

Ertaç, Seda; Koçkesen, Levent; Özdemir, Duygu

Working Paper

The role of verifiability and privacy in the strategic provision of performance feedback: Theory and experimental evidence

Working Paper, No. 1524

Provided in Cooperation with:

Koç University - TÜSİAD Economic Research Forum, Istanbul

Suggested Citation: Ertaç, Seda; Koçkesen, Levent; Özdemir, Duygu (2015) : The role of verifiability and privacy in the strategic provision of performance feedback: Theory and experimental evidence, Working Paper, No. 1524, Koç University-TÜSİAD Economic Research Forum (ERF), Istanbul

This Version is available at:

<https://hdl.handle.net/10419/129364>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

**THE ROLE OF VERIFIABILITY AND PRIVACY IN
THE STRATEGIC PROVISION OF PERFORMANCE
FEEDBACK: THEORY AND EXPERIMENTAL
EVIDENCE**

Seda Ertaç
Levent Koçkesen
Duygu Özdemir

Working Paper 1524
November 2015

This Working Paper is issued under the supervision of the ERF Directorate. Any opinions expressed here are those of the author(s) and not those of the Koç University-TÜSİAD Economic Research Forum. It is circulated for discussion and comment purposes and has not been subject to review by referees.

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
Rumelifeneri Yolu 34450 Sarıyer/Istanbul

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

October 2015

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 in addition to being naive, some agents also suffer from self-serving biases and engage in non-Bayesian social comparisons in their interpretation of performance feedback.

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

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

*Seda Ertac, Department of Economics, Koç University, Rumelifeneri Yolu Sariyer, Istanbul 34450 Turkey. E-mail: ertac@ku.edu.tr. Levent Koçkesen, Department of Economics, Koç University, Rumelifeneri Yolu Sariyer, Istanbul 34450 Turkey. E-mail: lkockesen@ku.edu.tr. Duygu Ozdemir, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom. E-mail: dozdem@essex.ac.uk. Ertac thanks the Scientific and Technological Research Council of Turkey (TUBITAK) for generous financial support (Grant No. 111K444) for the experiments in this paper.

[†]Corresponding author.

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.² Forced ranking systems may overcome this deficiency but they cause problems of their own, potentially undermining employee confidence and motivation.³

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.⁴ 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 model 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.⁵ 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

¹One 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).

²See 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.

³One commentator observes that “[i]f you look at complaints that people have about performance appraisals, it is that you always want to give everyone an A-plus. Forced rankings don’t allow that to happen. But the problem is they foster a lot of competition among peers.” (Wharton management professor Matthew Bidwell quoted in “Should Performance Reviews Be Fired?”). Many large companies, such as Microsoft and General Electric, have recently abandoned the forced ranking system (also called stack ranking, or “rank and yank”), while Amazon’s continued use of it has been subjected to recent media scrutiny (see Eichenwald (2012) on Microsoft and Kantor and Streitfeld (2015) on Amazon).

⁴In 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)).

⁵This threshold plays the role of chance or other unpredictable exogenous factors such as market conditions, organizational standards, etc.

perform the task again in the second period or not.⁶

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 obtains 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 achieves 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.⁷

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. In some situations, however, it might be possible to produce evidence of performance, perhaps at some cost, and turn unverifiable feedback into verifiable. This gives organizations some discretion regarding the level of verifiability of performance feedback. Similarly, 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, and this is usually a choice variable for organizations as well. 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 degrees of 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 3 we describe our experimental design and in Section 4 we analyze our baseline model and derive several theoretical predictions. We show that, in private-verifiable feedback, unique equi-

⁶More 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. If their stated belief is larger than a randomly drawn number, they get to perform in the second-period task, otherwise they receive a payoff determined by a chance-based lottery. We elaborate further on belief elicitation in Section 3.

⁷The 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.

librium outcome is full information revelation, while in private-unverifiable feedback information cannot be revealed at all. While these results on private feedback are simple applications of the ideas introduced in Grossman (1981), Milgrom (1981), and Crawford and Sobel (1982), the public feedback game has multiple receivers and a multi-dimensional state and our theoretical model provides some novel findings for this setting.

The crucial aspect of our model is that the principal would like to induce high beliefs on both agents but, more critically, might prefer to induce moderate beliefs on both agents rather than beliefs that are extreme opposites, such as one agent having a very high and the other having a very low self-confidence. Therefore, she may choose to give no information and induce moderately lower beliefs on both agents rather than informing one agent that he was successful and the other that he has failed. Indeed, we show that in addition to a fully informative equilibrium, the public-verifiable feedback game has also a partially informative equilibrium. In the public-unverifiable game, in contrast, there is always a completely uninformative equilibrium, but there may also be a partially informative one.

In sum, our main theoretical 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 these predictions. 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.

There is also evidence that, when feedback is public, agents’ beliefs about their likelihood of success are influenced by the feedback provided to the other agent. More precisely, they become more optimistic if the other agent received negative feedback and less optimistic if the other agent received positive feedback. We further find that this effect is significant only when own feedback is positive and stronger in public-unverifiable than in 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 baseline 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.⁸ 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, they give bad news even though they actually believe that it will be interpreted as such. Since some of them still choose to tell the truth in case of a bad outcome, we conclude that lying imposes some costs and these costs are individual specific.

⁸If messages have no intrinsic meaning, then the equilibrium strategy of the principal could be anything as long as it does not contain any information about performance. However, since messages have natural interpretations in our experiment, we argue that the focal equilibrium in unverifiable feedback mechanisms is the one in which the principal always gives positive feedback.

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 new 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 indicate 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 more optimistic compared with the Bayesian benchmark.

Overall, our results suggest that individuals incur costs from lying and these costs differ across individuals. Also, some agents seem to be naive, or suffer from a self-serving bias, and engage in non-Bayesian social comparisons in their interpretation of performance feedback.

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. As such, it lies at the intersection of three main bodies of literature: (1) theoretical studies of strategic communication; (2) experimental studies of strategic communication; and (3) theoretical and empirical studies of performance feedback.

Verifiable feedback mechanisms involve a strategic communication phase that is known as a “persuasion” game in the literature, pioneered by Grossman (1981) and Milgrom (1981), while unverifiable feedback mechanisms involve what is known as a “cheap talk” game, first studied by Crawford and Sobel (1982). These models and the main results have been later generalized and extended in several directions.⁹ 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 communi-

⁹For the literature on persuasion games see, for example, Seidmann and Winter (1997) on more general sufficient conditions for the existence and uniqueness of a fully revealing equilibrium, Mathis (2008) on partial verifiability, Giovannoni and Seidmann (2007) on the finite type case, and more recently, Hagenbach et al. (2014) on sufficient conditions for the existence of fully informative equilibria in a more general environment of Bayesian games augmented by a pre-game communication phase. The basic cheap talk model Crawford and Sobel (1982) has also been extended in many directions, including multidimensional states of the world, multiple senders, multiple receivers, and multiple rounds of communication. See Sobel (2013) for a recent survey of this large literature.

cation. 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).¹⁰

There is also some theoretical work on 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.

On the experimental side, Dickhaut et al. (1995) and Cai and Wang (2006) find support for the qualitative predictions of the basic model of Crawford and Sobel (1982). Similar to our findings, Cai and Wang (2006) also 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.¹¹ Battaglini and Makarov (2010) find overall support for the predictions of Farrell and Gibbons (1989) and also report evidence of over-communication.¹² 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.¹³

There are only a few experimental studies of strategic communication of verifiable information. Forsythe et al. (1989) study a game where sellers are better informed about the good than the buyers and have the option to reveal this information. They find that full information revelation is achieved, but only as subjects become more sophisticated over repeated rounds of play. Benndorf et al. (2013) run a labor-market experiment with a lemons structure where workers can reveal their productivity and find that revelation takes place less frequently than predicted in equilibrium.

All of these contributions study either cheap talk or persuasion games, usually by varying the pref-

¹⁰We 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.

¹¹Cai 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.

¹²In the same vein, 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 heterogeneous 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.

¹³See 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).

erences of the receivers (or senders) across treatments. Our work differs from them because we take preferences as fixed and vary both the verifiability of information and the audience. Furthermore, our experimental design involves a real effort performance feedback setting while the previous literature usually employed a neutral frame.

Previous theoretical work on performance feedback generally used principal-agent models to study optimal feedback mechanisms under the assumption of truthful feedback (see, for example, Ertac (2005), Ederer (2010), Aoyagi (2010)). In contrast, we focus on the equilibrium properties of games induced by strategic performance feedback. On the experimental side, Ertac (2011) finds that in response to negative feedback, subjects with high initial self-confidence are more optimistic than Bayesian in their belief updating, whereas initially underconfident subjects are more pessimistic. Eil and Rao (2011) shows that people are more likely to update their beliefs under positive feedback than negative in a relative-performance setting. In addition to the asymmetry in belief updating, Mobius et al. (2011) also documents conservatism, such that people respond little to both positive and negative signals when the context is ego-relevant. Our finding that agents respond to feedback more optimistically than they should is in line with and complements these findings, but we also find that agents take into account the strategic incentives of the principal in their belief updating.¹⁴

More relevant for our work is the scant experimental literature on strategic performance feedback. Ederer and Fehr (2009) study the effect of private unverifiable feedback on effort in a dynamic tournament setting. They find that feedback is partially informative and, although agents discount such unverifiable feedback, they still respond to it. Gürtler and Harbring (2010) study the effect of public verifiable feedback on effort in a tournament where the theory predicts full revelation. Experimental findings support this prediction qualitatively but show that full revelation does not happen. They also find that agents respond to no feedback as bad news, but the effect on effort is not as strong as the theoretical prediction. Rosaz (2012) studies a one principal/one agent setting in which the principal has limited ability to lie. She finds that the principal indeed manipulates the feedback but the agent increases effort in response. Finally, Mohnen and Manthei (2006) considers a one principal/one agent setting with unverifiable feedback and induced effort. They find that some principals tell the truth (and suffer from cost of lying) but deception is also widespread. They also find that agents' effort levels react to feedback as if it is informative.

