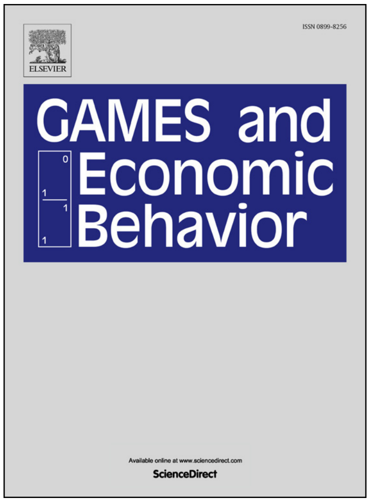
Accepted Manuscript



The Role of Verifiability and Privacy in the Strategic Provision of Performance Feedback: Theory and Experimental Evidence

Seda Ertac, Levent Koçkesen, Duygu Ozdemir

PII: S0899-8256(16)30088-4
DOI: http://dx.doi.org/10.1016/j.geb.2016.08.009
Reference: YGAME 2583

To appear in: *Games and Economic Behavior*

Received date: 9 October 2015

The Role of Verifiability and Privacy in the Strategic Provision of Performance Feedback: Theory and Experimental Evidence*

Seda Ertac† Koç University
Levent Koçkesen Koç University
Duygu Ozdemir University of Essex

May 2016

Abstract

We theoretically and experimentally analyze the role of verifiability and privacy in strategic performance feedback using a "one principal-two agent" context with real effort. We confirm the theoretical prediction that information transmission occurs only in verifiable feedback mechanisms and private-verifiable feedback is the most informative mechanism. Yet, subjects also exhibit some behavior that cannot be explained by our baseline model, such as telling the truth even when this will definitely hurt them, interpreting "no feedback" more optimistically than they should, and being influenced by feedback given to the other agent. We show that a model with individual-specific lying costs and naive agents can account for some, but not all, of these findings. We conclude that although agents do take into account the principal's strategic behavior to form beliefs in a Bayesian fashion, they are overly optimistic and interpret positive feedback to the other agent more pessimistically than they should.

Keywords: Lab experiments, Performance feedback, Strategic communication, Cheap talk, Disclosure, Persuasion, Multiple audiences, Lying

JEL Classification: C72, C92, D23, D82, D83, M12, M54.

---

*We thank David Gill, two anonymous referees, and the advisory editor for their comments and suggestions and Ayca Ebru Giritligil for making the BELIS laboratory available for our experiments. Ertac thanks the Scientific and Technological Research Council of Turkey (TUBITAK) for generous funding (Grant No. 111K444) for the experiments in this paper, and the Turkish Academy of the Sciences for general financial support.

†Corresponding author. Department of Economics, Koç University, Rumelifeneri Yolu Sariyer, Istanbul 34450, Turkey. E-mail: sertac@ku.edu.tr.

1 Introduction

This paper provides a theoretical and experimental study of the role of verifiability and privacy in the strategic communication of interim performance information. Performance feedback (also known as performance review or performance appraisal) is one of the most commonly used management practices. Almost every organization, be it a major corporation, a small company, a high school, or a hospital uses some form of performance feedback.[1] Although it is considered an indispensable part of any organization, performance feedback has also been the object of a heated debate. Employees usually dread it and many business experts and consultants are fierce opponents. One of the most critical voices, Samuel Culbert, states that "[i]t's a negative to corporate performance, an obstacle to straight-talk relationships, and a prime cause of low morale at work."(Culbert [2008]).

Ideally, performance feedback gives an unbiased report on past performance and provides guidance regarding how to improve future performance. This aspect, i.e., accuracy or unbiased communication, has been regarded as a crucial aspect of performance feedback. In practice, however, the accuracy of feedback may be tainted due to various biases that arise from the evaluator's self-interest. In particular, supervisors may be vague in their assessments or avoid giving negative feedback to their subordinates for strategic reasons.[2] Forced ranking systems may overcome this deficiency but they cause problems of their own, potentially undermining employee confidence and motivation.

Clearly, there are various pros and cons of performance feedback along a multitude of dimensions, but its effectiveness as a tool of communication seems to be one of the most contentious aspects. In this paper we focus on precisely this aspect. In a setting where feedback is given strategically by a supervisor, we theoretically and experimentally analyze how subordinates interpret the feedback they receive in forming an opinion of themselves and whether feedback communicates the actual performance information in a truthful manner.

In our experiment there is a supervisor (called principal) and two subordinates (called agents) who work for (potentially) two periods. In each period agents perform a real effort task and succeed if their performance is greater than a randomly determined threshold, which plays the role of chance or other unpredictable exogenous factors such as market conditions and organizational standards. The principal, and only the principal, observes the first-period performance (i.e., success or failure) of the agents and then decides whether and what type of feedback to provide to the agents. The agents observe the feedback (or lack thereof), update their beliefs about their likelihood of succeeding in the second period, and choose whether to perform the task again in the second period or not.[3]

The agents receive monetary payoff from their performances in the two periods, while the principal receives a payoff only from the agents' second-period performances. In addition, the principal's payoff depends on the minimum of the two agents' performances. That is, the principal ob-

[1] One source estimates that "97.2% of U.S. companies have performance appraisals, as do 91% of companies worldwide" (see "Should Performance Reviews Be Fired?"). Also see evidence cited in Murphy and Cleveland [1991].

[2] See Schraeder et al. [2007] for a summary of research in psychology, management, and organizational behavior. Culbert [2008] claims that "any critique [involved in performance review] is as much an expression of the evaluator's self-interests as it is a subordinate's attributes or imperfections." Longenecker et al. [1987] report (interview) evidence that the main concern of the executives in performance appraisals is not accuracy but rather to motivate and reward subordinates. Accordingly, they systematically inflate the ratings in order to increase performance. In the Forbes article titled "Ten Biggest Mistakes Bosses Make In Performance Reviews," the number 1 item is 'Too vague,' number 2 is 'Everything's perfect – until it's not and you're fired,' while number 8 is 'Not being truthful with employees about their performance' (Jackson [2012]).

[3] More precisely, subjects state the probability with which they believe they will succeed in the second-period task, which is elicited using a Becker-De Groot-Marschak type procedure.

tains an extra payoff only if both agents end up performing in the second-period task. This captures "weakest-link" type performance settings, where it is important that every agent achieve a certain level of performance. With such a payoff function, the principal prefers both agents to have a high perceived likelihood of success in the second-period task, i.e., to have high self-confidence. This also makes feedback a strategic choice: if the first-period performance is positively correlated with second-period performance, then the principal has an incentive to get the agents to believe that they succeeded in the first period task.[4]

We analyze the effectiveness of performance feedback mechanisms along two dimensions: (1) verifiability of the feedback; (2) privacy of the feedback. Our baseline scenario is truthful private feedback, in which each agent privately and truthfully learns whether he succeeded in the first period task or not. In the verifiable feedback case, the principal has to reveal the true performance or reveal no information at all, while in unverifiable feedback, she may lie about performance without incurring any monetary cost. The feedback may be private, in which case each agent receives feedback only about his own performance, or public, in which case both agents observe the feedback on each agent's performance. Therefore, in addition to the baseline scenario, we have four different treatments: (1) private-verifiable; (2) public-verifiable; (3) private-unverifiable; (4) public-unverifiable.

In reality, some performance measures are indeed objective and hence verifiable, while others are subjective and unverifiable. For example, a supervisor may have access to evaluations - by higher ranking administrators, co-workers, customers, or students - that can be reproduced if needed. Similarly, sales or productivity figures, customer ratings, exam grades of students, and long-term mortality rates after surgeries are all objectively measurable and verifiable performance measures. Subjective or judgmental evaluations by supervisors, on the other hand, are by their very nature unverifiable, i.e., cheap talk. Likewise, feedback is sometimes provided in a private manner, as in many performance review interviews, while in other cases it is public, as in 'employee of the month' types of feedback. The question of whether feedback should be provided publicly is especially relevant for contexts where it is important to preserve the "morale" of all agents. Given that most organizations have some freedom in determining their feedback mechanisms along the lines we consider, our results can have significant policy implications for firms and for educational settings.

In Section 4 we analyze a theoretical model and derive several predictions. Our main prediction is that information transmission occurs only in verifiable feedback mechanisms and private-verifiable feedback is the most informative feedback mechanism. Section 5.1 presents strong evidence in support of this prediction. We therefore conclude that, if effective communication is the main objective, organizations should try to provide measurable and verifiable forms of feedback and they would be better off if they do this privately.

We also find that positive and negative feedback have significant effects on beliefs in all treatments except private-unverifiable feedback, whereas giving no feedback has no significant effect on beliefs. Since "no feedback" must be interpreted as bad news, especially in verifiable feedback mechanisms, this finding contradicts the predictions of our model.

Our data provides evidence that when feedback is public, agents' beliefs about their likelihood of

[4] The experiment is designed so that the likelihood of success for each agent is independent of the likelihood of success for the other agent. This implies that the performance of the other agent is not informative about the likelihood of own success. Furthermore, feedback has no direct payoff consequences, which lets us isolate the communication phase involved in the feedback process from other strategic considerations.

success are influenced by the feedback provided to the other agent. More precisely, they become more optimistic if the other agent receives negative feedback and less optimistic if the other agent receives positive feedback. We further find that this effect is significant only when own feedback is positive, and stronger for public-unverifiable than for public-verifiable feedback. Since, in our experimental design, the other agent's performance has no informative content regarding own performance, these findings are also at odds with our model.

Finally, we find a positive effect of beliefs but no significant effect of feedback on performance. In other words, performance reviews are at most a weak instrument for boosting employee performance.

In Section 5.2, we analyze principals' behavior and find that, in all the treatments, some (but not all) subjects tell the truth. This goes against our prediction that in unverifiable feedback mechanisms, principals should always provide positive feedback. Furthermore, we find that principals expect positive feedback to be interpreted more optimistically and negative feedback more pessimistically than they actually are. In other words, some of them give bad news even though they actually believe that it will be interpreted as such, which leads us to conclude that lying imposes individual-specific costs.

In Section 6.1 we extend our baseline model to include individual-specific costs of lying and naive agents. We show that it can account for most of our empirical findings as well as some of the above discrepancies between the baseline model and the data. In particular, the extended model predicts that all principals will report truthfully if the agent is successful, but if the agent has failed, then some will still report truthfully but the rest will lie if they can, or give no feedback.

Interestingly, the model also shows that in public-unverifiable feedback, it is indeed rational for an agent who received positive feedback to be influenced adversely by the other agent's positive feedback. This is because, in equilibrium, the principal provides positive feedback to, say, agent 1 and negative feedback to agent 2 only when the outcome is success for agent 1 and failure for agent 2, whereas she provides positive feedback to both agents after all four possible outcomes, which includes failure for agent 1. A similar effect, however, does not exist if own feedback is negative, which is also in line with the evidence. This still does not explain why this effect also exists in public-verifiable feedback. However, the fact that it is stronger in public-unverifiable feedback and significant only when own feedback is positive indicates that agents do consider the principal's strategy in forming their beliefs.

We address this issue in more detail in Section 6.2 by comparing agents' actual post-feedback beliefs with hypothetical beliefs that a Bayesian agent would form if he perfectly predicted the (empirical) strategy used by the principals. Our analysis suggests that Bayesian updating plays a significant role in the formation of beliefs, but agents are, on average, overly optimistic in responding to their own feedback and interpret positive feedback to the other agent more pessimistically than they should.

Overall, we conclude that private-verifiable feedback is the most informative mechanism while unverifiable feedback is not informative, and public feedback interferes with the informativeness of positive feedback, especially when it is unverifiable.

2 Related Literature

To the best of our knowledge, this is the first comprehensive study that explores, both theoretically and experimentally, the impact of verifiability and audience on strategic information transmission in a realistic performance feedback context. Previous theoretical and experimental studies of performance feedback have mostly focused on the effects of *truthful* feedback on effort decisions and future performance. Theoretical work has generally used principal-agent models to study optimal information revelation mechanisms under the assumption of truthful feedback, taking into account the effects of the feedback on agents' actions (see, for example, Ertac [2005], Ederer [2010], Aoyagi [2010]). The experimental literature has mostly studied the motivational effects of truthful performance feedback in both organizational and educational settings and documented varying results. With flat wages, the majority of papers find that provision of relative performance feedback leads to higher effort on average, whereas evidence is more mixed in performance-pay settings.[5] Our major departure from this literature is that we consider *strategic* rather than truthful feedback and focus on the communication aspects.

