probit regression models)

Dependent variable: Decision to lie	All observations (1)		Second Order Belief Treatment (2)	
Second Order Belief treatment	.1121	(.0681)	-	-
Belief that the worker expects a lie	-	-	.427***	(.101)
Actual performance different from 22-28	.502***	(.051)	.428***	(.095)
Female	.015	(.068)	-.157	(.109)
Experienced	.203***	(.078)	.206	(.135)
Risk seeking	-.127	(.081)	.019	(.151)
Advantageous inequality averse	-.172**	(.067)	-.209*	(.113)
Disadvantageous inequality averse	-.029	(.074)	-.166	(.106)
Altruistic	-.028	(.071)	-.215*	(.113)
Observations	221		109	
LR Chi^2	84.28		65.34	
Prob > Chi^2	.0000		.0000	
Pseudo R^2	.2950		.4527	

Note: Marginal effects are reported. Standard errors in parentheses. *** indicates significance at the .01 level;,** at the .05 level, * at the .10 level.

Figures

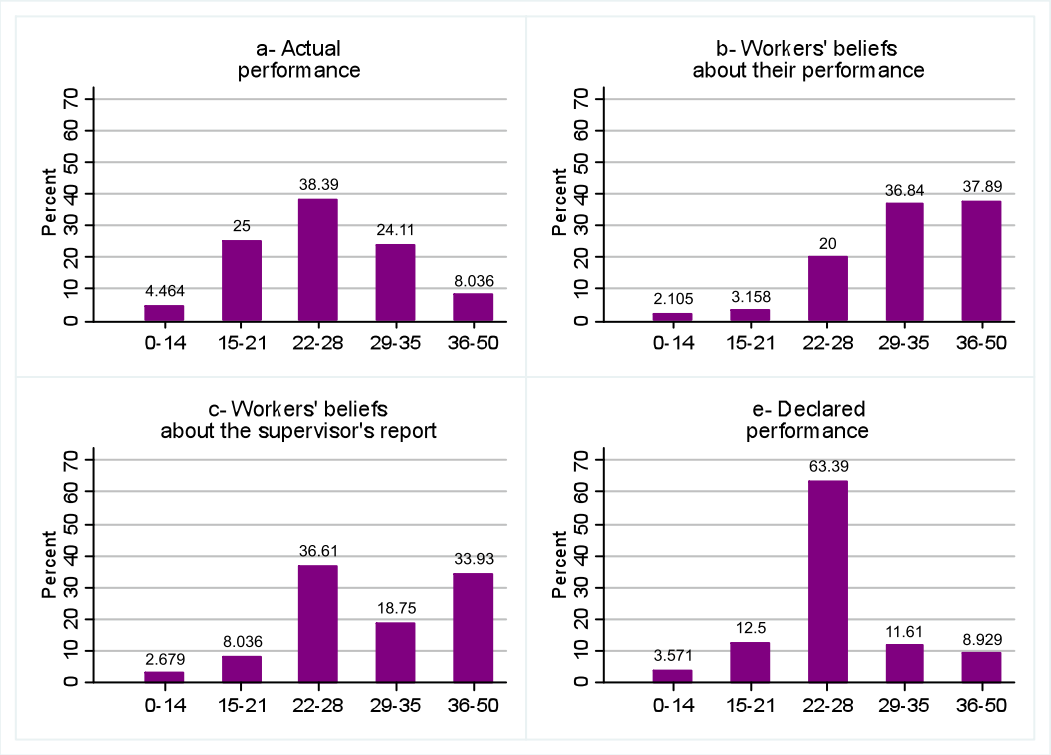


Fig.1-1. Distribution of performance categories and beliefs in the Baseline treatment