Our work is distinct from these studies along several dimensions: (1) We study a two-agent non-tournament setting where information about the other agent is strategically irrelevant. This allows us to test the audience effect in isolation from other strategic considerations. (2) We vary treatments along both the audience and the verifiability dimensions whereas the previous work did it only in one of these dimensions. (3) We focus on the strategic communication aspect and, using a recent method developed by Karni (2009) and Mobius et al. (2011), elicit beliefs in an incentive compatible way that is also robust to risk aversion. Previous work, in contrast, has focused on the effect of performance feedback on effort and therefore is not able to isolate the strategic communication aspect in a clean manner. (4) We use a real effort setting while the previous work induced ability and effort hypothetically. This obviously has a cost in terms of control over unobservables and makes the match with

¹⁴Besides these studies, which focus more on beliefs and self-confidence, there are many studies that explore the effects of *truthful* performance feedback on future performance. Overall evidence in this regard is mixed. See Eriksson et al. (2009), Azmat and Iriberry (2012), Azmat and Iriberry (2010), Bandiera et al. (2010), Blanes i Vidal and Nossol (2011), Kuhnen and Tymula (2012), Charness et al. (2010), and Gerhards and Siemer (2014).

theory more difficult, but creates an ego-relevant environment and increases the external validity of our results. 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.

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 5 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 10 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 1 principal and 2 agents, are formed. We use a “strangers” matching protocol, where new groups are randomly formed at the start of every period.

For participants in the role of agents, we use two different real-effort tasks: an addition task and a verbal task.¹⁵ The verbal task consists of general knowledge questions as well as verbal classification and number-letter matching questions. The addition task involves adding 4 or 5 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 (120 sec.). At the end of each round, the number of correct answers is compared to a “target score”, randomly determined for that specific period.¹⁶ 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,

¹⁵Details of the tasks are provided in Appendix B.

¹⁶The 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.

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 not do the task again, and 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 the next part.
4. Post-feedback beliefs: After seeing the message (or no message), subjects are allowed to update their previously reported beliefs.¹⁷
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. However, the principal does not have to report the actual outcome. The principal

¹⁷At this stage, the subjects can see their previously reported beliefs on the screen.

can choose to report success or failure, and could be lying. In addition, she has an option to send no message, withholding the information. This is intended to capture the differential effects of cost of lying versus cost of not providing any information.

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 5 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.

In addition to observing true outcomes and choosing what to reveal to their agents, principals are asked to guess agents' post-feedback reported minimum probability of chance at which they prefer the chance mechanism. That is, principals are asked to guess the posterior belief of each of their agents after receiving the feedback. These beliefs are incentivized by paying 10 ECU for each correct estimate.¹⁸ This is done in order to get better insight into the feedback strategy employed by the principals.

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 can meet or exceed the target in the first round, and 100 ECU, if the agent fails in the first round.¹⁹

For the second round, if the agent prefers to do the task rather than taking the chance mechanism, the payoff depends on whether he succeeds or fails in the second round exactly as in the first round. If, however, the agent chooses 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:

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

¹⁸If the guessed amount is in the ± 5 interval of agent's actual decision, the estimate is regarded as correct.

¹⁹1 ECU corresponded to 0.06 Turkish Liras at the time of the experiment.

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, and 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 round t , while e_{it} represents the entry of agent i to the performance stage (as opposed to taking the chance mechanism) in round t . 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 in the second round, both agents must end up doing the task. Therefore, because of the complementarity, the aim of the principal is to (1) convince both agents that they have high returns from the performance mechanism in the second round so that both enter, and (2) to maximize the post-feedback performance of the worse-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.²⁰ In order to mitigate potential order effects, we used 6 different configurations that differ in the sequence of treatments.²¹ Sessions lasted about 50 minutes, and subjects earned on average 20,23 TL, including a show-up fee.²²

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 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 $s > f$.

²⁰See Section B.3 in Appendix B for the survey questions.

²¹The configurations were as follows: TVU, TUV, VUT, UVU, VTU and UTV, where T, V and U correspond to Truthful, Verifiable and Unverifiable feedback mechanisms, respectively.

²²Earnings ranged between 15 TL and 28 TL, including the show-up fee.

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.²³ 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 , her 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 .²⁴ 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:

1. Private Feedback: A pure strategy of the principal is a pair of functions $\rho = (\rho_1, \rho_2)$, where $\rho_i : \Theta^2 \rightarrow M_i(\theta)$. 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 $\rho : \Theta \rightarrow M(\theta)$. 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

²³In the experiment $A_i = [0, 1]$.

²⁴It can be shown that $a_i^*(\mu_i)$ is strictly increasing in μ_i if u_i has increasing differences in (a_i, θ_i) and $\partial u_i / \partial a_i$ is strictly increasing in θ_i . For details of this argument see Van Zandt and Vives (2007).

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.²⁵ 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-Unverifiable; (2) Private-Verifiable; (3) Public-Unverifiable; (4) Public-Verifiable.

4.1 Private 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 $\rho_i(\theta_i, \theta_{-i}) = r_i$ implies that $\mu_i(r_i)$ puts probability one on θ_i , i.e., each agent can infer his type from the principal's report. An equilibrium is completely uninformative if $\mu_i(\rho_i(\theta_i, \theta_{-i})) = \mu_i(\rho_i(\theta'_i, \theta_{-i}))$ for all i, θ_i, θ'_i , and θ_{-i} , i.e., agents learn nothing about their own type in equilibrium.²⁶

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 of Proposition 1 is very easy.²⁷ 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) principal's payoff is monotonic in each agent's beliefs.²⁸

We have the following prediction:

Prediction 1. *In private-verifiable feedback:*

1. *Principal reports truthfully if the agent has succeeded and either reports truthfully or reports no information if the agent has failed;*
2. *Beliefs increase after positive feedback and decrease after negative or no feedback;*
3. *Beliefs increase conditional on success and decrease conditional on failure.*

If feedback is unverifiable, we arrive at exactly the opposite conclusion.

²⁵See Fudenberg and Tirole (1991) for the definition of perfect Bayesian equilibrium. Note that since our game has only two stages, perfect Bayesian equilibrium and sequential equilibrium are equivalent solution concepts.

²⁶Note that this definition does not rule out receiving information about the other agent's type in equilibrium. For example, it is possible that $\rho_1(s, f) = \rho_1(f, f) = f$ and $\rho_1(s, s) = \rho_1(f, s) = s$ so that agent 1 learns nothing about his own type but learns agent 2's type perfectly.

²⁷All proofs are relegated to Appendix A.

²⁸See Milgrom (1981) and Grossman (1981) for early results on this issue and Seidmann and Winter (1997) and Hagenbach et al. (2014) for more recent and general analyses.

Proposition 2. *If feedback is private and unverifiable 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.

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. For example, there is an equilibrium in which all types report success, beliefs after positive feedback are the same as the prior, while beliefs after any other report are weakly worse than the prior. There is also an equilibrium in which each type of the principal randomizes (with the same probabilities) among the three types of feedback and beliefs are equal to the prior after each feedback.

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 observed success always reports success. This implies that, in equilibrium, the principal who observed failure must also report success. This leads to the following prediction.

Prediction 2. *In private-unverifiable feedback:*

1. *Principal always provides positive feedback;*
2. *Beliefs do not change in response to positive feedback and decrease or stay the same after negative or no feedback;*
3. *Beliefs do not change conditional on actual state.*

4.2 Public Feedback

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 the private and public feedback for equilibrium analysis.

If feedback is verifiable, full information revelation is an equilibrium but, in contrast to private-verifiable feedback, there is also a partially informative equilibrium.

Proposition 3. *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 an equilibrium with full information 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.²⁹

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

²⁹Results 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.