Ederer and Fehr [2009] is one of the few experimental papers that study strategic performance feedback. They analyze the effect of private-unverifiable feedback on (induced) effort in a dynamic tournament with two agents. In their setting, the principal has an incentive to underreport the true performance difference between the agents. Hence, relative performance feedback should be completely uninformative and agents should not respond to feedback in equilibrium. In contrast, their results show that even though agents discount the information they receive from the principal, they still respond to it and some principals provide feedback that is close to the truth while others consistently underreport.[6]

In our private-unverifiable feedback treatment, we find a similar result to Ederer and Fehr [2009] in the sense that some principals tell the truth while others lie; however, in our case, agents heavily discount such feedback, which renders it uninformative. From a design perspective, our work is distinct from Ederer and Fehr [2009], as well as from the other papers in this literature, along several lines: (1) We vary treatments along the dimensions of both audience and verifiability and study their interaction, while Ederer and Fehr [2009] study only private-unverifiable feedback. This enables us to compare different feedback mechanisms along dimensions that may be discretionary in organizational and educational settings and to draw policy conclusions; (2) We study a non-tournament setting where information about the other agent's performance is irrelevant. These two aspects of our research allow us to uncover, both theoretically and experimentally, a novel finding: When feedback is public and the other agent receives positive feedback, agents interpret their own feedback more pessimistically - apparently by making inferences about the principal's strategy. Furthermore, this effect is stronger if feedback is unverifiable and own feedback is also positive; (3) We measure the impact of feedback directly on beliefs, rather than effort, using an incentive-compatible mechanism that is also robust to risk aversion. This allows us to isolate, in a clean manner, the strategic communi-

[5] See, among others, Azmat and Iriberri [2010], Azmat and Iriberri [2012], Bandiera et al. [2010], Blanes i Vidal and Nossol [2011], Charness et al. [2010], Eriksson et al. [2009], Gerhards and Siemer [2014], Gill et al. [2015], and Kuhnen and Tymula [2012].

[6] In a one principal/one agent setting with unverifiable feedback and induced effort, Mohnen and Manthei [2006] find similar results: Some principals tell the truth but deception is also widespread. Rosaz [2012] also studies unverifiable feedback in a one principal/one agent setting, but limits the principal's ability to lie. She finds that the principal indeed manipulates the feedback but the agent increases effort in response.

cation aspect of performance feedback, which is the main focus of this paper; (4) We use a real, rather than induced, effort setting, which creates an ego-relevant environment that should contribute to the external validity of our results.[7]

Gürtler and Harbring [2010] also study the effect of performance feedback on effort in a tournament setting, but unlike in Ederer and Fehr [2009], feedback is public and verifiable in their design. The theory, in this case, predicts that agents should interpret no feedback as bad news and full revelation of relative performance should occur. They find that although there is evidence that no feedback is regarded as bad news, the effect on effort is not as strong as the theory predicts.

Verifiable feedback mechanisms induce a strategic communication game that is known as a "disclosure" (or persuasion) game in the literature, pioneered by Grossman [1981] and Milgrom [1981], while unverifiable feedback mechanisms induce what is known as a "cheap talk" game, first studied by Crawford and Sobel [1982].[8] Therefore, our paper is also related to the literature that experimentally tests the predictions of cheap talk and disclosure games. In these strands of the literature, Dickhaut et al. [1995] and Cai and Wang [2006] find support for the qualitative predictions of the basic cheap talk model of Crawford and Sobel [1982], i.e., less information is transmitted as preferences of the sender and the receiver diverge, and Battaglini and Makarov [2010] find overall support for the predictions of Farrell and Gibbons [1989], which extends the basic model to the case of multiple receivers. Blume et al. [2001], Cai and Wang [2006], and Battaglini and Makarov [2010] find evidence for over-communication, i.e., a tendency for the senders to reveal more information than predicted by theory as well as a tendency for the receivers to rely on the information sent by the senders.[9] Drugov et al. [2013] test the two-receiver model by using five states rather than two and, similar to our setting, also run a private communication mode. They report evidence of a disciplining effect of public communication.[10]

The experimental literature on the strategic communication of verifiable information is smaller. Early work on experimental tests of disclosure games has studied disclosure in the context of markets, where the seller is better-informed and discloses quality to the buyer(s). Forsythe et al. [1989] find that full information revelation is achieved, but only as subjects become more sophisticated over repeated rounds of play. King and Wallin [1991] analyze a market setting where the seller may or may not be informed, which is unknown to the buyers, and find that full disclosure does not occur. Forsythe et al. [1999] find that imposing "anti-fraud" rules that constrain message sets to include the true state improves efficiency in comparison to cheap talk messages. More recently, Benndorf et al.

[7] This obviously has a cost in terms of control over unobservables and makes the match with theory more difficult. Since the focus of our work is not the effect of feedback on effort, we believe that the benefits of using real effort outweigh the costs.

[8] We discuss the relevant theoretical literature in Section 4.

[9] Cai and Wang [2006] explains this over-communication behavior using level-k behavior and quantal response equilibrium. Using information from eye-tracking technology, Wang et al. [2010] shows that senders look at payoffs in a way that is consistent with a level-k model. However, over-communication that persists over rounds, as in Blume et al. [2001] is difficult to explain by level-k reasoning. A potential explanation for over-communication on the part of senders is "lying aversion". Gneezy [2005] reports experimental evidence that subjects have a tendency to tell the truth even if it is against their material interests. Gneezy et al. [2013] study the same question using a new method and find that subjects are heterogenous with regard to their tendency to lie. See Charness and Dufwenberg [2006], Hurkens and Kartik [2009], Sutter [2009], Sanchez-Pages and Vorsatz [2009], Abeler et al. [2012], and Fischbacher and Föllmi-Heusi [2013] for further experimental evidence on lying aversion.

[10] See Crawford [1998] for an early survey of experimental work on strategic communication. There is also more recent experimental work on extensions of the basic cheap talk model to multiple dimensions and multiple senders, such as Lai et al. [2015] and Vespa and Wilson [2015].

[2015] find, in a labor-market experiment with a lemons structure where workers can reveal their productivity, that revelation takes place less frequently than predicted in equilibrium. The experimental context and decision settings used in these papers (e.g., asset markets, auction context) tend to include elements that may affect behavior independently of the basic strategic considerations in verifiable information disclosure. Jin et al. [2015] use a more direct test of the "no news is bad news" prediction in disclosure games, and find that receivers do not interpret no information sufficiently negatively. Hagenbach and Perez-Richet [2015] test the predictions of Hagenbach et al. [2014] by considering payoff structures for the sender that are not necessarily monotonic in the receiver's action, and find that whether the game is cyclic or acyclic matters for the receivers in forming skeptical beliefs and thereby for information transmission.

While some of our results, such as the tendency to tell the truth with unverifiable feedback and insufficient strategic discounting of no feedback with verifiable feedback are also reported in the existing experimental studies of strategic communication, our work is distinct along several dimensions. First, we elicit agents' beliefs and principal's expectations on agents' beliefs directly, while previous work has studied the effect of information on other strategic choices, which may be confounded by risk aversion or other factors specific to the decision environment. This allows us to more clearly focus on the motives behind giving feedback and its interpretation.[11] Second, previous work has tested the predictions of cheap talk or disclosure games usually by varying the preferences of the players, while we take the preferences as fixed and vary both verifiability and audience. As we have mentioned before, this allows us to study the interaction between these two dimensions and leads to novel findings. Third, our main purpose is to test informativeness of different performance feedback mechanisms in a real-effort context, while the previous work has either used a neutral framework to test game theoretical predictions or studied other specific environments such as auctions or labor markets.

3 Experimental Design

The experimental design is based on studying interim performance feedback in a one principal-two agent real effort context. The performance feedback technology available to the principal is the treatment variable, and we study five treatments in a within-subject design. Therefore, the experiment consists of five periods with each period corresponding to a different feedback mechanism, and within each period there are two rounds. To eliminate potential wealth effects, we use a random payment scheme, i.e., one of the ten rounds is chosen randomly and subjects are paid according to their payoffs in the chosen round.

At the start of the experiment, subjects are randomly assigned to the roles of either "Principal" or "Agent", and these roles do not change. In each period, 3-person groups, which consist of one principal and two agents, are formed. We use a "strangers" matching protocol, where new groups are randomly formed at the start of every period.

[11] This feature of the design relates the paper to the experimental literature on the effects of noisy but non-strategic feedback on beliefs. Our finding that agents respond to feedback more optimistically than they should is in line with and complements the findings in this literature that subjects may process information differently and exhibit biases of asymmetry or conservatism (in comparison to Bayesian updating) when the context is ego-relevant (Ertac [2011], Eil and Rao [2011], Mobius et al. [2011]). However, we study feedback that is provided strategically, which makes a difference because, as we will show, many principals indeed act strategically and agents take that into account in updating their beliefs.

For participants in the role of agents, we use two different real-effort tasks: an addition task and a verbal task (see Appendix B for details). The verbal task consists of general knowledge questions as well as verbal classification and number-letter matching questions. The addition task involves adding four or five two-digit numbers.

In each period, agents are randomly assigned to one of these tasks and perform the same task in both rounds of that period. For both tasks, subjects are asked to solve as many questions as possible within a limited time (2 mins.). At the end of each round, the number of correct answers is compared to a "target score", randomly determined for that specific period.[12] The same target score is employed in both rounds of the period. If a subject's score is greater than or equal to the target score, the subject is "successful", and has failed otherwise. Note that the target score is subject-specific and there is no common shock applied to the performance of subjects.

3.1 Belief elicitation

To elicit self-confidence, we use a crossover mechanism developed independently by Karni [2009] and Mobius et al. [2011], which is a Becker-De Groot-Marschak-type procedure for eliciting beliefs truthfully and independently of risk preferences. In this mechanism, subjects are presented with two alternative lotteries to determine their second-round payoff. In the performance-based lottery, the reward is based on the agent's second-round performance. That is, the agent receives the reward if his outcome is "successful" in the second-round performance stage. In the chance-based lottery, the agent earns the reward with probability X, regardless of his second-round performance. At the end of the first performance round, subjects are asked to report the minimum probability of winning in the chance-based lottery that would make them willing to choose the chance-based lottery as opposed to the performance-based one. The computer then draws X randomly. If the randomly drawn X is at least as large as the agent's stated minimum, the chance lottery applies. Otherwise, the agent is rewarded based on his second-round performance. This mechanism gives agents an incentive to truthfully report the subjective probability with which they think they will succeed in the second round. In order to study the within-person effect of performance feedback on beliefs, we ask the subjects to make this decision twice: once before and once after receiving feedback. To maintain incentive compatibility, we randomly choose either the pre-feedback or post-feedback beliefs to determine whether the performance or chance mechanism will be implemented.

The timeline of a period for agents is as follows:

1. Pre-feedback performance: Subjects perform the assigned task within 2 minutes.
2. Pre-feedback beliefs: Without receiving any information, subjects state the minimum probability of winning that would induce them to leave their second-round payoff to chance.
3. Feedback: Feedback is received, in the form of a message whose content changes between treatments, as will be explained in Section 3.2.
4. Post-feedback beliefs: After seeing the message (or no message), subjects are allowed to update their previously reported beliefs. (At this stage, the subjects can see their previously reported beliefs on the screen.)

[12] The target score is a number which is randomly chosen from the interval [4,13] at the beginning of each period. The range of the target score was determined based on data from a pilot session.

5. Performance/chance mechanism: If the self-reported probability of winning (either pre- or post-feedback, depending on which was selected) is higher than the probability of winning in the chance mechanism (drawn by the computer), then the subject performs the same type of task for two minutes again, as in the first round. Otherwise, they do not perform the task, and their second-round payoff is determined by chance, according to the winning probability drawn by the computer.