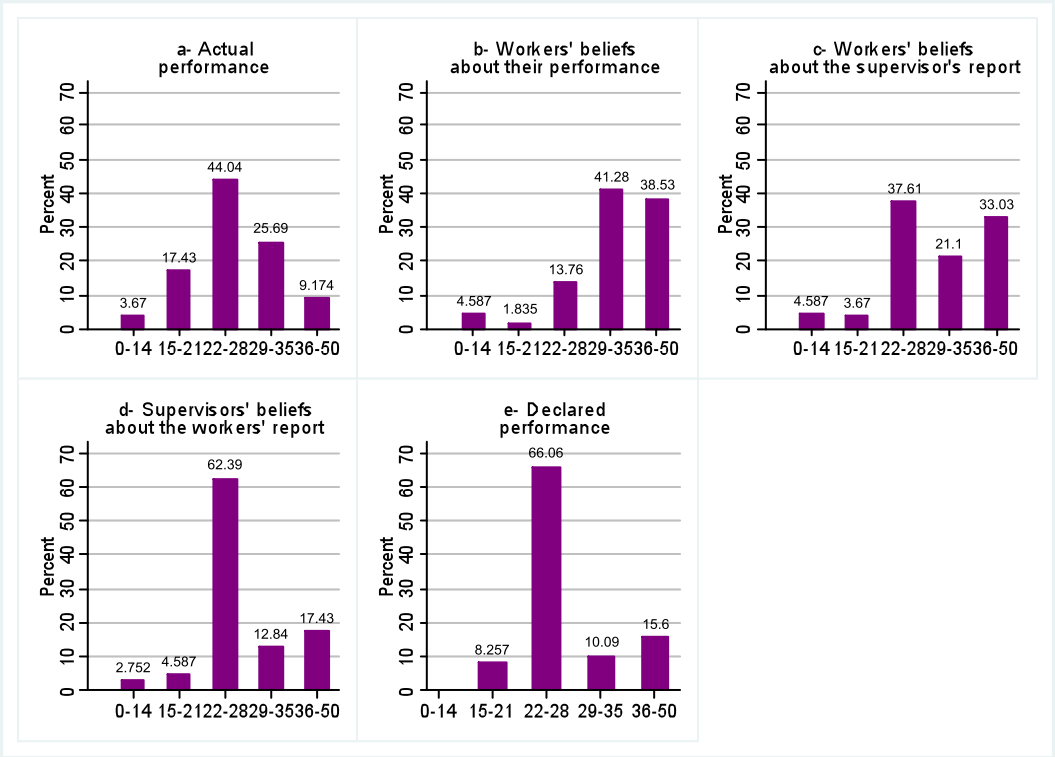


Fig.1-2. Distribution of performance categories and beliefs in the Second Order Belief treatment

Appendix 1. Instructions for the workers in the Baseline treatment (original in French)

We thank you for participating in this experiment. At the end of the session, you will be paid in cash in another room. It is forbidden to communicate with the other participants throughout the experiment.

Instructions for participants A and B

During this part, your earnings are expressed in points, with the following conversion rate:

3 points = €1

There are two types of participants in equal number: participants A and participants B. Your role has been randomly assigned to you at the beginning of the experiment by the color of the ticket you drew randomly in the bag upon arrival.

All the participants in this room are participants A. All the participants B are located 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 paragraphs. Each paragraph is composed of French words generated randomly and it has no meaning. Participants A are requested to count the occurrence of four letters in each paragraph. Each answer should be written in the box corresponding to the counted letter.

Here is an example of paragraph:

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

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

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

Once the 10 minutes have elapsed, we collect the forms and bring them to the participants B in the other room.

Participant B's role

Each participant B receives the form filled out by the participant A he is paired with. Nobody will ever be informed of the identity of the participant he is paired with.

Participants B are requested to count the number of correct answers in their participant A's form. In order to do it, we provide them with the list of correct answers to the different problems. Participants B indicate «V» if the answer of their participant A is correct and « F » if it is incorrect 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					

Once all participants B have finished, we call them one by one. In the payment room, they report their count verbally to the person in charge of payments and they receive the corresponding payoff. Once all the participants B have been paid, the participants A will receive the payoff corresponding to the report of their participant B.