$\mu_i^i(\theta) = 0$ otherwise, $\mu_i^i(\emptyset) = 1/3$, $\alpha_i(\theta) = \theta_i$, $\alpha_i(\emptyset) = 3$. Clearly, types (s, s) and (f, f) have no profitable deviations. If type (s, f) deviates and reports (s, f) , she receives a payoff of $7 \times 1 \times w(1) = 7w(1)$ while her payoff to \emptyset is given by $3 \times 3 \times w(1) = 9w(1) > 7w(1)$. 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.³⁰

Bayes' rule implies that beliefs as a function of feedback are given by $\mu_i(s, m_{-i}) = 1$, $\mu_i(f, m_{-i}) = 0$, and $\mu_i(\emptyset) = 0$ in the fully informative equilibrium. In the partially informative equilibrium, belief after no feedback is $\mu_i(\emptyset) = p(sf) / (1 - p(ss))$, which is higher than it is in the fully informative equilibrium, but still smaller than the prior.³¹ Conditional on the actual state, beliefs are $\mu_i(\theta_i = s) = 1$ and $\mu_i(\theta_i = f) = 0$ in the fully informative and

$$\begin{aligned}\mu(\theta_i = s) &= \frac{p(ss)}{p(ss) + p(sf)} + \frac{p(sf)}{p(ss) + p(sf)} \frac{p(sf)}{1 - p(ss)} \\ \mu_i(\theta_i = f) &= \frac{p(sf)}{1 - p(ss)}\end{aligned}$$

in the partially informative equilibrium. Note that beliefs after success (failure) are smaller (larger) in the partially informative equilibrium than those in the fully informative equilibrium, but they are still higher (lower) than the prior. Therefore we have the following prediction:

Prediction 3. *In public-verifiable feedback:*

1. *If both agents are successful, then principal reports truthfully. If one of the agents fails, then principal reports either truthfully or gives no information;*
2. *Beliefs increase after positive feedback and decrease after negative or no feedback;*
3. *Beliefs increase conditional on success and decrease conditional on failure;*
4. *Beliefs do not depend on the feedback provided to the other agent.*

If feedback is unverifiable, 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

³⁰This property follows more generally if $v(a, \theta)$ is symmetric, concave, and strictly supermodular in a . Proof is simple: For any $a \neq a'$,

$$v\left(\frac{a+a'}{2}, \frac{a+a'}{2}, \theta\right) \geq \frac{1}{2}v(a, a, \theta) + \frac{1}{2}v(a', a', \theta) > \frac{1}{2}v(a, a', \theta) + \frac{1}{2}v(a', a, \theta) = v(a, a', \theta)$$

where the first inequality follows from concavity, the second from supermodularity, and the equality from symmetry.

³¹In order to simplify notation we sometimes denote a state or a message without a comma separating its components, e.g., as sf instead of (s, f) and $s\emptyset$ instead of (s, \emptyset) . Also, from now on $p(\theta)$ will denote the prior probability assigned to state θ .

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.³²

As before, in unverifiable feedback, messages are completely arbitrary and hence we do not have a precise prediction regarding the principal's strategy and the agents' beliefs. However, it is again plausible that type (s, s) always reports (s, s) , which implies that, in equilibrium, all other types also report (s, s) .

We have the following prediction:

Prediction 4. *In public-unverifiable feedback:*

1. *Principal always provides positive feedback to both agents;*
2. *Beliefs do not change in response to positive feedback and decrease or stay the same after negative or no feedback;*
3. *Beliefs do not change conditional on actual state.*

The above analysis also implies that

Prediction 5. *Private-verifiable feedback is the most informative mechanism.*

5 Results

The main focus of our study is whether performance feedback is informative and whether that 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.³³ 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,

³²If we change principal's payoff function to $v(a, \theta) = a_1 + a_2$ in Example 1, then there is a partially informative equilibrium given by $\rho(s, s) = \rho(f, f) = \emptyset$, $\rho(s, f) = (s, f)$, $\rho(f, s) = (f, s)$, $\mu_i^i(\emptyset) = 1/2$, $\mu_i^i(s, \theta_{-i}) = 1$, $\mu_i^i(f, \theta_{-i}) = 0$, $\alpha_i(\emptyset) = 4$, $\alpha_i(s, \theta_{-i}) = 7$, $\alpha_i(f, \theta_{-i}) = 1$.

³³The 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).

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.³⁴

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. Our model in Section 4 predicts that there will be information transmission in verifiable feedback and no information transmission in unverifiable feedback cases, and this is exactly what Figure 1 shows.³⁵ 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.³⁶ Likewise, we find 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 10%). 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).³⁷ Therefore, the data confirm the theoretical predictions 1 through 4 in Section 4 with regard to information transmission (see item 3 in each of those predictions). The model also predicts that private-verifiable feedback will be the most informative mechanism (Prediction 5) and Figure 1 and Table 1 show that it is indeed the closest mechanism to truthful feedback. Therefore, in terms of predicting the extent of information transmission, our model does a very good job.

³⁴In 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 Koc University students. Related regressions are available upon request

³⁵In Figure 1 VPr stands for verifiable-private, VPu for verifiable-public, UPr for unverifiable-private, and UPu for unverifiable-public treatments. From now on we will use these abbreviations in presenting our results.

³⁶The p -values of Wilcoxon signed rank test for the hypothesis of a zero change in beliefs in each treatment is as follows: Truthful with $p = 0.0002$, verifiable-private with $p = 0.0004$, verifiable-public with $p = 0.0129$, unverifiable-private with $p = 0.1101$, unverifiable-public with $p = 0.7872$ for success; truthful with $p = 0.0002$, verifiable-private with $p = 0.0228$, verifiable-public with $p = 0.0167$, unverifiable-private with $p = 0.6956$, unverifiable-public with $p = 0.1476$ for failure cases.

³⁷Note 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.

Figure 1: Information Transmission in Different Treatments

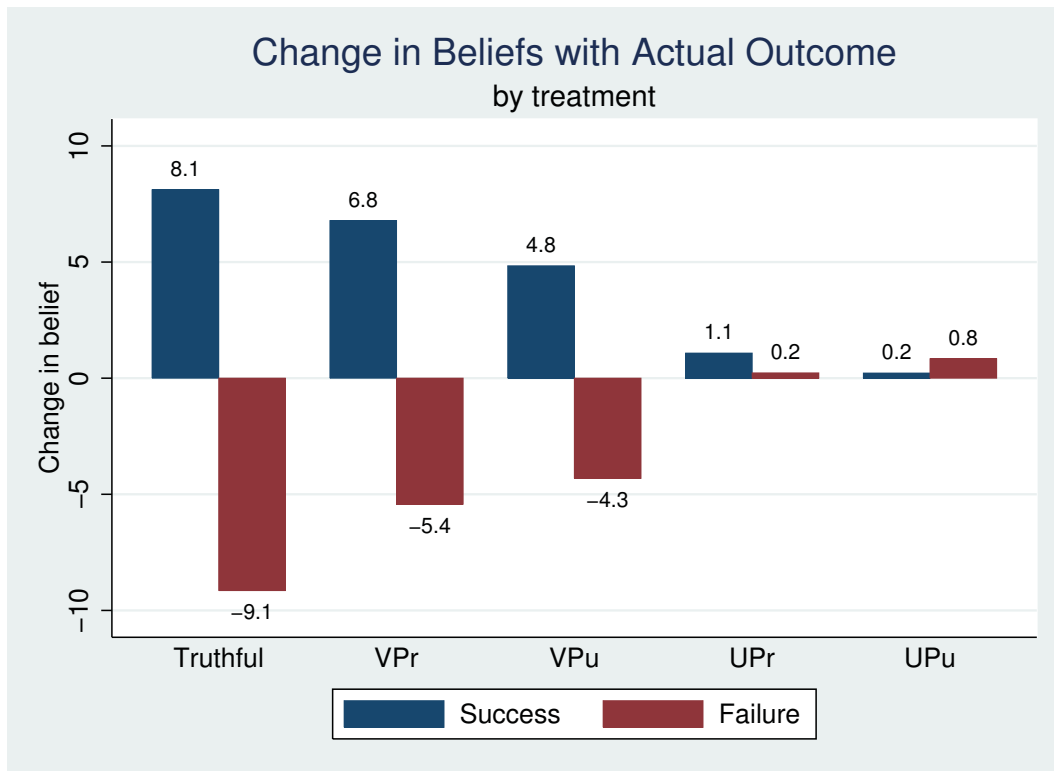


Table 1: Information Transmission in Different Treatments

Dependent var:	(1)	(2)
Change in beliefs	Success	Failure
VPr	-3.786 (2.789)	4.321 (3.784)
VPu	-5.039* (2.661)	5.796 (3.896)
UPr	-8.822*** (2.845)	10.88*** (3.831)
UPu	-11.12*** (2.740)	10.74*** (4.094)
Session	YES	YES
N	195	252
χ^2	27.15	22.27

GLS Regressions, standard errors in parentheses

Baseline is truthful feedback

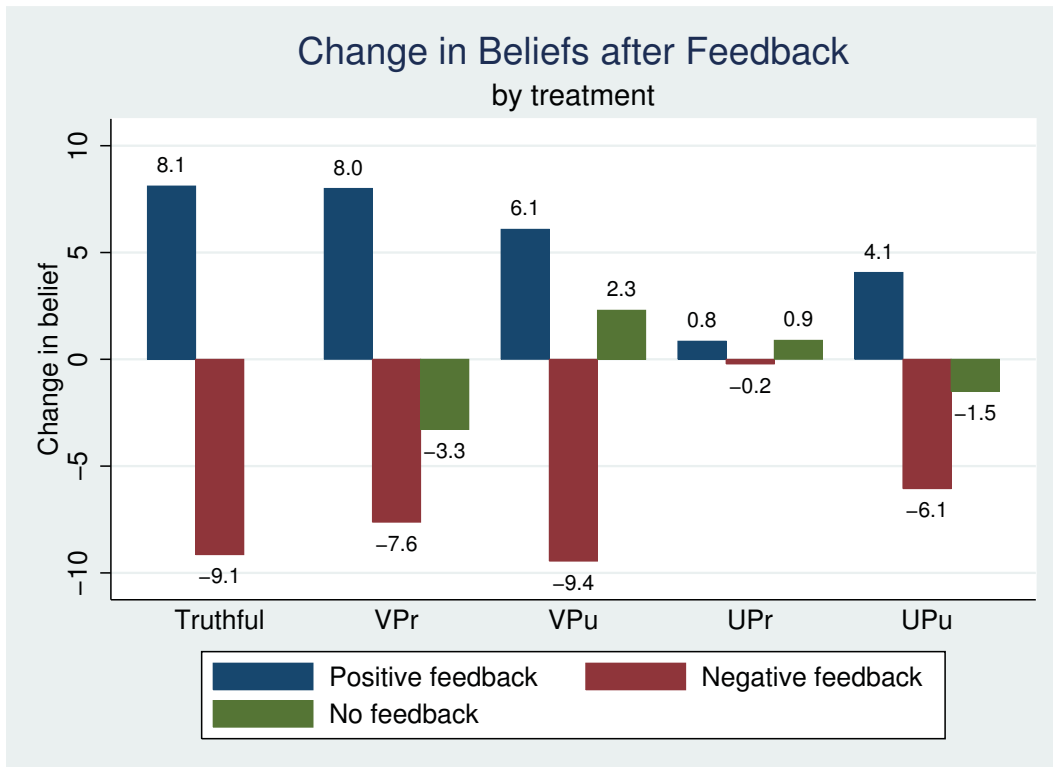
* $p < .1$, ** $p < .05$, *** $p < .01$

5.1.2 Impact of Feedback on Beliefs

In terms of how feedback affects beliefs, our model predicts that under verifiable feedback, beliefs will increase after positive feedback and decrease after negative or no feedback. Under unverifiable feedback, on the other hand, beliefs will not change in response to positive feedback and will decrease or stay the same after negative or no feedback (see item number 2 in predictions 1 through 4).