3.2 Feedback mechanism

Note that after the first round, agents do not have exact knowledge of whether they were successful, although they will have subjective beliefs. Principals, on the other hand, observe the true first-round outcomes (success or failure) of the two agents they have been matched with. After stating their priors, agents may receive a message about whether they were successful in the first round. There are five types of feedback mechanisms used throughout the experiment, which differ in the provider, audience, and content of the feedback. In terms of content, we have the following types:

1. Truthful feedback: In this mechanism, subjects receive an accurate message (success or failure) from the computer. This is the baseline mechanism in our design.

2. Verifiable feedback: In this mechanism, performance feedback is reported by the principal. The principal can choose either to transmit the true outcome (success or failure), or to withhold the information. Sent messages always have to be correct, and agents know that there can be no deception.

3. Unverifiable feedback (cheap talk): As in the verifiable mechanism, the feedback comes from the principal, but she does not have to report the actual outcome, i.e., she can lie. In addition, she has an option to send no message.

Within the verifiable and unverifiable mechanisms, we also employ two different feedback types that differ in the audience of the messages:

1. Private feedback: In this mechanism, the principal reports the feedback independently and privately to the agents, and agents only see the message targeted to them.

2. Public feedback: In this mechanism, the principal has to announce the feedback publicly. That is, each agent observes the other agent's message, in addition to his own.

This design leaves us with five different feedback treatments, which are implemented within-subject: truthful feedback, private-verifiable feedback, public-verifiable feedback, private-unverifiable feedback, and public-unverifiable feedback. In the public-verifiable case, the principal has to decide either to release the truthful outcome to both of the agents publicly, or to withhold the information. On the other hand, in the public-unverifiable case, the feedback for each agent is chosen separately from the three options explained above (success, failure or no information) and the messages for both agents are delivered publicly to all.

Finally, in order to get a better insight into the feedback strategy employed by the principals, they are asked to guess agents' post-feedback beliefs.

3.3 Payoffs

The payoffs of participants in the role of agents depend on their performance outcomes as well as their decisions. To incentivize performance in the first round, we use differential rewards based on a performance target: 300 ECU (experimental currency unit), if the agent succeeds, and 100 ECU, if he fails. (1 ECU = 0.06 Turkish Liras (TL).)

In the second round, if the agent ends up doing the task, his payoff depends on whether he succeeds or fails, exactly as in the first round. If, however, the agent ends up with the chance mechanism, then his second-round earnings are 300 ECU with probability X, and 100 ECU with probability $(1-X)$, where X is the randomly chosen probability of winning.

The principal's payoff, on the other hand, depends on the second-round entry behavior and performance outcomes of the two agents. For the principal, we use a payoff function in which the performances of the two agents are complements. Specifically, the payoff function is:

$$V_t = \begin{cases} 100, & t = 2n-1 \\ 50 + 10(g_{1t} + g_{2t}) + \min\{q_{1t}, q_{2t}\}, & t = 2n \end{cases}$$

where $n \in \{1,2,3,4,5\}$ is the period number, q_{it} is the return from the second-round performance of agent i in period t, and g_{it} is an indicator variable that takes the value of 1 if the principal's guess in period t for agent i is correct, i.e., in the ± 5 interval of agent's actual belief, 0 otherwise. Return from the performance of agent i in period t is equal to

$$q_{it}(c_{it}, e_{it}) = \begin{cases} 20 * c_{it}, & e_{it} = 1 \\ 0, & e_{it} = 0 \end{cases}$$

where c_{it} denotes the number of correct answers of agent i in period t, while e_{it} represents the entry of agent i to the performance stage (as opposed to taking the chance mechanism). In the first round the principal's payoff is a constant amount, 100 ECU. The second-round payoff is composed of three elements: a constant amount, 50 ECU, an extra 10 ECU for each correct guess about the agents' beliefs, and the minimum of the returns from both agents. As can be seen from the above payoff function, for the principal to earn an extra return over the fixed endowment, both agents must end up doing the task. This, together with complementarity, implies that the principal should aim to (1) convince both agents that they are likely to succeed in the second round task, and (2) maximize the post-feedback performance of the worst-performing agent in the second round.

3.4 Procedures

The experiment was programmed using the z-Tree experimental software (Fischbacher [2007]), and implemented at the Koç University and Bilgi University computer labs in the Spring term of 2013. We collected data from 132 subjects in total (72 subjects from Koç University and 60 subjects from Bilgi University). The experiment was conducted in 13 sessions, with 8 sessions at Koç University and 5 sessions at Bilgi University. Our sample consists of 68 male and 64 female participants, who are mostly undergraduates. At the end of each session, we conducted a survey to collect demographic data such as age, gender, major and GPA (see Appendix B). In order to mitigate potential order effects, we used 6 different configurations that differ in the sequence of treatments.[13] Sessions lasted about 50 minutes, and subjects earned between 15 TL and 28 TL (on average 20.23 TL), including a show-up fee.

4 Theory and Predictions

In this section we will analyze a stylized model of our experimental design and derive theoretical predictions that will form the basis for the empirical analysis in Section 5. There are two agents, indexed by $i = 1,2$, and a principal, denoted by P. For each agent i, a state of the world θ_i is realized and observed only by the principal. In our experimental design, this state corresponds to either "success" or "failure", denoted by s and f, respectively. We assume that states are independently distributed across agents and the probability of success for agent i is equal to $p_i \in (0,1)$. We will also assume for simplicity of exposition that s and f are real numbers with $s > f$.

After observing (θ_1,θ_2), the principal provides feedback to the agents. As we have explained in Section 3, this feedback might be verifiable, in which case, the principal cannot lie but still choose to give no information, or might be unverifiable, i.e., might be cheap talk, in which case the principal can lie about the actual state of the world or provide no information. Feedback is either private, in which case the principal provides feedback on θ_i to each agent i independently and privately, or public, in which case both agents observe the common feedback about (θ_1,θ_2). After receiving feedback, each agent independently chooses an action and the game ends. In our experimental design, this action corresponds to the choice made by the agent in the belief elicitation round. As we have explained before, our belief elicitation mechanism is designed so that it is optimal for each agent to choose the probability with which he believes that he will be successful in the second-round task.

Payoff function of agent i is given by $u_i(a_i,\theta_i)$, where $a_i \in A_i$ is the action choice of agent i and A_i is a compact and convex set of real numbers. Principal's payoff function is $v(a,\theta)$, where $a = (a_1, a_2)$ and $\theta = (\theta_1,\theta_2)$. We assume that players are expected payoff maximizers. If agent i believes that he is successful, i.e., $\theta_i = s$, with probability μ_i, his expected payoff is equal to $U_i(a_i,\mu_i) = \mu_i u_i(a_i, s) + (1-\mu_i)u_i(a_i, f)$. We assume that, for each $\mu_i \in [0,1]$ there is a unique maximizer of $U(a_i,\mu_i)$, denoted $a_i^*(\mu_i)$, which is in the interior of A_i and strictly increasing in μ_i. From now on, whenever we say that agent i has high beliefs we mean that μ_i is high.

We also assume that the principal's payoff function is strictly increasing in a_i, $i = 1,2$. This makes feedback a strategically important choice for the principal because she has an incentive to induce a high belief by each agent. This, of course, may render her feedback unreliable in equilibrium and the extent to which this happens may depend on the feedback technology itself, i.e., whether the feedback is private or public and verifiable or not. The main theoretical issue we deal with in this section is the informativeness of the feedback provided by the principal in these different cases.

Denote the set of states as $\Theta = \{f,s\}$ and the set of messages that can be potentially sent by the principal as $M = \{f,s,\emptyset\}$, where $\emptyset$ denotes no information. Let $M(\theta)$ be the set of messages that are feasible when the the state is $\theta = (\theta_1,\theta_2)$. The following describes the set of strategies available to the principal under different treatments:

[13] The configurations were as follows: TVU, TUV, VUT, UVT, VTU and UTV, where T, V and U correspond to Truthful, Verifiable and Unverifiable feedback mechanisms, respectively.

1. Private Feedback: A pure strategy of the principal is a pair of functions $\rho = (\rho_1, \rho_2)$, where $\rho_i : \Theta^2 \to M_i(\theta)$. If feedback is unverifiable, then $M_i(\theta) = M$, i.e., there are no restrictions on the feasible messages. If feedback is verifiable, then $M_i(\theta) = \{\theta_i, \emptyset\}$, i.e., principal either tells the truth or provides no information to an agent.

2. Public Feedback: A pure strategy of the principal is a function $\rho : \Theta \to M(\theta)$. If feedback is unverifiable, then $M(\theta) = M^2$, i.e., there are no restrictions on the feasible messages. If feedback is verifiable, then $M(\theta) = \{\theta, \emptyset\}$, i.e., principal either tells the truth or provides no information to both agents.

After observing feedback r, agent i forms beliefs on the state of the world $\mu_i(r) \in [0,1]^2$ and chooses an action $\alpha_i(r) \in A_i$. Let μ_i^i denote the probability that agent i's beliefs put on the event $\theta_i = s$ and μ_i^{-i} the probability on $\theta_{-i} = s$. Let $\mu = (\mu_1, \mu_2)$ and $\alpha = (\alpha_1, \alpha_2)$ denote, respectively, an agent belief profile and strategy profile. An assessment is composed of a strategy for each player and beliefs by the agents: (ρ, α, μ).

An assessment is a perfect Bayesian equilibrium (PBE) if strategies are optimal given beliefs and beliefs are formed by using Bayes' rule whenever possible. In what follows we will analyze the set of pure strategy perfect Bayesian equilibria of each extensive form game defined by one of the four possible feedback mechanisms: (1) private-verifiable; (2) public-verifiable; (3) private-unverifiable; (4) public-unverifiable.

A verifiable feedback mechanism induces a game of strategic communication known as a "disclosure game", pioneered by Grossman [1981] and Milgrom [1981], while an unverifiable feedback mechanism induces a "cheap talk game", introduced by Crawford and Sobel [1982]. These basic models and the main results have been later generalized and extended in several directions.[14] Most relevant for us are Farrell and Gibbons [1989] and Koessler [2008], both of which consider a two-receiver, two-state, and two-action model and analyze public and private communication. Farrell and Gibbons [1989] consider only the cheap talk case, while Koessler [2008] extends it to verifiable messages. Our model differs from theirs in that the state is multidimensional (which is formally equivalent to four states) and there is a continuum of actions. None of our results on public feedback follows directly from the analyses in these two papers, but the reasoning behind the existence of partially informative equilibrium in public-verifiable feedback is similar to the case of the mutual subversion in Koessler [2008], and the partially informative equilibrium in public-unverifiable feedback resembles the mutual discipline case in Farrell and Gibbons [1989].[15]

4.1 Verifiable Feedback

Each agent has (or updates) his beliefs regarding the other agent's type as well as his own type. However, since types are independent and only own type affects payoffs, what matters strategically is only beliefs on own type. Accordingly, we say that an equilibrium is fully informative if each agent can

[14] For the literature on disclosure games see Seidmann and Winter [1997], Mathis [2008], Giovannoni and Seidmann [2007], and Hagenbach et al. [2014]. The basic cheap talk model in Crawford and Sobel [1982] has also been extended in many directions. See Sobel [2013] for a recent survey of this large literature.

[15] We should also mention Goltsman and Pavlov [2011], which generalizes Crawford and Sobel [1982] to the case of two receivers with different preferences and compares public with private feedback. Again, our model's state space and payoff structure are different in a way that makes direct application of their results impossible.

infer his type from the principal's report and completely uninformative if agents learn nothing about their own type.

Our first result shows that if feedback is verifiable, then agents receive perfect information about their own types.

Proposition 1. *If feedback is private and verifiable, then all equilibria are fully informative.*

Proof. All proofs are relegated to Appendix A. □

Proof of Proposition 1 is very easy. If feedback is verifiable and the principal learns that an agent is successful, then she can simply send the message that he is successful and induce the best beliefs and the highest action on the part of that agent. Since feedback is verifiable, the other type of the principal, i.e., the type who observed that the agent has failed cannot mimic this feedback. This full revelation result is well known in the literature and follows from two aspects of our model: (1) every type has a message that only that type can send; (2) the principal's payoff is monotonic in each agent's beliefs.