Earnings

The earnings of the participants A and B depend on the number reported by the participant B.

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

The following table reports the potential payoffs of the participants A and the participants B.

(*See the Table at the end of the instruction set*).

If you have any questions regarding these instructions, please raise your hand. We will answer to your questions in private. Please do not to start before the experimentalist has given the start.

(*Instructions distributed after the 10 minutes have elapsed)*

Please answer to the following question. If your answer is correct (+ or – 1), you earn € 1 more.

«How many correct answers do you think the participant B who is matched with you will report in the payment room?».

_____ correct answers.

(*Instructions distributed after the answer to the previous question has been collected)*

Please answer to the following question.

«How many correct answers do you think you have provided when counting the letters?».

_____ correct answers.

Post-experimental questionnaire (*Note that this questionnaire was also to be completed by the participants B but at the very beginning of the session)*

You will receive an additional €2 for filling out this form that will be added to your earnings of the previous part.

1- What is your age?

2- What is your gender?

Male O
Female O

3- What is your education attainment?

High school or less O
Baccalaureat O
Baccalaureat +2 O
Baccalaureat +3 O
Baccalaureat +4 O
Baccalaureat +5 O
Beyond Baccalaureat +5 O

4- Are you?

Employee O
Unemployed O
Student O

5- If you are a student, where are you studying ?

Itech O
EM Lyon O
EC Lyon O
University (economics) O
Other O

6- What is your major?

Economics O
Engineering O
Business O

Political Sciences	O
Psychology	O
MASS	O
Mathematics	O
Computer sciences	O
Sociology	O
Biology	O
Chemistry	O
Medecine	O
Other	O

7- Did you already participate in an experiment ? Yes O
No O

8- If your answered yes at question 7, how many times have you participated since last September ?

At the end of the session, you will be called one by one in the payment room. It is still forbidden to talk to other participants.

Appendix 2. Instructions for the supervisors in the Baseline treatment (original in French)
(Note that the demographic questionnaire was distributed to participants B before they received these instructions)

We thank you for participating in this experiment.
This session consists of three independent parts. You have received the instructions for the first part. You will receive the instructions for the next part after completion of this part.

At the end of the session, one of the first two parts will be randomly selected for payment. Your earnings in the selected part and in the third part will be added up. You will be paid in cash in another room.
It is forbidden to communicate with the other participants throughout the experiment.

Part 1

The attached form displays ten decisions. Each decision is a paired choice between "Option A" and "Option B". You will make ten choices, but only one of them will be used in the end to determine your earnings if this part is selected for payment.

Here is a ten-sided dice that will be used to determine your payoff. The faces are numbered from 1 to 10 (the "0" face of the dice will serve as 10). In the payment room, you will throw this die twice :

- once to select one of the ten decisions to be used,
- and a second time to determine what your payoff is for the option you chose, A or B, for the particular decision selected.

Even though we ask you to make ten decisions, only one of these will end up affecting your earnings. However, you do not know in advance which decision will be used. Obviously, each decision has an equal chance of being used in the end.

- Look at Decision 1.

Option A pays 2 € if the throw of the dice is 1, and it pays 1.6 € if the throw is 2-10.
Option B yields 3.85 € if the throw of the dice is 1 and it pays 0.1 € if the throw is 2-10.

- Look at Decision 2.

Option A pays 2 € if the throw of the dice is 1 or 2, and it pays 1.6 € if the throw is 3-10.
Option B yields 3.85 € if the throw of the dice is 1 or 2 and it pays 0.1 € if the throw is 3-10.

The other decisions are similar, except that as you move down the table, the chances of a higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the dice will not be needed since each option pays the highest payoff for sure, so your choice here is between 2 € and 3.85 €.