Figure 2 shows the impact of positive, negative, and no feedback on beliefs under different feedback mechanisms and indicates that support for our predictions in this case is mixed. 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.³⁸ On the other hand, the change in beliefs after no feedback is not significantly different from zero in any of the treatments.³⁹ Therefore, no feedback is not interpreted as the bad news that it is under verifiable feedback, while in contrast to our prediction, positive feedback leads to an increase in beliefs in public-unverifiable feedback.

Figure 2: Impact of Feedback on Beliefs



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,

³⁸The p -values of Wilcoxon signed rank test for each treatment are as follows: Truthful with $p = 0.0002$ and $p = 0.0002$, verifiable-private with $p = 0.0002$ and $p = 0.0466$, verifiable-public with $p = 0.0021$ and $p = 0.0033$, unverifiable-public with $p = 0.0050$ and $p = 0.09$, unverifiable-private with $p = 0.2359$ and $p = 0.4871$, for positive and negative feedback cases respectively.

³⁹The p -values of Wilcoxon signed rank test for each treatment are $p = 0.2109$, $p = 0.9804$, $p = 0.6741$, $p = 0.7098$, for verifiable-private, verifiable-public, unverifiable-private, and unverifiable-public, respectively.

the change in beliefs under verifiable private 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)). This supports our predictions except for verifiable-public feedback, where theory predicts no difference from truthful feedback.⁴⁰ When subjects receive negative feedback, none of the treatments, except unverifiable-private feedback, is different from truthful feedback. This finding does not contradict our predictions either. Under no feedback, on the other hand, we find that there is no difference across treatments, and this contradicts our prediction that beliefs should decrease in response to no feedback in verifiable feedback mechanisms.

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)
VPr	-2.318 (2.945)	2.391 (5.519)	-0.630 (5.605)	-0.759 (2.842)	2.926 (5.460)
VPu	-5.017* (2.879)	1.673 (5.063)	3.504 (5.624)		
UPr	-9.383*** (2.599)	12.22** (5.552)		-7.697*** (2.479)	12.12** (5.445)
UPu	-6.500** (2.778)	4.101 (6.176)	0.449 (5.828)		
P-P (VPu)				-6.090* (3.565)	
P-N (VPu)				-1.170 (3.524)	
N-P (VPu)					-2.029 (5.992)
N-N (VPu)					7.033 (7.129)
P-P (UPu)				-6.909** (3.143)	
P-N (UPu)				3.376 (5.078)	
N-P (UPu)					-5.040 (8.665)
N-N (UPu)					12.99 (9.353)
Session	YES	YES	YES	YES	YES
N	207	149	91	207	149
χ^2	21.79	15.75	.	23.93	18.99

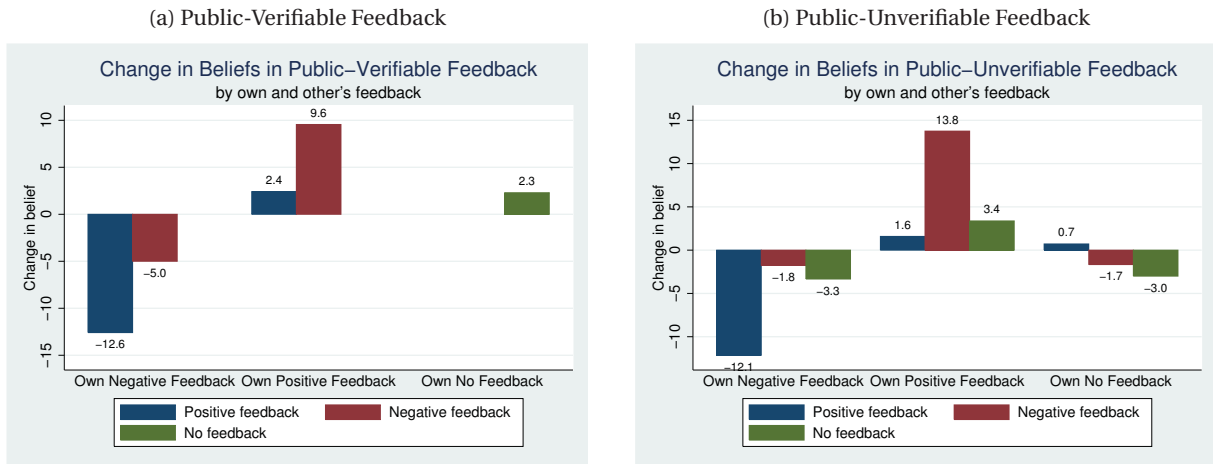
GLS regressions, standard errors in parentheses. In columns (1), (2), (4), and (5) the baseline is truthful feedback, in (3) it is unverifiable-private

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

P-P, P-N, N-P, N-N refers to the message pair sent to agents

⁴⁰Note, however, that the coefficient of VPu is only marginally significant.

Figure 3: Change in Beliefs in Public Feedback



The above findings suggest that verifiable public feedback can have different effects than verifiable 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.

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. Therefore, we conclude that the data contradict our prediction that in public-verifiable feedback beliefs should not depend on the other agent's feedback (see item number 4 in Prediction 3). Column (4) of Table 2 also shows that the less optimistic response to verifiable positive feedback than truthful positive feedback comes from observations where own positive feedback is accompanied with positive feedback for the other agent.⁴¹ We will further discuss the significance of these results in Section 6 within the context of a model with lying costs and naive agents.

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 Becker-De Groot-Marschak mechanism was skewed toward inducing subjects

⁴¹These results are robust to taking the dependent variable to be the posterior beliefs and controlling for the prior beliefs as a regressor.

to enter.⁴² 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. Therefore, the principal has an additional incentive to induce higher beliefs.⁴³ We also check the impact of feedback on performance, both overall and in each treatment separately, and find no significant effect. Overall, although our experiment is not designed to analyze this issue, we may conclude that inducing higher beliefs is good for performance but interim performance feedback is not an effective tool in this respect.

Table 3: Impact of Beliefs on Performance

	(1)
	Post-feedback perf.
Pre-feedback Performance	0.442*** (0.0489)
Change in beliefs	0.0135* (0.00775)
Session	YES
N	392
χ^2	130.6

GLS regressions, standard errors in parentheses

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

5.2 Analysis of Principals' Behavior

We now turn to explore the principals' side of the analysis. 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) and this difference is statistically significant at 5% level in a test of proportions ($p = 0.045$).

⁴²Note that this does not affect the incentive compatibility of the mechanism.

⁴³Remember that for the principal two issues are relevant: (1) whether the agents work at all; (2) how well they perform.

Table 4: Feedback to Individual Agents under Different Treatments

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

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

Columns VPr and VPu of Table 4 show that principals prefer to transmit information 44% of the time when the outcome is bad in the verifiable-private case, while the frequency of transmission is 57% under verifiable-public 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 test of proportions shows that reports significantly change according to the actual outcome when talk is cheap ($p = 0.0010$ in private-unverifiable and $p = 0.0006$ in public-unverifiable). 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, and others report no information. Overall, these findings suggest that some principals prefer to tell the truth even if this might adversely affect their payoff. Therefore, we conclude that our predictions regarding principals' behavior in unverifiable feedback cases (item 1 in predictions 2 and 4) are not supported. We will comment more on principals' strategy in Section 6 and discuss how it can be interpreted using a model with individual specific lying costs and naive agents.