In public feedback, the principal cannot change her reporting strategy regarding one agent's performance without the other agent observing this change. This creates the main difference between private and public feedback for equilibrium analysis. Indeed, if feedback is public and verifiable, then full information revelation is an equilibrium but, in contrast to private feedback, there is also a partially informative equilibrium.

Proposition 2. *If feedback is public and verifiable, then in equilibrium there is either full information revelation or $\rho(s,s) = (s,s)$ and $\rho(\theta) = \emptyset$ for all $\theta \neq (s,s)$.*

It is easy to construct a fully revealing equilibrium by specifying strategies $\rho(\theta) = \theta$ for all θ and beliefs as $\mu_i(\emptyset) = 0$ for $i = 1,2$. The following example shows that there is also a partially informative equilibrium.[16]

Example 1. Let $s = 7$, $f = 1$, $p_i = 1/2$, and payoff functions be $u_i(a_i,\theta_i) = \theta_i a_i - \frac{1}{2}a_i^2$ and $v(a,\theta) = a_1 a_2 w(\min\{\theta_1,\theta_2\})$, where w is a strictly increasing function with $w(1) > 0$. It can be shown that the following assessment is an equilibrium: $\rho(s,s) = (s,s)$, $\rho(\theta) = \emptyset$ for all $\theta \neq (s,s)$, $\mu_i^i(\theta) = 1$ if $\theta_i = s$ and $\mu_i^i(\theta) = 0$ otherwise, $\mu_i^i(\emptyset) = 1/3$, $\alpha_i(\theta) = \theta_i$, $\alpha_i(\emptyset) = 3$. What makes this example work is the form of the principal's payoff function, which is similar to the one in our experiment and has the property that intermediate beliefs by both agents is better for the principal than extreme beliefs. It is easy to show that this property exists as long as principal's payoff function is symmetric, concave, and strictly supermodular in a.[17]

The above results and Bayes' rule imply the following prediction:

Prediction A. *In verifiable feedback:*

1. *Beliefs increase conditional on success and decrease conditional on failure;*

[16] Results in Milgrom [1981] and Seidmann and Winter [1997] imply that there is an equilibrium with full information revelation in our game. However, their uniqueness result does not apply to the public feedback case because the action and the type spaces are multidimensional. In fact, the example shows that there is an equilibrium with less than full information revelation.

[17] We should note that the partially informative equilibrium constructed in the example depends on the common knowledge assumption on p_i, which may not be satisfied in the experiments. We are grateful to a referee for pointing this out.

2. *Beliefs increase after positive feedback and decrease after negative or no feedback;*

3. *Beliefs do not depend on the feedback provided to the other agent;*

4. *If feedback is private, the principal reports truthfully to the agent who succeeds and either reports truthfully or gives no feedback to the agent who fails. If feedback is public, principal reports truthfully if both agents succeed and either reports truthfully or gives no feedback if one of the agents fails.*

4.2 Unverifiable Feedback

If feedback is unverifiable and private, then there is no information transmission in equilibrium.

Proposition 3. *If feedback is private and unverifiable, then all equilibria are completely uninformative.*

Proof of this result is also simple. If some message induces higher beliefs for some agent, then all types of the principal would have an incentive to send that message, contradicting the hypothesis that this message induces higher beliefs. This again simply follows from the fact that principal's payoff is monotonic in each agent's beliefs.

If feedback is public, then there is always a completely uninformative equilibrium and never a fully informative equilibrium. Furthermore, in any equilibrium, types (s,s) and (f,f) must always give the same feedback.

Proposition 4. *If feedback is public and unverifiable, then there is a completely uninformative equilibrium. In any equilibrium $\rho(s,s) = \rho(f,f)$ and hence fully informative equilibrium does not exist.*

Therefore, equilibrium is at most partially informative and whether feedback provides any information at all, depends on the payoff function of the principal. For instance, in Example 1 all equilibria are completely uninformative. Since principal's payoff function in that example is similar to the one in the experiment, we expect feedback to be uninformative in the experiment as well.

Since messages have no intrinsic meaning and are completely costless in our model, there is no precise prediction regarding the principal's strategy and agents' beliefs after feedback. However, in our experiment, as well as in real life, reports have a natural meaning and hence it is plausible to expect that a principal who observes success always reports success. This implies that, in equilibrium, the principal who observes failure must also report success. Therefore, we have the following prediction:

Prediction B. *In unverifiable feedback:*

1. *Beliefs do not change conditional on actual state;*

2. *Beliefs do not change in response to positive feedback and decrease or stay the same after negative or no feedback;*

3. *Principal always provides positive feedback.*

The above analysis also implies the following prediction:

Prediction C. *Private-verifiable feedback is the most informative mechanism and private-unverifiable feedback is not informative.*

5 Results

The main focus of our study is whether performance feedback is informative and whether this depends on the verifiability and privacy of the feedback. Section 5.1 mainly presents our results on this issue. Our model also produces theoretical predictions regarding the principal's behavior in different treatments. We therefore present a summary of the principals' behavior in Section 5.2 and discuss how it fits with the theoretical predictions.

5.1 Analysis of Agents' Behavior

We start with some summary statistics about task performance. On average (in both rounds), subjects attempted to solve 8.79 questions in the addition task and 10.47 questions in the verbal task, and correctly solved 7.08 and 7.82 questions, respectively. The answers to a survey question that asks whether it is important for subjects to succeed independently of its monetary payoff reveal that a majority of subjects do care about success per se.[18] This shows that we have managed to create an ego-relevant performance environment for subjects in our experiment, which is important for analyzing belief updating in a realistic fashion.

We first examine the initial (pre-feedback) beliefs of the subjects who have been assigned the role of an agent. Pre-feedback beliefs show that most agents prefer to perform in the second round: Average belief is 0.66 while the median is 0.7. In other words, on average, they believe that they will succeed with probability 0.66 if they were to perform the task. Since only 51% of the subjects successfully pass the target score upon entry, we conclude that participants overestimate their performance, i.e., they are overconfident. This is consistent with results from other real-effort experiments in the literature (e.g., Hoelzl and Rustichini [2005]), and highlights the benefit of using real effort, because in reality, overconfidence or self-serving biases may influence how agents interpret feedback given by the principal.[19]

5.1.1 Information Transmission

We start by analyzing how beliefs change conditional on the actual performance outcome of the agent. If there is information transmission, then beliefs should move up for successful agents and down for unsuccessful ones. Figure 1 shows that there is information transmission in verifiable feedback and no information transmission in unverifiable feedback cases. Wilcoxon sign-rank tests indicate that the actual outcome has a significant effect on agents' beliefs in truthful and verifiable feedback treatments, while it has no significant impact in unverifiable feedback treatments.[20]

[Figure 1 about here.]

[18] The mean assessment of subjects is 3.73 on a 1-5 scale and 75.84% of them choose either the important or very important option (Appendix B.3, question 10).

[19] In order to ensure that there is no selection bias, we test whether prior beliefs are independent of treatments and order configurations, and find no significant differences. Neither do we find differences in the prior beliefs of Bilgi and Koç University students. Related regressions are available upon request

[20] The p-values of Wilcoxon signed rank test for the hypothesis of a zero change in beliefs in each treatment is as follows: Truthful with $p = 0.0002$, private-verifiable with $p = 0.0004$, public-verifiable with $p = 0.013$, private-unverifiable with $p = 0.110$, public-unverifiable with $p = 0.787$ for success; truthful with $p = 0.0002$, private-verifiable with $p = 0.023$, public-verifiable with $p = 0.017$, private-unverifiable with $p = 0.696$, public-unverifiable with $p = 0.148$ for failure cases.

Likewise, regression analysis shows that private-verifiable feedback is not significantly different from truthful feedback under either success or failure, while public-verifiable feedback is significantly different from truthful feedback only under success (and only at the 10% level). In contrast, both types of unverifiable feedback lead to a change in beliefs that is further away from the effect of truthful feedback (see Table 1).[21]

[Table 1 about here.]

Finally, Figure 1 and Table 1 show that private-verifiable feedback is the closest mechanism to truthful feedback.

We can summarize our findings as follows:

Result 1. *Verifiable feedback is informative while unverifiable feedback is not. Private-verifible feedback is the most informative mechanism. These results confirm Predictions A.1, B.1, and C.*

5.1.2 Impact of Feedback on Beliefs

Figure 2 and Wilcoxon signed-rank tests show that both positive and negative feedback have significant (positive and negative, respectively) effects on beliefs in all treatments except the private-unverifiable feedback treatment.[22] On the other hand, the change in beliefs after no feedback is not significantly different from zero in any of the treatments.[23]

[Figure 2 about here.]

The regressions in Table 2 further explore the differences in the impact of feedback on beliefs across treatments. The table shows that in terms of direction, agents tend to discount the principal's feedback in all the treatments: positive feedback is interpreted less optimistically than truthful positive feedback and negative feedback less pessimistically. However, in response to positive feedback, the change in beliefs under private-verifiable feedback is not significantly different from that under truthful feedback, while all the other treatments induce significantly lower beliefs than truthful positive feedback (see column (1)).[24] When subjects receive negative feedback, none of the treatments, except private-unverifiable feedback, is different from truthful feedback. Under no feedback, on the other hand, we find that there is no significant difference across treatments.

[Table 2 about here.]

We can summarize our findings as follows:

Result 2. *In all feedback mechanisms positive feedback increases and negative feedback decreases beliefs, but in private-unverifiable feedback these changes are insignificant. In all mechanisms, no feedback leads to only insignificant changes in beliefs. Predictions A.2 and B.2 are confirmed except that*

[21] Note that we collect data over different rounds from the same subject in all of the treatments. Thus, to account for correlation, we use random effects model in regressions that use multiple observations from the same subject.

[22] The p-values of Wilcoxon signed rank test for each treatment are as follows: Truthful with $p = 0.0002$ and $p = 0.0002$, private-verifiable with $p = 0.0002$ and $p = 0.047$, public-verifiable with $p = 0.002$ and $p = 0.003$, public-unverifiable with $p = 0.005$ and $p = 0.090$, private-unverifiable with $p = 0.236$ and $p = 0.487$, for positive and negative feedback cases respectively.

[23] The p-values of Wilcoxon signed rank test for each treatment are $p = 0.211$, $p = 0.980$, $p = 0.674$, $p = 0.710$, for private-verifiable, public-verifiable, private-unverifiable, and public-unverifiable, respectively.

[24] Note, however, that the coefficient of public-verifiable feedback is only marginally significant.

no feedback does not decrease beliefs in verifiable feedback and positive feedback increases beliefs in public-unverifiable feedback.

The above findings suggest that public feedback can have quite different effects than private feedback. To explore this further, we look at whether beliefs are affected by the feedback provided to the other agent in public feedback treatments. Figures 3a and 3b show, for each type of own feedback received, whether beliefs respond to the other person's feedback in verifiable and unverifiable cases, respectively. We can see that in both treatments beliefs are affected adversely when the other agent has received positive feedback as opposed to negative feedback. Secondly, the magnitude of this effect is larger under unverifiable feedback than under verifiable feedback.

[Figure 3 about here.]

Columns (4) and (5) of Table 2 test whether, in public feedback treatments, the other agent's feedback makes a difference in belief updating, when own feedback is positive and negative, respectively. The results support the conclusions we have drawn from Figure 3 and further show that the adverse effect of the other agent's positive feedback is significant only if own feedback is positive as well and that the effect is significant only at the 10% level in public-verifiable feedback. Column (4) of Table 2 also shows that the less optimistic response to verifiable positive feedback, i.e., the negative coefficient of public-verifiable feedback in column (1), comes from observations where own positive feedback is accompanied with positive feedback to the other agent.[25]

Therefore, we conclude that:

Result 3. *In public feedback, beliefs are affected adversely when the other agent also receives positive feedback. This effect is stronger if own feedback is also positive and larger under unverifiable feedback. Therefore, Prediction A.3 is rejected.*

5.1.3 Impact of Beliefs and Feedback on Performance

Although it is not the focus of our study, we also examine how beliefs and feedback affect second-round performance. Note that in our experiment, only the agents whose posterior beliefs are larger than a randomly determined threshold perform in the second round and the rest simply receive a randomly determined payoff. In order to minimize ability-based selection and to be able to observe the effect of beliefs on the second-round performance for a relatively unbiased set of subjects, the random device in the belief elicitation mechanism was skewed toward inducing subjects to enter.[26] Consequently, 87% of the subjects performed in the second round. Table 3 shows that, controlling for the first round performance, higher beliefs lead to higher second-round performance, and hence the principal has an additional incentive to induce higher beliefs. We also checked the impact of feedback on performance, both overall and in each treatment separately, and found no significant effect. (These results are available upon request.) Overall, although our experiment is not designed to analyze this issue, we have the following result.