If you have any question, please raise your hand. Your questions will be answered in private.

Please indicate for the 10 decisions if you choose option A or option B.

	Your decisions
Decision 1 Option A: 1 chance over 10 to be paid 2 € et 9 chances over10 to win 1.6 € Option B: 1 chance over 10 to be paid 3.85 € et 9 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 2 Option A: 2 chances over 10 to be paid 2 € et 8 chances over 10 to be paid 1.6 € Option B: 2 chances over 10 to be paid 3.85 € et 8 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 3 Option A: 3 chances over 10 to be paid 2 € et 7 chances over 10 to be paid 1.6 € Option B: 3 chances over 10 to be paid 3.85 € et 7 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 4 Option A: 4 chances over 10 to be paid 2 € et 6 chances over 10 to be paid 1.6 € Option B: 4 chances over 10 to be paid 3.85 € et 6 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 5 Option A: 5 chances over 10 to be paid 2 € et 5 chances over 10 to be paid 1.6 € Option B: 5 chances over 10 to be paid 3.85 € et 5 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 6 Option A: 6 chances over 10 to be paid 2 € et 4 chances over 10 to be paid 1.6 € Option B: 6 chances over 10 to be paid 3.85 € et 4 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 7 Option A: 7 chances over 10 to be paid 2 € et 3 chances over 10 to be paid 1.6 € Option B: 7 chances over 10 to be paid 3.85 € et 3 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 8 Option A: 8 chances over 10 to be paid 2 € et 2 chances over 10 to be paid 1.6 € Option B: 8 chances over 10 to be paid 3.85 € et 2 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 9 Option A: 9 chances over 10 to be paid 2 € et 1 chances over 10 to be paid 1.6 € Option B: 9 chances over 10 to be paid 3.85 € et 1 chances over 10 to be paid 0.1 €	Option A O Option B O
Decision 10 Option A: 10 chances over 10 to be paid 2 € et 0 chance over 10 to be paid 1.6 € Option B:10 chances over 10 to be paid 3.85 € et 0 chance over 10 to be paid 0.1 €	Option A O Option B O

Part 2 (*instructions distributed after completion of part 1)*

During this part, your earnings are expressed in points, with the following conversion rate:

5 points = €1.

The situation is the following one. There are two roles: participant X and participant Y. The participant X chooses one of two possible payment distributions between the participant Y and himself (option A or option B) in each of 21 decisions. The participant Y has no decision to make; he can only accept the participant X's decisions.

The matching of participants X and participants Y in this room has been randomly determined in advance. You will never be informed of the identity of the participant with whom you have been matched.

You will make 21 decisions in the role of a participant X. However, your role (X or Y) will be assigned at the end of the session in the payment room if this part is selected for payment.

If this part is selected for payment in the payment room:

- you will toss a coin to determine your role (X or Y). If the coin indicates tail, you will be a participant X. If the coin indicates head, you will be a participant Y;
- then, you will randomly draw one of 21 tags numbered between 1 to 21 to select one of your 21 decisions. If you are a participant X, you win the amount corresponding to the option you have chosen. If you are a participant Y, you win the amount corresponding to the option chosen by the participant X you are matched with.

Even though we ask you to make 21 decisions, only one of these will end up affecting your earnings. However, you do not know in advance which decision will be used. Obviously, each decision has an equal chance of being used in the end.

For example, look at Decision 1.

Decision 1 Option A: Your payment is 5 points and the participant Y's payment is 5 points. Option B: Your payment is 20 points and the participant Y's payment is 0 point.	Option A O Option B O

You make a decision in the role of a participant X:
If you choose option A, you and the participant Y earn 5 points each. If you choose option B, you receive 20 points and the participant Y receives 0 point.

The 20 other decisions are similar, except that as you move down the table, the earnings of the participant X decrease and the earnings of the participant Y increase.