Table 5: Feedback to Both Agents in Private Feedback

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

Number of agents receiving each message type, percentages in parentheses

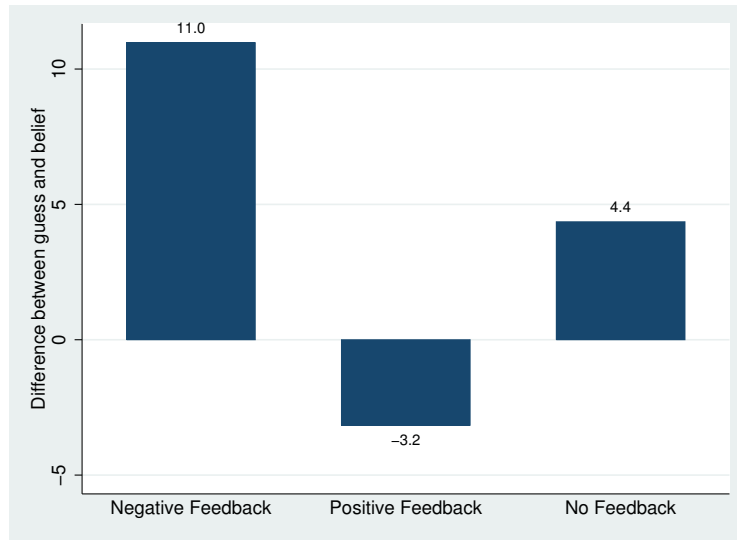
Table 6: Feedback to Both Agents in Public Feedback

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

Number of agents receiving each message type, percentages in parentheses

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 ($p = 0.002$ in a Wilcoxon test). Thus, principals generally overestimate the response of agents' beliefs to the feedback, especially in the case where they send a negative message. The expectation of a pessimistic response to the 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: Difference between Principal's Guess and Actual Belief



Finally, we examine principals' expectations regarding how agents' beliefs will be influenced by feedback given to other agents. Table 7 shows that principals expect that positive feedback given to the other agent will influence the agent's beliefs negatively in public-unverifiable feedback, but expect no significant impact in public-verifiable feedback. 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: Principal's Expectations in Public Feedback

	(1) guess	(2) guess (VPu)	(3) guess (UPu)
Other Positive Feedback	-3.151 (3.198)	3.708 (5.692)	-10.20** (4.176)
VPu	10.19*** (2.960)		
N	75	31	44
χ^2	14.73	0.424	5.959

GLS Regressions, standard errors in parentheses

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

Other Positive Message = 1 if other receives positive feedback

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 has also received 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 it suggests that individuals suffer from cost of lying and this cost varies among them.⁴⁴

The second finding might be due to naivety 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 somehow 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 providing no information) for the principals and naivety on the part of the agents. We will 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 cost of lying and cost of no information is small enough; and (5) there are some who prefer providing no information to lying.⁴⁵

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.

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.

⁴⁴See the discussion in Section 2.

⁴⁵These assumptions are equivalent to the following: (1) $F_{\theta|r}(x) = 1$ 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 $G(\cdot|r_{-i}, \theta_{-i})$ such that $G(x|r_{-i}, \theta_{-i}) > 0$ for all r_{-i}, θ_{-i} and $x > 0$; (5) $G(v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i})|r_{-i}, \theta_{-i}) < 1$ for all r_{-i}, θ_{-i} .

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, the the principal gives positive feedback, while if he has failed, then those principals with small costs of “no feedback” report no information while those with large costs report failure. Proportion of principals who report no information increases with 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 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 provide 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.⁴⁶ This leads to the following prediction.

Prediction 6. *If lying is costly and some agents are naive, then under private-verifiable feedback:*

1. *All principals report truthfully if the agent has succeeded and some report truthfully and some no information if the agent has failed;*
2. *Beliefs increase after positive feedback and decrease after negative or no feedback;*
3. *Beliefs after negative feedback are smaller than beliefs after no feedback;*
4. *Beliefs increase conditional on success and decrease conditional on failure.*

Prediction 6 is similar to its counterpart in our original model, Prediction 1, with the exception of item 3, which is novel and follows from the existence of naive agents in the new model. Table 5 strongly supports item 1 in Prediction 6, and Figure 2 and the other evidence presented in Section 5.1.2 provide support for predictions (2) and (3), while Figure 1 supports item 4. However, the decrease in beliefs after no feedback is not statistically significant.

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 $q(\emptyset|sf) = q(\emptyset|fs) = 0$.

⁴⁶Please see Appendix A for the calculation of beliefs in this section.

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 provide no information. The fraction of principals who provide no information 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' actions 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 report no information. 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.⁴⁷

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 7. *If lying is costly and some agents are naive, then under public-verifiable feedback:*

1. *All principals report truthfully if both agents have succeeded. If at least one of the agents has failed, then some principals tell the truth and some report no information;*
2. *Beliefs increase after positive feedback and decrease after negative feedback;*
3. *Beliefs after negative feedback are smaller than beliefs after no feedback;*
4. *Beliefs decrease conditional on failure but the magnitude of change is smaller than it is under private feedback;*
5. *Beliefs do not depend on the feedback provided to the other agent.*

Our empirical findings give strong support to item 1 (see Table 6) as well as items 2 and 3 (see Figure 2). Note that item 3 is novel in the new model. Also note that in the data, beliefs after no feedback increase, 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 . Prediction (4) is supported as well (see Figure 1) but prediction (5) is rejected (see Figure 3a and Table 2 columns (4) and (5)).

Overall, theoretical predictions are very close to the data except that in the data beliefs are more pessimistic if own positive feedback is accompanied by positive feedback to the other agent. Therefore, we conclude agents make non-Bayesian social comparisons in forming their beliefs or they believe that the difficulty of the tasks are correlated.

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

⁴⁷We show in Appendix A that in a symmetric equilibrium where $q(\emptyset|sf) = q(\emptyset|fs)$, if beliefs about the other agent is uniform, then beliefs after no feedback must decrease.

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 not giving information report truthfully, those with small costs of lying report success, and those with larger costs of lying but small costs of no information give no feedback.

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

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

Using this result and Bayes' rule we have the following prediction.

Prediction 8. *If lying is costly and some agents are naive, then under private-unverifiable feedback*

1. *All principals report truthfully if the agent has succeeded; Some principals report success, some no information, and some truthfully if the agent has failed;*
2. *Beliefs increase after positive feedback at a magnitude smaller than those under verifiable feedback;*
3. *Beliefs decrease after negative and no feedback at a magnitude similar to those under verifiable negative feedback;*
4. *Beliefs increase conditional on success at a magnitude smaller than those in private-verifiable feedback.*

Except for a few outliers, Table 5 gives strong support to item 1. Item 2 is also supported but item 3 is only partially supported: beliefs after negative feedback indeed decrease, but the magnitude of decrease is much smaller under unverifiable than under verifiable feedback (see Figure 2) and beliefs after no feedback do not decrease. Finally, item 4 is supported (see Figure 1). Overall, there is a good fit in terms of the principal's behavior and agents' beliefs except that agents do not interpret negative or no feedback as pessimistically as they should.

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 9. *If lying is costly and some agents are naive, then under public-unverifiable feedback:*

1. *All principals report truthfully to the agent who has succeeded but, to the agent who has failed, some principals report success, some no information and some failure;*
2. *Beliefs after positive feedback are smaller than those under verifiable feedback;*

3. *Beliefs decrease after negative and no feedback at a magnitude similar to those under verifiable negative feedback;*
4. *Beliefs conditional on success are smaller than those in private-verifiable feedback;*
5. *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 of the other agent.*

In the new model, some principals report truthfully and some no information when the agent has failed, while they always provide positive feedback in the original model. Also, agents respond to feedback and there is some information transmission in the new model, whereas none existed in the original model. We also have a new prediction, item 5, which 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.

Except for a few outliers, Table 6 gives strong support to item 1. Item 2 is also supported but item 3 is partially supported because the decrease in beliefs after negative and no feedback is smaller compared with verifiable feedback (see Figure 2). Item 4 also receives support (see Figure 1). Perhaps most remarkably, item 5 is strongly supported (see Figure 3b and Table 2 columns (4) and (5)).

We can also see that the effect of the other agent's feedback is larger under public-unverifiable feedback compared with that under public-verifiable feedback (compare Figure 3a and Figure 3b). The reason is that under verifiable feedback we only see the effect of social comparison (or belief in correlated tasks), while under unverifiable feedback there is also the effect coming from the principal's strategy.

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 under 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.⁴⁸ 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.

⁴⁸In calculating principals' strategy we eliminated some outliers in tables 5 and 6: in VPr row SF column No Info, Info; in UPr row FS column SNo; in UPu row SS column NoS, row FS columns SF and SNo, row SF columns FF and NoF.

Figure 5: Information Transmission: Actual vs. Bayesian

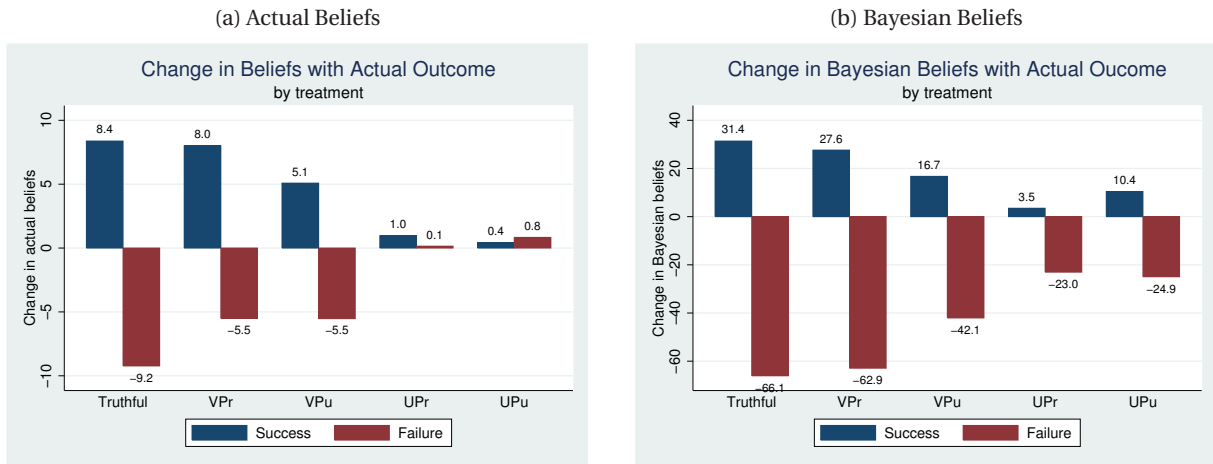


Figure 5 plots the average change in actual and Bayesian beliefs in each treatment conditional on the actual outcome of the agent. We can see that the direction of change is the same in actual and Bayesian beliefs, except under unverifiable feedback when the actual outcome is failure. Also, compared with the Bayesian case, overall information transmission is much weaker when the actual state is failure.⁴⁹

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.