Result 4. *Inducing higher beliefs increases performance but interim performance feedback is not an effective tool in this respect.*

[25] These results are robust to taking the dependent variable to be the posterior beliefs and controlling for the prior beliefs as a regressor.

[26] Note that this does not affect the incentive compatibility of the mechanism.

[Table 3 about here.]

5.2 Analysis of Principals' Behavior

We now turn to explore the principals' side. We first categorize the messages sent by the principals under different feedback mechanisms, depending on the actual outcome (Table 4). As expected, if the actual outcome is success and the principal can privately convey it, a positive message is transmitted in almost all cases, both verifiable (97%) and unverifiable (94%). The percentage of positive messages under public feedback when the actual outcome is success is somewhat lower (82% in verifiable and 85% in unverifiable). This difference between private and public reporting is statistically significant only in the verifiable case ($p = 0.045$ in verifiable and $p = 0.265$ in unverifiable feedback, according to a test of proportions).

[Table 4 about here.]

Table 4 shows that principals prefer to transmit information 44% of the time when the outcome is failure in the private-verifiable case, while the frequency of transmission is 57% under public-verifiable feedback. So in both cases, around half of the time the bad outcome is revealed. This might be either because in equilibrium verifiable negative feedback and no feedback are interpreted similarly, or more likely because some principals have a preference for reporting truthfully.

Similarly, Table 4 shows that when the outcome is failure principals lie and give positive feedback in 54% of the cases in private-unverifiable and 38% of the cases in public-unverifiable feedback. On the other hand, when the outcome is success, they report truthfully in 94% of the cases in private-unverifiable and 85% of the cases in public-unverifiable feedback. This is again consistent with lying aversion. A Pearson chi-square test shows that reports significantly change according to the actual outcome when talk is cheap ($p = 0.0003$ in private-unverifiable and $p = 0.0002$ in public-unverifiable feedback). This confirms that principals consider the actual outcome in reporting, rather than sending random or always positive signals regardless of the true state.

Although the number of observations is small, Table 5 and 6 provide further detail that may help identify the reporting strategies used by the subjects. It seems that when the outcome is success principals always report truthfully, whereas when the outcome is failure, some report truthfully, some lie if they can, and others report no information.

[Table 5 about here.]

[Table 6 about here.]

Our design also allows us to observe the expectations of the principals regarding how agents will update their beliefs. This can potentially give insights into the rationale behind the principals' strategy. As shown in Figure 4, principals expect the positive feedback they send to be interpreted more optimistically than it actually is (although this is not significant in a Wilcoxon test), and negative messages to be evaluated significantly more pessimistically ($p = 0.003$ in a Wilcoxon test). Thus, principals generally overestimate the response of agents' beliefs to the feedback, especially when the feedback is negative. The expectation of a pessimistic response to negative feedback reveals that at least some principals take into account its adverse effect on beliefs but provide negative feedback anyway, which is consistent with an aversion to lying.

[Figure 4 about here.]

Finally, we examine principals' expectations regarding how agents' beliefs will be influenced by the feedback given to the other agent. As Table 7 shows, principals expect that a positive feedback to the other agent will adversely influence the beliefs of an agent when his own feedback is also positive and when feedback is public and unverifiable, but expect no significant impact if own feedback is negative or feedback is verifiable.[27] Interestingly, this is a feature of the equilibrium of the model with lying costs and naive agents, which will be analyzed in Section 6.

[Table 7 about here.]

We can summarize our findings as follows.

Result 5. *Some principals prefer to tell the truth even when they know that this might adversely affect their payoff. Prediction A.4 is confirmed but B.3 is rejected.*

6 Discussion

Overall, our theoretical model in Section 4 does a good job in terms of explaining the relative informativeness of different feedback mechanisms. There are, however, three major discrepancies between our theoretical predictions and empirical findings: (1) Some principals report truthfully even when they believe that this may hurt them; (2) Agents do not interpret "no feedback" as pessimistically as the theory suggests; (3) Positive feedback is interpreted less optimistically if the other agent also receives positive feedback and this effect is stronger in public-unverifiable than in public-verifiable feedback.

The finding that some principals have a tendency to tell the truth is in line with previous empirical studies of strategic communication and suggests that individuals suffer from cost of lying and this cost varies among them. The second finding might be due to naiveté in belief formation, i.e., agents interpret the feedback literally and when they receive "no information", they keep their priors more or less unchanged. Another finding that supports the naive agent hypothesis is that, even in private-verifiable feedback, a significant fraction of principals provide no information when the agent has failed. Since "no information" and negative feedback must both be interpreted in the same (pessimistic) way in private-verifiable feedback, this is not rational if there is even a minimal preference for telling the truth. If, however, principals believe that some of the agents are naive, then this may be optimal. Indeed, Figure 4 and the preceding discussion have indicated that principals expect agents to respond to feedback in a somewhat naive way. Therefore, we conclude that at least some agents are naive and that principals expect them to act naively. The third finding could be due to the fact that agents make (non-Bayesian) social comparisons in forming their beliefs or they believe that the difficulty of the tasks are correlated in such a way that if the other agent has succeeded, then probability of own success in the next task is smaller. Another possible explanation of this finding is that agents are rational and such beliefs simply follow from the principals' strategy and Bayes' rule.

In the next section we extend our theoretical model to allow for individual-specific cost of lying (and cost of withholding information) for the principals and naiveté on the part of the agents. We will

[27] Note, however, that the number of observations is small in some of these regressions.

see that such an extension can account for most of our empirical findings as well as some of the above discrepancies between the predictions of the original model and the data.

6.1 Cost of Lying and Naive Agents

Suppose that lying or providing no information has an individual specific cost associated with it. Let $c(r|\theta)$ be the cost of sending report r when the state is θ and assume that it is distributed according to the probability distribution $F_{r|\theta}$ in the population. Also assume that (1) telling the truth is costless; (2) there are some individuals for whom the cost of lying is small; (3) there are some who always prefer to tell the truth; (4) there are some for whom the difference between the cost of lying and cost of withholding information is small enough; and (5) there are some who prefer withholding information to lying.[28]

A fraction $\eta \in (0,1)$ of agents are naive, i.e., they believe that the state is exactly equal to the principal's report and if the report is "no information", then they keep their prior unchanged. Let $q_i(r|\theta)$ denote the fraction of principals with type θ who send report r to agent i in private feedback, and $q(r|\theta)$ denote the same fraction in public feedback.

Before we present our results, we should briefly discuss the few existing theoretical studies of cheap talk games with lying costs and naive agents. Kartik et al. [2007] show that if the message space is not bounded, then there is a fully revealing equilibrium. Our message space is bounded, which makes full information revelation impossible in the unverifiable feedback case. Kartik [2009] assumes that the sender has a convex cost of lying and characterizes a class of monotone equilibria in which low types separate while high types pool. Chen [2011] analyzes a related model in which the sender is honest and the receiver is naive with positive probabilities and shows that dishonest senders exaggerate the state of the world. Our results do not immediately follow from these two studies because we assume both cost of lying and naive agents and allow cost of lying to differ among senders. Also, we allow sending "no information" and analyze verifiable messages as well as cheap talk.

6.1.1 Verifiable Feedback with Lying Cost and Naive Agents

As the following result shows, under private-verifiable feedback, equilibrium behavior is uniquely determined.

Proposition 5. *If feedback is private and verifiable, then for any i and θ_{-i}*

$$q_i(s|s,\theta_{-i}) = 1, \ q_i(\emptyset|f,\theta_{-i}) > 0, \ q_i(f|f,\theta_{-i}) > 0.$$

Therefore, in equilibrium, if the agent is successful, then the principal gives positive feedback, while if he has failed, then those principals with small costs of withholding information give no feedback while those with large costs report failure. Proportion of principals who give no feedback increases in the fraction of naive agents and the extra benefit of letting the agent keep the prior beliefs. Note that in our model of Section 4, which assumed lying is costless, behavior of the principal when

[28] These assumptions are equivalent to the following: (1) $F_{\theta|\theta}(x) = 1$ for all $x \geq 0$; (2) $F_{r|\theta}(x) > 0$ for all r,θ and $x > 0$; (3) $F_{r|\theta}(v(a_1(1),a_2(1),\theta) - v(a_1(0),a_2(0),\theta) < 1$; (4) $c(s,r_{-i}|f,\theta_{-i}) - c(\emptyset,r_{-i}|f,\theta_{-i})$ is a non-negative random variable with probability distribution $G\big(.|r_{-i},\theta_{-i}\big)$ such that $G\big(x|r_{-i},\theta_{-i}\big) > 0$ for all r_{-i},θ_{-i} and $x > 0$; (5) $G\big(v(a_i(1),a_{-i}(\mu_{-i}(r_{-i})),f,\theta_{-i}) - v(a_i(0),a_{-i}(\mu_{-i}(r_{-i})),f,\theta_{-i})|r_{-i},\theta_{-i}\big) < 1$ for all r_{-i},θ_{-i}.

the agent has failed was indeterminate, i.e., sending negative feedback and no feedback were both compatible with equilibrium. In the current model, principal's behavior is unique given his cost of lying. Also note that if there were no naive agents, then in equilibrium we would not observe any principal who provides no feedback.

Proposition 5 and Bayes' rule imply that beliefs significantly increase after positive feedback and decrease after negative feedback, while beliefs after no feedback decrease but at a magnitude smaller than beliefs after negative feedback. Beliefs conditional on success increase and conditional on failure decrease. (See Appendix A for the calculation of beliefs in this section.)

Equilibrium behavior is also unique in public-verifiable feedback.

Proposition 6. *If feedback is public and verifiable, then* $q(ss|ss) = 1$*,* $q(\emptyset|ff) > 0$*, and* $q(ff|ff) > 0$*. If*

$$\eta^2 v(a_1(p), a_2(p), sf) + 2\eta(1-\eta)v(a_1(p), a_2(0), sf) + (1-\eta)^2 v(a_1(0), a_2(0), sf) > v(a_1(1), a_2(0), sf) \quad (1)$$

then, $q(\emptyset|sf) > 0$ *and* $q(\emptyset|fs) > 0$*. If condition (1) does not hold, then there is an equilibrium in which* $q(\emptyset|sf) = q(\emptyset|fs) = 0$*.*

This result shows that if both agents are successful, then the principal truthfully reports it. If both have failed, then some tell the truth while others give no feedback. The fraction of principals who provide no feedback increases with the prior and the proportion of naive agents.

Behavior of the principal when only one of the agents has succeeded depends on condition (1), which is likely to hold if the fraction of naive agents is high and the agents' action are complements. Since in our experiment there are strong complementarities between the agents' actions, we expect this condition to hold and hence some principals with types sf and fs to give no feedback. This is exactly the type of behavior we observe in the data (see Table 6).

Therefore, we assume that condition (1) holds, in which case Bayes' rule implies that beliefs increase after positive feedback and decrease after negative feedback. Direction of change in beliefs after no feedback is ambiguous, but they decrease less than they do in private-verifiable feedback. If beliefs about the other agent is uniform, then average beliefs conditional on failure is smaller than the prior but it is not clear whether beliefs conditional on success is greater than the prior.