If you have any question, please raise your hand. Your questions will be answered in private.

Please indicate for each of the 21 decisions if you choose option A or option B.

	Your decision
Decision 1 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 20 pts and the participant Y's payoff is 0 pt.	 Option A O Option B O
Decision 2 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 19 pts and the participant Y's payoff is 1 pt.	 Option A O Option B O
Decision 3 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 18 pts and the participant Y's payoff is 2 pts.	 Option A O Option B O
Decision 4 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 17 pts and the participant Y's payoff is 3 pts.	 Option A O Option B O
Decision 5 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 16 pts and the participant Y's payoff is 4 pts.	 Option A O Option B O
Decision 6 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 15 pts and the participant Y's payoff is 5 pts.	 Option A O Option B O
Decision 7 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 14 pts and the participant Y's payoff is 6 pts.	 Option A O Option B O
Decision 8 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 13 pts and the participant Y's payoff is 7 pts.	 Option A O Option B O
Decision 9 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 12 pts and the participant Y's payoff is 8 pts.	 Option A O Option B O
Decision 10 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 11 pts and the participant Y's payoff is 9 pts.	 Option A O Option B O
Decision 11 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 10 pts and the participant Y's payoff is 10 pts.	 Option A O Option B O
Decision 12 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 9 pts and the participant Y's payoff is 11 pts.	 Option A O Option B O
Decision 13 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 8 pts and the participant Y's payoff is 12 pts.	 Option A O Option B O
Decision 14 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 7 pts and the participant Y's payoff is 13 pts.	 Option A O Option B O
Decision 15 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 6 pts and the participant Y's payoff is 14 pts.	 Option A O Option B O
Decision 16 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 5 pts and the participant Y's payoff is 15 pts.	 Option A O Option B O
Decision 17 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 4 pts and the participant Y's payoff is 16 pts.	 Option A O Option B O
Decision 18 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 3 pts and the participant Y's payoff is 17 pts.	 Option A O Option B O
Decision 19 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 2 pts and the participant Y's payoff is 18 pts.	 Option A O Option B O
Decision 20 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 1 pt and the participant Y's payoff is 19 pts.	 Option A O Option B O
Decision 21 Option A: Your payoff is 5 pts and the participant Y's payoff is 5 pts. Option B: Your payoff is 0 pt and the participant Y's payoff is 20 pts.	 Option A O Option B O

Part 3 (instructions for participants A and B) (*instructions distributed after completion of part 2)*

During this part, your earnings are expressed in points, with the following conversion rate:

3 points = €1.

There are two types of participants in equal number: participants A and participants B. Your role has been assigned to you at the beginning of the experiment by the color of the ticket you drew randomly in the bag upon arrival.

All the participants in this room are participants B. All the participants A are located 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 of counting the occurrence of letters in paragraphs. 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 four letters. Each answer should be written in the box corresponding to the counted letter.

Here is an example of paragraph:

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

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

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

Once the 10 minutes have elapsed, we collect the forms and bring them to the participants B in the other room.

Participant B's role

Each participant B receives the form filled out by the participant A he is paired with. Nobody will ever be informed of the identity of the participant he is paired with.

Participants B are requested to count the number of correct answers in their participant A's form. In order to do it, we provide them with the list of correct answers to the different problems. Participants B indicate «V» if the answer of their participant A is correct and « F » if it is incorrect 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					

Once all participants B have finished, we call them one by one. In the payment room, they report their count verbally to the person in charge of payments and they receive the corresponding payoff. Once all the participants B have been paid, the participants A will receive the payoff corresponding to the report of their participant B.

Earnings

The earnings of the participants A and B depend on the number reported by the participant B.

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

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

Please put all the documents you have received in the basket before leaving the room for payment.

If you have any question regarding these instructions, please raise your hand. We will answer to your questions in private. Please do not to start before the experimentalist has given the start.

Table of payoffs of participants A and participants B

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