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 a positive feedback to the other agent more pessimistically than it is justified by Bayesian updating alone.

In Table 8, we present regression results which show that Bayesian updating plays a significant role in the formation of actual beliefs and that it explains about 20% of the total variation in them. 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.

Overall, we conclude that agents' beliefs are consistent with the strategy employed by the princi-

⁴⁹We 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 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 apply to the the other graphs in this section.

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

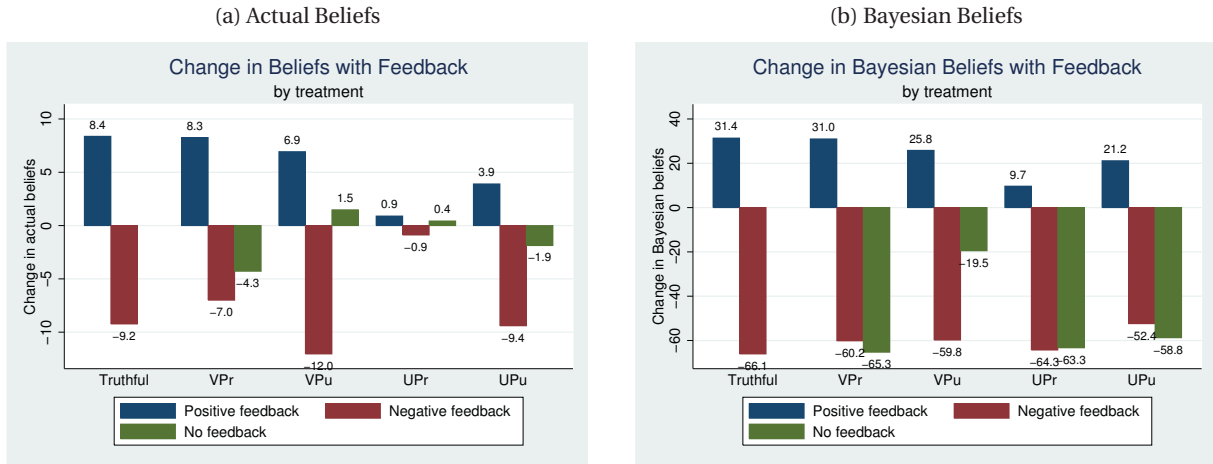


Figure 7: Change in Beliefs with Other's Feedback: Actual vs. Bayesian

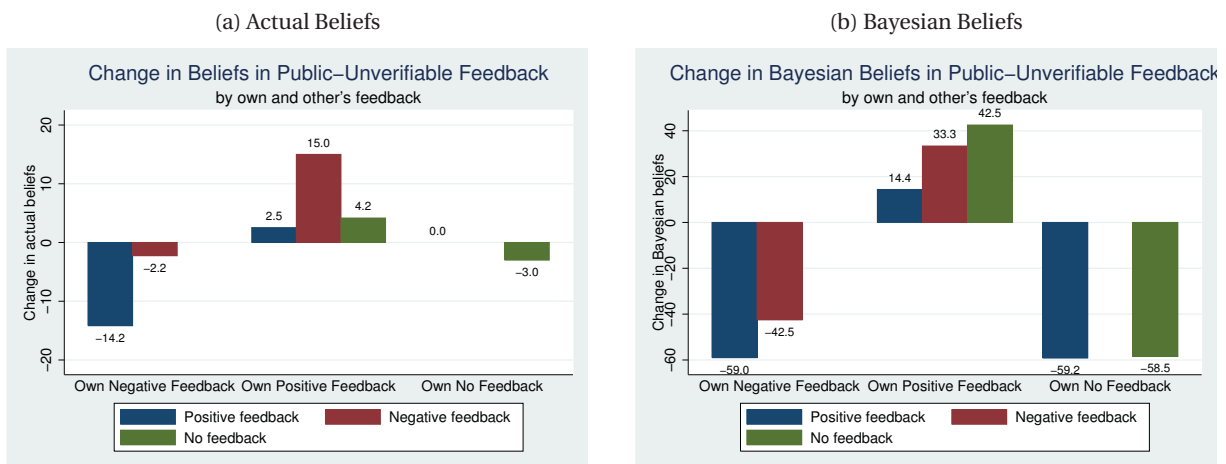


Table 8: Post-Feedback Beliefs

	(1)	(2)	(3)
	Post-Feedback Belief	Post-Feedback Belief	Post-Feedback Belief
Bayesian Belief	0.198*** (0.0241)	0.154*** (0.0216)	0.156*** (0.0216)
Pre-Feedback Belief		0.655*** (0.0571)	0.649*** (0.0591)
VPr			1.723 (2.724)
VPu			-0.433 (3.028)
UPr			0.362 (2.819)
UPu			-1.278 (2.749)
N	386	386	386
χ^2	67.14	213.7	232.5
R^2	0.205	0.544	0.545

GLS Regressions, standard errors in parentheses

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

pals 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. Bayesian updating does not explain the entire variation in beliefs and it seems that some of the agents are naive or suffer from self-serving bias and engage in non-Bayesian social comparisons in their interpretation of performance feedback.

7 Conclusion

In this paper, we have employed 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 predicted that information about the 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.

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 public-unverifiable than in public-verifiable feedback. We have analyzed a model with individual-specific lying costs and naive agents and showed that it can account for these findings except that other agent's feedback should still have no impact on beliefs about own performance in *public-verifiable* feedback. Also we find that agents do not interpret negative or no feedback as pessimistically as they should under *private-unverifiable* feedback.

Overall, our results suggest that individuals suffer from lying costs and that these costs differ across individuals. Also, some agents are naive or suffer from a self-serving bias and engage in non-

Bayesian social comparisons in their interpretation of performance feedback. Although these factors render even unverifiable feedback somehow informative, credible revelation of performance requires verifiability.

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].
- Abeler, Johannes; Becker, Anke, and Falk, Armin. Truth-telling: A representative assessment. *IZA Discussion Paper*, 2012.
- Aoyagi, Masaki. Information feedback in a dynamic tournament. *Games and Economic Behavior*, 70 (2):242–260, 2010.
- Azmat, Ghazala and Iriberry, Nagore. 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.
- Azmat, Ghazala and Iriberry, Nagore. The provision of relative performance feedback information: An experimental analysis of performance and happiness. *CEP Discussion Paper*, 2012.
- Bandiera, Oriana; Larcinese, Valentino, and Rasul, Imran. Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on performance. *Working Paper*, May 2010.
- Battaglini, Marco and Makarov, Uliana. Cheap talk with multiple audiences: An experimental analysis. *Working Paper*, 2010.
- Benndorf, Volker; Kübler, Dorothea, and Normann, Hans-Theo. Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *Working Paper*, September 2013.
- Blanes i Vidal, Jordi and Nossol, Mareike. Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10):1721–1736, 2011.
- Cai, Hongbin and Wang, Joseph Tao-Yi. Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56:7–36, 2006.
- Charness, Gary and Dufwenberg, Martin. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006.
- Charness, Gary; Masclet, David, and Villeval, Marie Claire. Competitive preferences and status as an incentive: Experimental evidence. *Working Paper*, 2010.
- Chen, Ying. Perturbed communication games with honest senders and naive receivers. *Journal of Economic Theory*, 146(2):401–424, 2011.
- Crawford, Vincent. A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78:286–298, 1998.

- Crawford, Vincent P. and Sobel, Joel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.
- Culbert, Samuel A. 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].
- Dickhaut, JohnW.; McCabe, KevinA., and Mukherji, Arijit. An experimental study of strategic information transmission. *Economic Theory*, 6(3):389–403, 1995.
- Drugov, Mikhail; Hernán-González, Roberto; Kujal, Praveen, and Martinez, Marta Troya. Cheap talk with two audiences: An experiment. *Working Paper*, December 2013.
- Ederer, Florian. Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*, 19(3):733–769, 2010.
- Ederer, Florian and Fehr, Ernst. Deception and incentives: How dishonesty undermines effort provision. *Working Paper*, May 2009.
- Eichenwald, Kurt. Microsoft’s lost decade. *Vanity Fair*, August 2012. URL <http://www.vanityfair.com/news/business/2012/08/microsoft-lost-mojo-steve-ballmer>. [Online; accessed 16-September-2015].
- Eil, David and Rao, Justin M. The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138, 2011.
- Eriksson, Tor; Poulsen, Anders, and Villeval, Marie Claire. Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688, 2009.
- Ertac, Seda. Social comparisons and optimal information revelation: Theory and experiments. *Working Paper*, 2005.
- Ertac, Seda. 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.
- Farrell, Joseph and Gibbons, Robert. Cheap talk with two audiences. *The American Economic Review*, 79(5):1214–1223, 1989.
- Fischbacher, Urs. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.
- Fischbacher, Urs and Föllmi-Heusi, Franziska. Lies in disguise-an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547, 2013.
- Forsythe, Robert; Isaac, R. Mark, and Palfrey, Thomas R. Theories and tests of “blind bidding” in sealed-bid auctions. *RAND Journal of Economics*, 20:214–238, 1989.
- Fudenberg, Drew and Tirole, Jean. Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260, 1991.