In summary, we have the following predictions:

Prediction D. *If lying is costly and some agents are naive, then in verifiable feedback:*

1. *If feedback is private, beliefs increase conditional on success and decrease conditional on failure. If feedback is public, beliefs decrease conditional on failure but the magnitude of change is smaller than it is under private feedback;*
2. *Beliefs increase after positive feedback and decrease after negative feedback. If feedback is private, beliefs also decrease after no feedback;*
3. *Beliefs after negative feedback are smaller than beliefs after no feedback;*
4. *If feedback is public, beliefs do not depend on the feedback provided to the other agent;*

5. *If feedback is private, all principals report truthfully to the agent who succeeds while some report truthfully and some give no feedback to the agent who fails. If feedback is public, all principals report truthfully if both agents succeed, while some tell the truth and some give no feedback if one of the agents fails.*

Our empirical findings verify prediction D.1 (see Figure 1) as well as D.2 and D.3 (see Figure 2), except that the decrease in beliefs after no feedback is not statistically significant in private feedback. Note that prediction D.3 is novel in the new model and follows from the existence of naive agents. Also note that in the data, beliefs after no feedback increase in public-verifiable feedback, which cannot be explained with our original model. In the model with lying costs, this could happen if each agent assigns a disproportionately high likelihood to the event that he has succeeded and the other has failed, i.e., agent 1 believes that the state is sf while agent 2 believes that it is fs. Finally, while prediction D.5 is verified (see Tables 5 and 6), D.4 is rejected (see Figure 3a and Table 2 columns (4) and (5)). Overall, empirical observations are very close to theoretical predictions except that in the data beliefs are somewhat more pessimistic if own positive feedback is accompanied by positive feedback to the other agent.

6.1.2 Unverifiable Feedback with Lying Cost and Naive Agents

The most significant difference between the models with and without lying costs appears under unverifiable feedback. In particular, and unlike the original model, the model with lying costs and naive agents uniquely pins down the principal's behavior under private-unverifiable feedback. If the agent is successful, the principal sends positive feedback and if he has failed, then those with high costs of lying and withholding information report truthfully, those with small costs of lying report success, and those with larger costs of lying but small costs of withholding give no feedback.

Proposition 7. *If feedback is private and unverifiable, then for any i and θ_{-i}*

$$q_i(s|s,\theta_{-i}) = 1,\ q_i(s|f,\theta_{-i}) > 0,\ q_i(\emptyset|f,\theta_{-i}) > 0,\ q_i(f|f,\theta_{-i}) > 0.$$

Equilibrium behavior under public-unverifiable feedback may not be unique. However, if v is separable, i.e., $v(a,\theta) > v(a',\theta)$ implies $v(a,\theta') > v(a',\theta')$, then the following is true.

Proposition 8. *If v is separable and feedback is public and unverifiable, then $q(ss|ss) = 1$, $q(ss|\theta) > 0$ for some $\theta \neq ss$ and $q_i(\theta|\theta) > 0$ for all θ.*

A natural extension of the private-unverifiable feedback equilibrium to public case along the lines suggested by Proposition 8 is the following: (1) Type ss sends ss; (2) Type ff sends ss, $\emptyset\emptyset$, or ff; (3) Type sf sends ss, $s\emptyset$, or sf; (4) Type fs sends ss, $\emptyset s$, or fs. This is exactly the type of behavior we observe in the data. Therefore, we assume that this is the equilibrium that our subjects play.

Using Bayes' rule to derive the beliefs, we have the following prediction.

Prediction E. *If lying is costly and some agents are naive, then in unverifiable feedback:*

1. *Beliefs conditional on success are smaller than those in private-verifiable feedback;*

2. *Beliefs after positive feedback are smaller than those in verifiable feedback;*

3. *Beliefs decrease after negative and no feedback at a magnitude similar to those in verifiable negative feedback;*

4. *Beliefs after positive feedback are smaller if the other agent receives positive feedback as well, while beliefs after negative feedback are not affected by the feedback to the other agent;*

5. *All principals report truthfully to the agent who succeeds but, to the agent who fails, some principals report success, some no information and some failure.*

Our empirical findings verify predictions E.1 (see Figure 1) and E.2. Prediction E.3 is not supported because the decrease in beliefs after negative or no feedback is smaller compared with verifiable feedback (see Figure 2). Except for a few outliers, Tables 5 and 6 give strong support to E.5. Perhaps most remarkably, item E.4 is strongly supported (see Figure 3b and Table 2 columns (4) and (5)). Note that this prediction is novel to the new model and follows from the fact that feedback (s, f) is given only by the principal who observed (s, f) whereas (s, s) is given by types (s, s), (s, f), (f, s), and (f, f). Therefore, a Bayesian agent 1 who receives feedback (s, f) is sure that he succeeded, while if he receives feedback (s, s), then he assigns a positive probability that he failed. Finally, the new model, as well as the original one, predicts the private-verifiable feedback to be the most informative mechanism, which is supported by our findings.

Overall, the model fits the data quite well, and in some respects better than the original model, but there are still two deviations of the model's predictions from what we observe in the data: (1) Agents interpret other agent's success pessimistically even in verifiable feedback; (2) Agents do not interpret negative or no feedback as pessimistically as they should, particularly in private-unverifiable feedback.

6.2 Are Agents Bayesian?

Suppose that agents know (or predict) the strategy employed by the principals in our experiment and use Bayes' rule to update their beliefs. How would their beliefs change upon observing feedback? How do actual beliefs compare with such Bayesian beliefs?

In order to answer these questions, we estimate the principals' strategy using the data in Table 5 and 6 and then use each agents' pre-feedback beliefs, the feedback they received, and Bayes' rule to calculate post-feedback beliefs.[29] Before we start presenting our findings, we should stress that we are subjecting the agents to quite a stringent test. A perfect fit between the actual and Bayesian beliefs requires not only that they use Bayes' rule correctly to update their beliefs but also that they predict the principals' strategy perfectly.

Figure 5 plots the average change in actual and Bayesian beliefs in each treatment conditional on the actual outcome of the agent. We can see that the direction of change is the same in actual and Bayesian beliefs, except under unverifiable feedback when the actual outcome is failure. Also, compared with the Bayesian case, overall information transmission is much weaker when the actual state is failure.[30]

[29] In calculating principals' strategy we eliminated some outliers in tables 5 and 6: in private-verifiable, row SF column No Info,Info; in private-unverifiable, row FS column SNo; in public-unverifiable, row SS column NoS, row FS columns SF and SNo, row SF columns FF and NoF.

[30] We should note that the scale of the graphs in 5a and 5b are different because agents update their beliefs by amounts that are much smaller than the theoretical ones. For example, under truthful feedback, Bayes' rule requires that beliefs go

[Figure 5 about here.]

Figure 6 plots average change in actual and Bayesian beliefs in response to feedback. We again see that the direction of change in beliefs is the same in actual and Bayesian beliefs (except those in public-verifiable and private-unverifiable treatments after no feedback). This figure also supports our conclusion from the previous section that agents do not interpret negative or no feedback as pessimistically as they should, especially when they are unverifiable.

[Figure 6 about here.]

Finally, we compare the change in beliefs in response to the other agent's feedback in public-unverifiable feedback. As Figure 7 shows, as long as the direction of change in the beliefs are concerned, agents on average act in a Bayesian manner. However, and as we have discovered before, they seem to interpret positive feedback to the other agent more pessimistically than is justified by Bayesian updating alone.

[Figure 7 about here.]

In Table 8, we present regression results which show that Bayesian updating plays a significant role in the formation of actual beliefs and explains about 20% of their total variation. We also see that, together with the prior, Bayesian updating explains about half of the total variation in the posterior beliefs. Furthermore, the relationship between actual and Bayesian beliefs do not depend on the feedback mechanism in a significant way.

[Table 8 about here.]

Overall, we conclude that agents' beliefs are consistent with the strategy employed by the principals and Bayesian updating, except that they respond to negative or no feedback more optimistically and interpret positive feedback to the other agent more pessimistically than they should. However, Bayesian updating does not explain the entire variation in beliefs. This could be due to agents' inability to correctly anticipate the principal's strategy, their naiveté, or other biases they suffer in processing information, such as self-serving biases and non-Bayesian social comparisons.

7 Conclusion

In this paper, we employ a theoretical model and data from a laboratory experiment to examine the role of verifiability and privacy in strategic interim performance feedback. Our baseline theoretical model predicts that information about agents' performances can be credibly revealed only when the performance information is verifiable and, furthermore, private-verifiable feedback is the most informative mechanism. These predictions are strongly supported by our empirical analysis.

up to 1 after success and down to 0 after failure, whereas in reality they go up to 0.77 and down to 0.57, respectively. This is simply because in the theoretical model beliefs refer to the probability that they have been successful in the task they have just finished, while in the experiment they measure the probability with which they believe they will be successful in the next task. We expect the latter to be strictly increasing in the former but not necessarily identical with it. Also note that numbers in Figure 5a are slightly different than those in Figure 1. This is because we had to drop a few observations for which we could not apply Bayes' rule in calculating beliefs. In order to maintain comparability between the actual and Bayesian beliefs we also dropped those observations in calculating the average change in actual beliefs. These comments also apply to the the other graphs in this section.

However, the baseline model cannot account for some interesting features of the data: (1) many principals tell the truth even when they believe this may hurt them; (2) agents do not interpret "no feedback" as pessimistically as they should; and (3) positive feedback is interpreted less optimistically if the other agent has also received positive feedback, and this effect is stronger in the case of public-unverifiable than in the case of public-verifiable feedback. We then analyze a model with individual-specific lying costs and naive agents, and show that it can account for many of these findings. We also find that while many agents do take into account the principal's strategic behavior to form beliefs in a Bayesian fashion, some are naive and act in a non-Bayesian manner, particularly when informed about other agents' feedback. From a more practical point of view, we conclude that credible communication of interim performance requires verifiability and it is best to keep feedback private.

References

Should performance reviews be fired? Knowledge@Wharton, April 27 2011. URL https://knowledge.wharton.upenn.edu/article/should-performance-reviews-be-fired/. [Online; accessed 24-August-2015].

Johannes Abeler, Anke Becker, and Armin Falk. Truth-telling: A representative assessment. *IZA Discussion Paper*, 2012.

Masaki Aoyagi. Information feedback in a dynamic tournament. *Games and Economic Behavior*, 70 (2):242–260, 2010.

Ghazala Azmat and Nagore Iriberri. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94 (7):435–452, 2010.

Ghazala Azmat and Nagore Iriberri. The provision of relative performance feedback information: An experimental analysis of performance and happiness. *CEP Discussion Paper*, 2012.

Oriana Bandiera, Valentino Larcinese, and Imran Rasul. Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on performance. *Working Paper*, May 2010.

Marco Battaglini and Uliana Makarov. Cheap talk with multiple audiences: An experimental analysis. *Working Paper*, 2010.

Volker Benndorf, Dorothea Kübler, and Hans-Theo Normann. Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *European Economic Review*, 75:43–59, 2015.

Jordi Blanes i Vidal and Mareike Nossol. Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10):1721–1736, 2011.

Andreas Blume, Douglas V DeJong, Yong-Gwan Kim, and Geoffrey B Sprinkle. Evolution of communication with partial common interest. *Games and Economic Behavior*, 37(1):79–120, 2001.

Hongbin Cai and Joseph Tao-Yi Wang. Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56:7–36, 2006.

Gary Charness and Martin Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006.

Gary Charness, David Masclet, and Marie Claire Villeval. Competitive preferences and status as an incentive: Experimental evidence. *Working Paper*, 2010.

Ying Chen. Perturbed communication games with honest senders and naive receivers. *Journal of Economic Theory*, 146(2):401–424, 2011.

Vincent Crawford. A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78:286–298, 1998.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.

Samuel A. Culbert. Get rid of the performance review! *The Wall Street Journal*, October 20 2008. URL http://www.wsj.com/articles/SB122426318874844933. [Online; accessed 16-September- 2015].

JohnW. Dickhaut, KevinA. McCabe, and Arijit Mukherji. An experimental study of strategic information transmission. *Economic Theory*, 6(3):389–403, 1995.

Mikhail Drugov, Roberto Hernán-González, Praveen Kujal, and Marta Troya Martinez. Cheap talk with two audiences: An experiment. *Working Paper*, December 2013.