- Gerhards, Leonie and Siemer, Neele. Private versus public feedback-the incentive effects of symbolic awards. *Working Paper*, 2014.
- Giovannoni, Francesco and Seidmann, Daniel J. Secrecy, two-sided bias and the value of evidence. *Games and Economic Behavior*, 59:296–315, 2007.
- Gneezy, Uri. Deception: The role of consequences. *The American Economic Review*, 95(1):384–394, 2005.
- Gneezy, Uri; Rockenbach, Bettina, and Serra-Garcia, Marta. Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293 – 300, 2013.
- Goltsman, Maria and Pavlov, Gregory. How to talk to multiple audiences. *Games and Economic Behavior*, 72:100–122, 2011.
- Grossman, Sanford J. The informational role of warranties and private disclosure about product quality. *Journal of law and economics*, 24(3):461–483, 1981.
- Gürtler, Oliver and Harbring, Christine. Feedback in tournaments under commitment problems: Experimental evidence. *Journal of Economics & Management Strategy*, 19(3):771–810, 2010.
- Hagenbach, Jeanne; Koessler, Frédéric, and Perez-Richet, Eduardo. Certifiable pre-play communication: Full disclosure. *Econometrica*, 82:1093–1131, 2014.
- Hoelzl, Erik and Rustichini, Aldo. Overconfident: Do you put your money on it? *The Economic Journal*, 115(503):305–318, 2005.
- Hurkens, Sjaak and Kartik, Navin. Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, 12(2):180–192, 2009.
- Jackson, Eric. 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-terribly/>. [Online; accessed 24-August-2015].
- Kantor, Jodi and Streitfeld, David. Inside amazon: Wrestling big ideas in a bruising workplace. *The New York Times*, August 15 2015. URL http://www.nytimes.com/2015/08/16/technology/inside-amazon-wrestling-big-ideas-in-a-bruising-workplace.html?_r=0.
- Karni, Edi. A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606, 2009.
- Kartik, Navin. Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395, 2009.
- Kartik, Navin; Ottaviani, Marco, and Squintani, Francesco. Credulity, lies, and costly talk. *Journal of Economic theory*, 134(1):93–116, 2007.
- Koessler, Frédéric. Lobbying with two audiences: Public vs private certification. *Mathematical Social Sciences*, 55(3):305–314, 2008.

- Kuhnen, Camelia M and Tymula, Agnieszka. Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113, 2012.
- Lai, Ernest K.; Lim, Wooyoung, and Wang, Joseph Tao-yi. An experimental analysis of multidimensional cheap talk. *Working Paper*, April 2015.
- Longenecker, Clinton O; Sims Jr, Henry P, and Gioia, Dennis A. Behind the mask: The politics of employee appraisal. *The Academy of Management Executive*, pages 183–193, 1987.
- Mathis, Jérôme. Full revelation of information in sender-receiver games of persuasion. *Journal of Economic Theory*, 143:571–584, 2008.
- Milgrom, Paul R. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391, 1981.
- Mobius, Markus M; Niederle, Muriel; Niehaus, Paul, and Rosenblat, Tanya S. Managing self-confidence: Theory and experimental evidence. *NBER Working Paper*, 2011.
- Mohnen, Alwine and Manthei, Kathrin. Is honesty the best policy? An experimental study on the honesty of feedback in employer-employee relationships. *Working Paper*, 2006.
- Murphy, Kevin R and Cleveland, Jeanette N. *Performance appraisal: An organizational perspective*. Allyn & Bacon, 1991.
- Rosaz, Julie. Biased information and effort. *Economic Inquiry*, 50(2):484–501, 2012.
- Sanchez-Pages, Santiago and Vorsatz, Marc. Enjoy the silence: an experiment on truth-telling. *Experimental Economics*, 12(2):220–241, 2009.
- Schraeder, Mike; Becton, J. Bret, and Portis, Ron. A critical examination of performance appraisals. *The Journal for Quality and Participation*, 30(1):20–25, Spring 2007.
- Seidmann, Daniel J and Winter, Eyal. Strategic information transmission with verifiable messages. *Econometrica*, 65(1):163–169, 1997.
- Sobel, Joel. Giving and receiving advice. In Acemoglu, Daron; Arellano, Manuel, and Dekel, Eddie, editors, *Advances in Economics and Econometrics*. Cambridge University Press, 2013.
- Sutter, Matthias. Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119(534):47–60, 2009.
- Van Zandt, Timothy and Vives, Xavier. Monotone equilibria in Bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360, 2007.
- Vespa, Emanuel and Wilson, Alistair J. Communication with multiple senders: An experiment. *Working Paper*, April 2015.
- Wang, Joseph Tao-yi; Spezio, Michael, and Camerer, Colin F. 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.

A Proofs and Supporting Results (For online publication only)

Proof of Proposition 1. Suppose, for contradiction, that there is an equilibrium in which $\mu_i(\rho_i(\theta_i, \theta_{-i}))$ puts probability less than one on type θ_i . If $\theta_i = s$, then principal with type (s, θ_{-i}) can deviate and report s to agent i while keeping her report to the other agent unchanged. This increases agent i 's beliefs and hence his action, without changing the other agent's action, and benefits the principal. If $\theta_i = f$, then the principal with type $\theta' = (s, \theta'_{-i})$, for some θ'_{-i} , must be sending $\mu_i(\rho_i(f, \theta_{-i}))$ with positive probability. This implies that $\mu_i(\rho_i(f, \theta_{-i}))$ puts probability less than one on $\theta_i = s$ as well. But then principal with type θ' can deviate and report s to agent i while keeping her report to the other agent unchanged. This increases agent i 's beliefs and benefits the principal. \square

Proof of Proposition 2. Suppose, for contradiction, that agent i learns something about his own type from the principal's equilibrium report. In other words, suppose that there exist i , θ and θ' such that $\mu_i^i(\rho_i(\theta)) > \mu_i^i(\rho_i(\theta'))$. This implies that $\alpha_i(\rho_i(\theta)) > \alpha_i(\rho_i(\theta'))$. But then, principal with type θ' can deviate and report $\rho_i(\theta)$ to agent i while keeping her report to the other agent unchanged. This increases agent i 's action, without changing the other agent's action, and benefits the principal. \square

Proof of Proposition 3. We will first show that $\rho(s, s) = (s, s)$ in any equilibrium. Suppose, for contradiction, that there exists a $\theta \neq (s, s)$ such that $\rho(\theta) = \rho(s, s)$. This implies that $\mu_i^i(\rho(s, s)) < 1$ for some $i = 1, 2$. But then, type (s, s) has a profitable deviation to $r = (s, s)$. Therefore, we conclude that $\rho(s, s) = (s, s)$ in any equilibrium. Suppose now that $\rho(f, f) = (f, f)$. If $\rho(s, f) \neq \rho(f, s)$, then there is full information revelation. If $\rho(s, f) = \rho(f, s) = \emptyset$, then type (f, f) has a profitable deviation to \emptyset . Suppose, alternatively, that $\rho(f, f) = \emptyset$ but $\rho(s, f) \neq \emptyset$. If $\rho(f, s) \neq \emptyset$ as well, then there is full information revelation. If $\rho(f, s) = \emptyset$, then type (f, s) has a profitable deviation. Similarly, if $\rho(f, f) = \emptyset$ but $\rho(f, s) \neq \emptyset$. \square

Proof of Proposition 4. It is easy to construct a completely uninformative equilibrium.⁵⁰ It is also easy to see that there is no equilibrium in which the agents receive full information. In fact, we can show that type (f, f) must always send the same feedback with some other type. Because, otherwise, at least one of the agents must have beliefs that put positive probability on s after some message r and type (f, f) would have a profitable deviation to r . In order to prove that $\rho(s, s) = \rho(f, f)$, suppose, for contradiction, that $\rho(f, f) \neq \rho(s, s)$. Then, the previous argument implies that $\rho(f, f) = \rho(f, s)$ or $\rho(f, f) = \rho(s, f)$. Suppose first that $\rho(f, f) = \rho(f, s)$. This implies that $\rho(s, f) = \rho(s, s)$ since otherwise type (s, f) has a profitable deviation to $\rho(s, s)$. But this implies that type (f, f) has a profitable deviation to $\rho(s, s)$. We obtain a similar contradiction if $\rho(f, f) = \rho(s, f)$. \square

Proof of Proposition 5. Fix i and θ_{-i} and let $\theta = (s, \theta_{-i})$. For any $r \in M_{-i}(\theta)$

$$v(a_i(1), a_{-i}(\mu_{-i}(r)), s, \theta_{-i}) \geq v(a_i(\mu_i^S(\emptyset)), a_{-i}(\mu_{-i}(r)), s, \theta_{-i}).$$

Therefore, no feedback does not increase the payoff for type s but it has a cost, which implies that

⁵⁰This is well known in the literature: in a cheap-talk game, there is always a babbling equilibrium.