Florian Ederer. Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*, 19(3):733–769, 2010.

Florian Ederer and Ernst Fehr. Deception and incentives: How dishonesty undermines effort provision. *Working Paper*, May 2009.

David Eil and Justin M Rao. The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138, 2011.

Tor Eriksson, Anders Poulsen, and Marie Claire Villeval. Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688, 2009.

Seda Ertac. Social comparisons and optimal information revelation: Theory and experiments. *Working Paper*, 2005.

Seda Ertac. Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545, 2011.

Joseph Farrell and Robert Gibbons. Cheap talk with two audiences. *The American Economic Review*, 79(5):1214–1223, 1989.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.

Urs Fischbacher and Franziska Föllmi-Heusi. Lies in disguise-an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547, 2013.

Robert Forsythe, R. Mark Isaac, and Thomas R. Palfrey. Theories and tests of "blind bidding" in sealed-bid auctions. *RAND Journal of Economics*, 20:214–238, 1989.

Robert Forsythe, Russell Lundholm, and Thomas Rietz. Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence. *Review of Financial Studies*, 12(3):481–518, 1999.

Leonie Gerhards and Neele Siemer. Private versus public feedback-the incentive effects of symbolic awards. *Working Paper*, 2014.

David Gill, Zdenka Kissová, Jaesun Lee, and Victoria L Prowse. First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Available at SSRN 2641875*, 2015.

Francesco Giovannoni and Daniel J. Seidmann. Secrecy, two-sided bias and the value of evidence. *Games and Economic Behavior*, 59:296–315, 2007.

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

Uri Gneezy, Bettina Rockenbach, and Marta Serra-Garcia. Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293 – 300, 2013.

Maria Goltsman and Gregory Pavlov. How to talk to multiple audiences. *Games and Economic Behavior*, 72:100–122, 2011.

Sanford J Grossman. The informational role of warranties and private disclosure about product quality. *Journal of law and economics*, 24(3):461–483, 1981.

Oliver Gürtler and Christine Harbring. Feedback in tournaments under commitment problems: Experimental evidence. *Journal of Economics & Management Strategy*, 19(3):771–810, 2010.

Jeanne Hagenbach and Eduardo Perez-Richet. Communication with evidence in the lab. Technical report, Mimeo, 2015.

Jeanne Hagenbach, Frédéric Koessler, and Eduardo Perez-Richet. Certifiable pre-play communication: Full disclosure. *Econometrica*, 82:1093–1131, 2014.

Erik Hoelzl and Aldo Rustichini. Overconfident: Do you put your money on it? *The Economic Journal*, 115(503):305–318, 2005.

Sjaak Hurkens and Navin Kartik. Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, 12(2):180–192, 2009.

Eric Jackson. Ten biggest mistakes bosses make in performance reviews, January 9 2012. URL http://www.forbes.com/sites/ericjackson/2012/01/09/ten-reasons-performance-reviews-are-done [Online; accessed 24-August-2015].

Ginger Zhe Jin, Michael Luca, and Daniel Martin. Is no news (perceived as) bad news? an experimental investigation of information disclosure. Technical report, National Bureau of Economic Research, 2015.

Edi Karni. A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606, 2009.

Navin Kartik. Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395, 2009.

Navin Kartik, Marco Ottaviani, and Francesco Squintani. Credulity, lies, and costly talk. *Journal of Economic theory*, 134(1):93–116, 2007.

Ronald R King and David E Wallin. Voluntary disclosures when seller's level of information is unknown. *Journal of Accounting Research*, pages 96–108, 1991.

Frédéric Koessler. Lobbying with two audiences: Public vs private certification. *Mathematical Social Sciences*, 55(3):305–314, 2008.

Camelia M Kuhnen and Agnieszka Tymula. Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113, 2012.

Ernest K. Lai, Wooyoung Lim, and Joseph Tao-yi Wang. An experimental analysis of multidimensional cheap talk. *Working Paper*, April 2015.

Clinton O Longenecker, Henry P Sims Jr, and Dennis A Gioia. Behind the mask: The politics of employee appraisal. *The Academy of Management Executive*, pages 183–193, 1987.

Jérôme Mathis. Full revelation of information in sender-receiver games of persuasion. *Journal of Economic Theory*, 143:571–584, 2008.

Paul R Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391, 1981.

Markus M Mobius, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. Managing self-confidence: Theory and experimental evidence. *NBER Working Paper*, 2011.

Alwine Mohnen and Kathrin Manthei. Is honesty the best policy? An experimental study on the honesty of feedback in employer-employee relationships. *Working Paper*, 2006.

Kevin R Murphy and Jeanette N Cleveland. *Performance appraisal: An organizational perspective.* Allyn & Bacon, 1991.

Julie Rosaz. Biased information and effort. *Economic Inquiry*, 50(2):484–501, 2012.

Santiago Sanchez-Pages and Marc Vorsatz. Enjoy the silence: an experiment on truth-telling. *Experimental Economics*, 12(2):220–241, 2009.

Mike Schraeder, J. Bret Becton, and Ron Portis. A critical examination of performance appraisals. *The Journal for Quality and Participation*, 30(1):20–25, Spring 2007.

Daniel J Seidmann and Eyal Winter. Strategic information transmission with verifiable messages. *Econometrica*, 65(1):163–169, 1997.

Joel Sobel. Giving and receiving advice. In Daron Acemoglu, Manuel Arellano, and Eddie Dekel, editors, *Advances in Economics and Econometrics*. Cambridge University Press, 2013.

Matthias Sutter. Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119(534):47–60, 2009.

Emanuel Vespa and Alistair J. Wilson. Communication with multiple senders: An experiment. *Working Paper*, April 2015.

Joseph Tao-yi Wang, Michael Spezio, and Colin F. Camerer. Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3):984–1007, 2010.

Table 1: Information Transmission in Different Treatments

Dependent var: Change in beliefs	(1) Success	(2) Failure
Private-Verifiable	-3.786	4.321
	(2.789)	(3.784)
Public-Verifiable	-5.039*	5.796
	(2.661)	(3.896)
Private-Unverifiable	-8.822***	10.883***
	(2.845)	(3.831)
Public-Unverifiable	-11.123***	10.745***
	(2.740)	(4.094)
Session	YES	YES
N	195	252
χ^2	27.145	22.272

GLS Regressions, standard errors in parentheses

Baseline is truthful feedback

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2: Impact of Feedback on Beliefs

	(1) Change in belief (Positive feedback)	(2) Change in belief (Negative Feedback)	(3) Change in belief (No Feedback)	(4) Change in belief (Own positive feedback)	(5) Change in belief (Own negative feedback)
Private-Verifiable	-2.318	2.391	-1.079	-0.759	2.926
	(2.945)	(5.519)	(4.408)	(2.842)	(5.460)
Private-Unverifiable	-9.383***	12.224**	-0.449	-7.697***	12.116**
	(2.599)	(5.552)	(5.828)	(2.479)	(5.445)
Public-Verifiable	-5.017*	1.673	3.055		
	(2.879)	(5.063)	(4.617)		
Public-Unverifiable	-6.500**	4.101			
	(2.778)	(6.176)			
PUBLIC-VERIFIABLE					
Other positive feedback				-6.090*	-2.029
				(3.565)	(5.992)
Other negative feedback				-1.170	7.033
				(3.524)	(7.129)
PUBLIC-UNVERIFIABLE					
Other positive feedback				-6.909**	-5.040
				(3.143)	(8.665)
Other negative feedback				3.376	12.992
				(5.078)	(9.353)
Session	YES	YES	YES	YES	YES
N	207	149	91	207	149
χ^2	21.788	15.753	.	23.927	18.988

Standard errors in parentheses

GLS Regressions for Different Feedback Mechanisms

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 3: Impact of Beliefs on Performance

	(1) Post-feedback Performance
Pre-feedback Performance	0.442***
	(0.049)
Change in beliefs	0.013*
	(0.008)
Session	YES
N	392
χ^2	130.599

GLS regressions, standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4: Feedback to Individual Agents under Different Treatments

	Private-Verifiable (message)			Public-Verifiable (message)			Private-Unverifiable (message)				Public-Unverifiable (message)			
Actual	Info	No Info	Total	Info	No Info	Total	S	F	No	Total	S	F	No	Total
S	31	1	32	31	7	38	30	0	2	32	29	2	3	34
	(96.88)	(3.13)	(39.02)	(81.58)	(18.42)	(46.34)	(93.75)	(0.00)	(6.25)	(39.02)	(85.29)	(5.88)	(8.82)	(45.95)
F	22	28	50	25	19	44	27	17	6	50	15	10	15	40
	(44.00)	(56.00)	(60.98)	(56.82)	(43.18)	(53.66)	(54.00)	(34.00)	(12.00)	(60.98)	(37.50)	(25.00)	(37.50)	(54.05)
Total	53	29		56	26		57	17	8		44	12	18	
	(64.63)	(35.37)		(68.29)	(31.71)		(69.51)	(20.73)	(9.76)		(59.46)	(16.22)	(24.32)	

Number of agents receiving each message type, percentages in parentheses.

S= *Successful*, F= *Failed* and No= *No message.*

Table 5: Feedback to Both Agents in Private Feedback

	Private-Verifiable (message)				Private-Unverifiable (message)								
Actual	Both Info	Info,No Info	No Info, Info	Both No Info	SS	SF	FS	FF	No-No	SNo	NoS	FNo	NoF
SS	5	0	0	0	6	0	0	0	0	0	0	0	0
	(100.00)	(0.00)	(0.00)	(0.00)	(100.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SF	7	4	1	0	7	1	0	0	0	1	0	0	0
	(58.33)	(33.33)	(8.33)	(0.00)	(77.78)	(11.11)	(0.00)	(0.00)	(0.00)	(11.11)	(0.00)	(0.00)	(0.00)
FS	5	0	5	0	4	0	5	0	0	2	0	0	0
	(50.00)	(0.00)	(50.00)	(0.00)	(36.36)	(0.00)	(45.45)	(0.00)	(0.00)	(18.18)	(0.00)	(0.00)	(0.00)
FF	4	1	0	9	5	1	0	5	1	1	2	0	0
	(28.57)	(7.14)	(0.00)	(64.29)	(33.33)	(6.67)	(0.00)	(33.33)	(6.67)	(6.67)	(13.33)	(0.00)	(0.00)

Number of agents receiving each message type, percentages in parentheses

The abbreviations in message part represents message pairs.

S= *Successful*, F= *Failed* and No= *No message.* SNo refers to (*Successful, No message*) message pair.

Table 6: Feedback to Both Agents in Public Feedback

	Public-Verifiable (message)		Public-Unverifiable (message)								
Actual	Both Info	Both No Info	SS	SF	FS	FF	No-No	SNo	NoS	FNo	NoF
SS	8	0	3	0	0	0	0	0	1	0	0
	(100.00)	(0.00)	(75.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(25.00)	(0.00)	(0.00)
SF	6	3	4	3	0	1	0	4	0	0	1
	(66.67)	(33.33)	(30.77)	(23.08)	(0.00)	(7.69)	(0.00)	(30.77)	(0.00)	(0.00)	(7.69)
FS	9	4	7	1	3	0	0	1	1	0	0
	(69.23)	(30.77)	(53.85)	(7.69)	(23.08)	(0.00)	(0.00)	(7.69)	(7.69)	(0.00)	(0.00)
FF	5	6	1	0	0	1	5	0	0	0	0
	(45.45)	(54.55)	(14.29)	(0.00)	(0.00)	(14.29)	(71.43)	(0.00)	(0.00)	(0.00)	(0.00)

Number of agents receiving each message type, percentages in parentheses

The abbreviations in message part represents message pairs.

S= *Successful*, F= *Failed* and No= *No message.* SNo refers to (*Successful, No message*) message pair.

Table 7: Principal's Expectations in Public Feedback-Extended

	Own Positive Feedback			Own Negative Feedback		
	(1)	(2)	(3)	(4)	(5)	(6)
	Guess	Guess (Public-Verifiable)	Guess (Public-Unverifiable)	Guess	Guess (Public-Verifiable)	Guess (Public-Unverifiable)
Other Positive Feedback	-3.151	3.708	-10.195**	-1.550	-0.333	-3.323
	(3.198)	(5.692)	(4.176)	(8.969)	(7.903)	(17.649)
Public-Verifiable	10.188***			4.722		
	(2.960)			(3.169)		
N	75	31	44	37	25	12
χ^2	14.735	0.424	5.959	2.250	0.002	0.035

Standard errors in parentheses

GLS Regressions for Different Feedback Mechanisms

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: Post-Feedback Beliefs

	(1) Post-Feedback Belief	(2) Post-Feedback Belief	(3) Post-Feedback Belief
Bayesian Belief	0.198***	0.154***	0.156***
	(0.024)	(0.022)	(0.022)
Pre-Feedback Belief		0.655***	0.649***
		(0.057)	(0.059)
Private-Verifiable			1.723
			(2.724)
Public-Verifiable			-0.433
			(3.028)
Private-Unverifiable			0.362
			(2.819)
Public-Unverifiable			-1.278
			(2.749)
N	386	386	386
χ^2	67.137	213.684	232.519
R^2	0.205	0.544	0.545

GLS Regressions, standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure 1: Information Transmission in Different Treatments

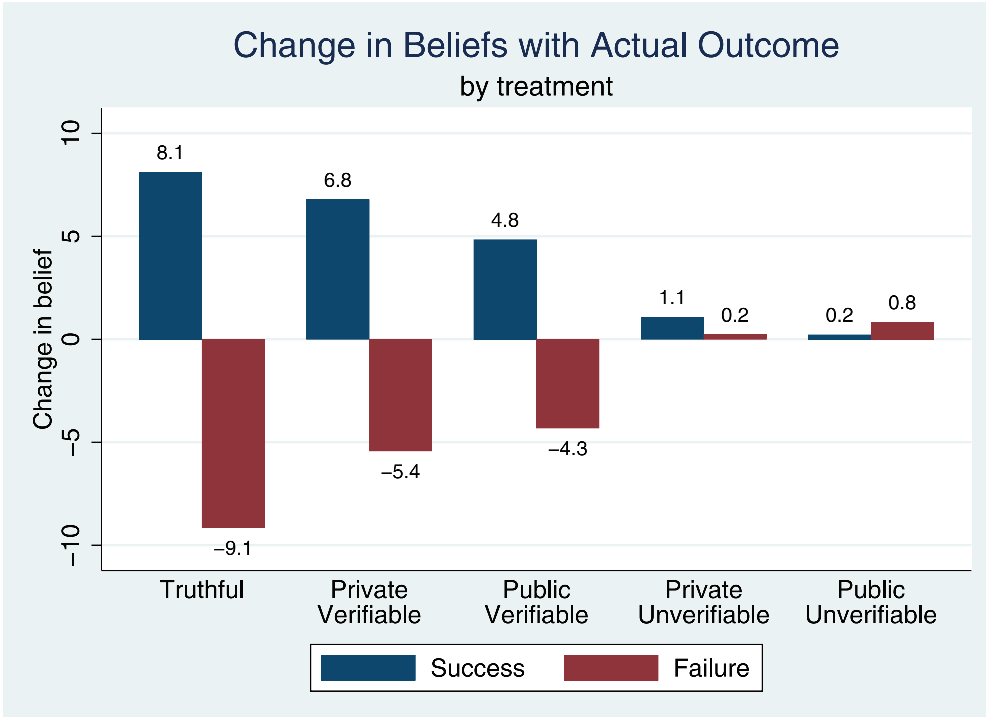


Figure 2: Impact of Feedback on Beliefs

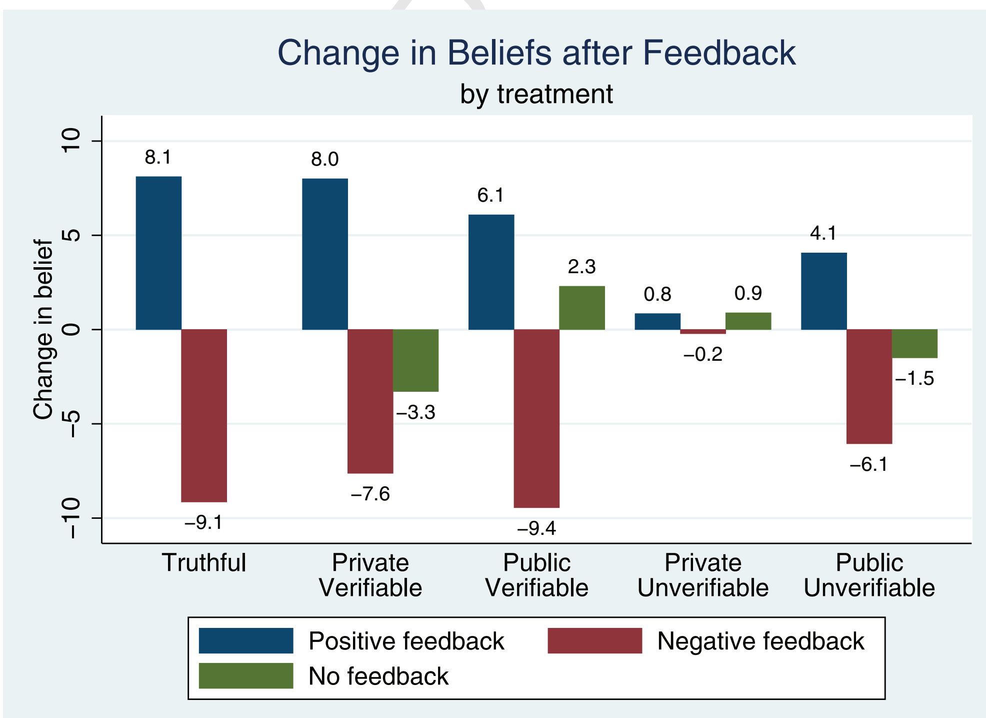
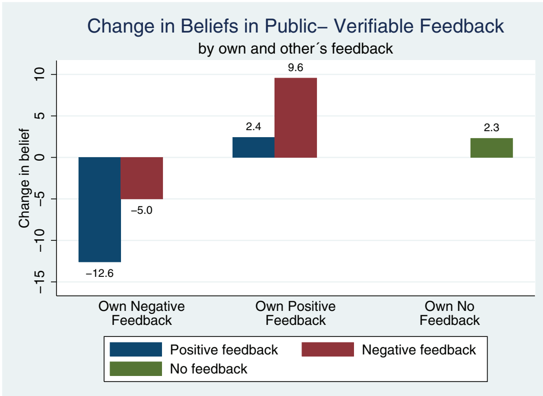


Figure 3: Change in Beliefs in Public Feedback

(a) Public-Verifiable Feedback



(b) Public-Unverifiable Feedback

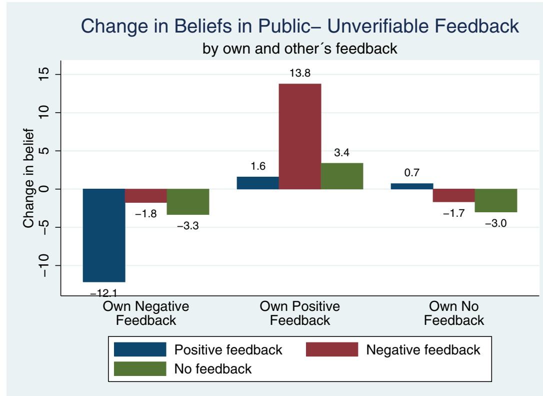


Figure 4: Difference between Principal's Guess and Actual Belief

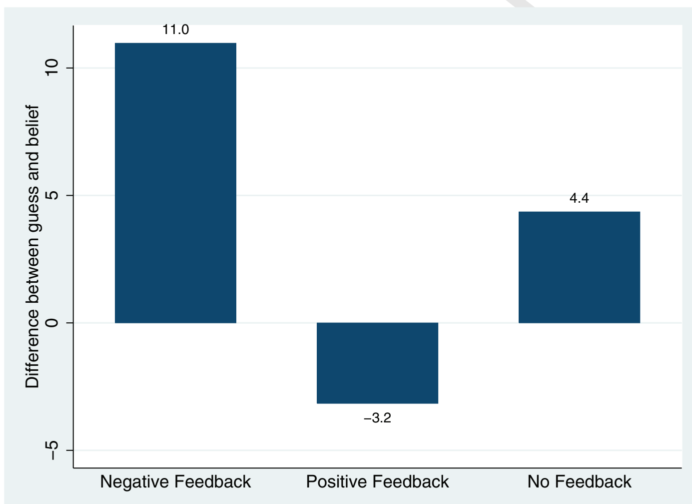


Figure 5: Information Transmission: Actual vs. Bayesian

(a) Actual Beliefs (b) Bayesian Beliefs

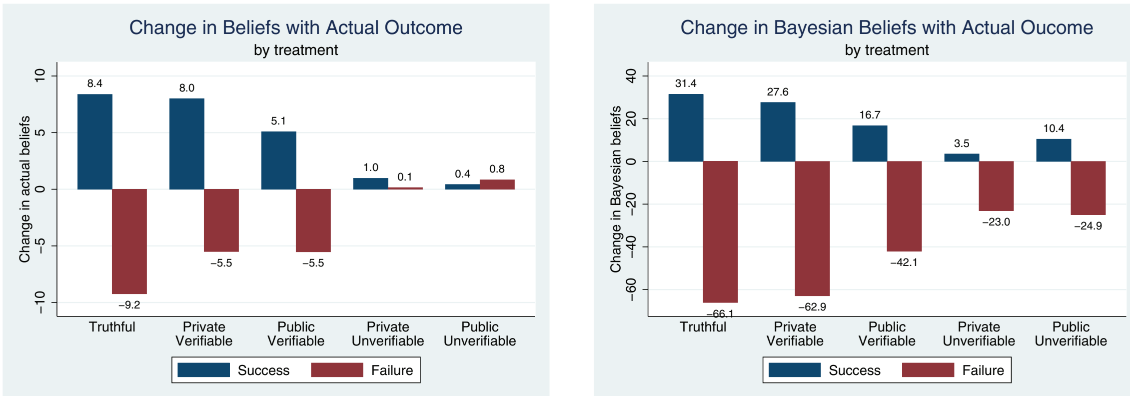


Figure 6: Change in Beliefs with Feedback: Actual vs. Bayesian

(a) Actual Beliefs (b) Bayesian Beliefs

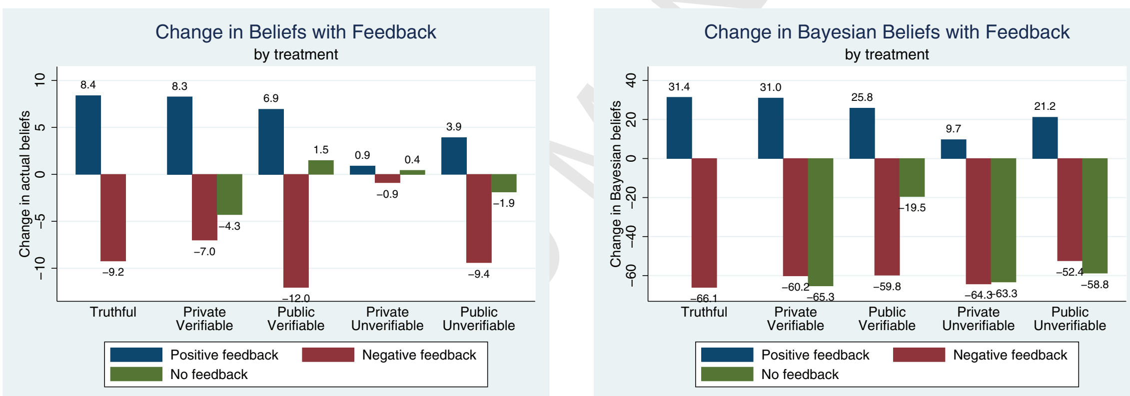


Figure 7: Change in Beliefs with Other's Feedback: Actual vs. Bayesian

(a) Actual Beliefs (b) Bayesian Beliefs