$q_i(s|s, \theta_{-i}) = 1$. Now let $\theta = (f, \theta_{-i})$. For any $r \in M_{-i}(\theta)$,

$$(1 - \eta)v(a_i(\mu_i^S(\emptyset)), a_{-i}(\mu_{-i}(r)), f, \theta_{-i}) + \eta v(a_i(p), a_{-i}(\mu_{-i}(r)), f, \theta_{-i}) \\ > v(a_i(0), a_{-i}(\mu_{-i}(r)), f, \theta_{-i}),$$

and hence there exists a positive measure of principals with small enough costs $c(\emptyset, r|f, \theta_{-i})$ who prefer sending no information. This implies that $q_i(\emptyset|f, \theta_{-i}) > 0$ and $\mu_i^S(\emptyset) = 0$. Since there is also a positive measure of principals who always tell the truth, we also have $q_i(f|f, \theta_{-i}) > 0$. \square

Beliefs in private-verifiable feedback with lying cost and naive agents: Bayes' rule implies that beliefs of the sophisticated agents are $\mu_i^S(s) = 1$, $\mu_i^S(f) = 0$, $\mu_i^S(\emptyset) = 0$ while average beliefs are $\mu_i(s) = 1$, $\mu_i(f) = 0$, $\mu_i(\emptyset) = \eta p$. Average beliefs conditional on state are $\mu_i(\theta_i = s) = 1$ and

$$\mu_1(\theta_1 = f) = \left(\frac{p(ff)}{p(ff) + p(fs)} q_1(\emptyset|ff) + \frac{p(fs)}{p(ff) + p(fs)} q_1(\emptyset|fs) \right) \eta p \\ \mu_2(\theta_2 = f) = \left(\frac{p(ff)}{p(ff) + p(sf)} q_2(\emptyset|ff) + \frac{p(sf)}{p(ff) + p(sf)} q_2(\emptyset|sf) \right) \eta p$$

Proof of Proposition 6. Since $\mu_i(ss) = 1$ and sending no information is costly, $q(ss|ss) = 1$. Since there are some naive agents and principals with small costs of not telling the truth as well as those with large costs, $q(\emptyset|ff) > 0$ and $q(ff|ff) > 0$. The left hand side of (1) is equal to the worst payoff that type sf expects from sending \emptyset . Therefore, if this condition holds, then there will be some principals who will find sending \emptyset profitable. If it does not hold, then there is an equilibrium in which $q(\emptyset|sf) = q(\emptyset|fs) = 1$, which is supported with beliefs $\mu_i^S(\emptyset) = 0$. \square

Beliefs in public-verifiable feedback with lying cost and naive agents: Beliefs are given by $\mu_i^S(s, m_{-i}) = 1$, $\mu_i^S(f, m_{-i}) = 0$ for all m_{-i} and

$$\mu_1^S(\emptyset) = \frac{q(\emptyset|sf)p(sf)}{q(\emptyset|sf)p(sf) + q(\emptyset|fs)p(fs) + q(\emptyset|ff)p(ff)} \\ \mu_2^S(\emptyset) = \frac{q(\emptyset|fs)p(fs)}{q(\emptyset|fs)p(fs) + q(\emptyset|sf)p(sf) + q(\emptyset|ff)p(ff)}$$

Note that beliefs of agent 1 after no feedback decrease (i.e., are smaller than the prior) if and only if

$$q(\emptyset|sf) \frac{p(sf)}{p(ss) + p(sf)} < q(\emptyset|fs) \frac{p(fs)}{p(ff) + p(fs)} + q(\emptyset|ff) \frac{p(ff)}{p(ff) + p(fs)}.$$

In particular, this is the case in a symmetric equilibrium where $q(\emptyset|sf) = q(\emptyset|fs)$ and beliefs about the other agent is uniform, i.e., $p(ss) = p(sf)$ and $p(fs) = p(ff)$. Average beliefs are $\mu_i(s, m_{-i}) = 1$, $\mu_i(f, m_{-i}) = 0$ for all m_{-i} and $\mu_i(\emptyset) = (1 - \eta)\mu_i^S(\emptyset) + \eta p$. Also, note that beliefs after no information are larger than those under private-verifiable feedback. Average beliefs conditional on the actual state

are

$$\begin{aligned}\mu_1(\theta_1 = s) &= \frac{p(ss)}{p(ss) + p(sf)} + \frac{p(sf)}{p(ss) + p(sf)} [q(sf|sf) + q(\emptyset|sf)\mu_1(\emptyset)] \\ \mu_1(\theta_1 = f) &= \frac{p(ff)q(\emptyset|ff) + p(fs)q(\emptyset|fs)}{p(ff) + p(fs)} \mu_1(\emptyset) \\ \mu_2(\theta_2 = s) &= \frac{p(ss)}{p(ss) + p(fs)} + \frac{p(fs)}{p(ss) + p(fs)} [q(fs|fs) + q(\emptyset|fs)\mu_2(\emptyset)] \\ \mu_2(\theta_2 = f) &= \frac{p(ff)q(\emptyset|ff) + p(fs)q(\emptyset|sf)}{p(ff) + p(sf)} \mu_2(\emptyset)\end{aligned}$$

Proof of Proposition 7. Fix i and θ_{-i} , let $\theta = (s, \theta_{-i})$ and r_{-i} be the message sent to agent $-i$. Suppose, for contradiction, that there exists $r_i \neq s$ such that $q_i(r_i|s, \theta_{-i}) > 0$. This implies that

$$v(a_i(\mu_i(r_i)), a_{-i}(\mu_{-i}(r_{-i})), s, \theta_{-i}) > v(a_i(\mu_i(s)), a_{-i}(\mu_{-i}(r_{-i})), s, \theta_{-i}),$$

for otherwise type s would not find sending r_i optimal. Therefore, $a_i(\mu_i(r_i)) > a_i(\mu_i(s))$ and hence for all r_{-i}

$$v(a_i(\mu_i(r_i)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > v(a_i(\mu_i(s)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}).$$

Since cost of r_i is smaller than cost of s for type f , this implies that $q_i(s|f, \theta_{-i}) = 0$. By assumption, there exist s types whose lying costs are so large that they tell the truth, i.e., $q_i(s|s, \theta_{-i}) > 0$. Therefore, $\mu_i^S(s) = 1$, which contradicts

$$v(a_i(\mu_i(r_i)), a_{-i}(\mu_{-i}(r_{-i})), s, \theta_{-i}) > v(a_i(\mu_i(s)), a_{-i}(\mu_{-i}(r_{-i})), s, \theta_{-i}).$$

Also, by assumption, there exist f types whose lying costs are so large that they tell the truth, i.e., $q_i(f|f, \theta_{-i}) > 0$. This implies that $\mu_i^S(f) = 0$. Now, suppose, for contradiction, that $q_i(s|f, \theta_{-i}) = 0$. This implies that $\mu_i^S(s) = 1$ and hence for any r_{-i}

$$v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}).$$

Assume first that $q_i(\emptyset|f, \theta_{-i}) = 0$ so that $q_i(f|f, \theta_{-i}) = 1$. Since

$$F(v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i})) > 0$$

deviation to s is profitable for some principals with small enough costs $c_{sr_{-i}|f\theta_{-i}}$, contradicting $q_i(f|f, \theta_{-i}) = 1$. Assume now that $q_i(\emptyset|f, \theta_{-i}) > 0$ so that $\mu_i^S(\emptyset) = 0$. Expected payoff to s is greater than the expected payoff to f and \emptyset , i.e.,

$$x = v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > 0$$

$$y = v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - (1 - \eta)v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i})$$

$$- \eta v(a_i(\mu_i(p)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > 0$$

which imply that $F(x) > 0$ and $G(y) > 0$. These imply that for some principals reporting s is strictly

better than reporting f and \emptyset , contradicting $q_i(s|f, \theta_{-i}) = 0$.

Note that $q_i(s|s, \theta_{-i}) = 1$ and $q_i(f|f, \theta_{-i}) > 0$ imply $\mu_i^S(s) > p$. Suppose, for contradiction, that $q_i(\emptyset|f, \theta_{-i}) = 0$. Then

$$x = (1 - \eta)v(a_i(\mu_i^S(\emptyset)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) + \eta v(a_i(\mu_i(p)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) \\ - v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) > 0$$

and hence $F_{\emptyset r_{-i}|f \theta_{-i}}(x) > 0$, i.e., there exist principals for whom \emptyset is better than f . Also

$$G(v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - v(a_i(0), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) | r_{-i}, \theta_{-i}) < 1$$

implies that there exist principals for whom

$$(1 - \eta)v(a_i(\mu_i^S(\emptyset)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) + \eta v(a_i(\mu_i(p)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - c_{\emptyset r_{-i}|f \theta_{-i}} \\ > (1 - \eta)v(a_i(\mu_i^S(s)), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) + \eta v(a_i(1), a_{-i}(\mu_{-i}(r_{-i})), f, \theta_{-i}) - c_{sr_{-i}|f \theta_{-i}}$$

so that they prefer \emptyset to s . This proves that $q_i(\emptyset|f, \theta_{-i}) > 0$. \square

Beliefs in private-unverifiable feedback with lying cost and naive agents: Bayes' rule implies that beliefs of the sophisticated agents are $\mu_i^S(f) = 0$, $\mu_i^S(\emptyset) = 0$

$$\mu_1^S(s) = \frac{p(ss) + p(sf)}{p(ss) + p(sf) + p(fs)q_1(s|fs) + p(ff)q_1(s|ff)} \\ \mu_1^S(f) = \frac{p(ss) + p(fs)}{p(ss) + p(fs) + p(sf)q_2(s|sf) + p(ff)q_2(s|ff)}$$

while average beliefs are $\mu_i(s) = (1 - \eta)\mu_i^S(s) + \eta$, $\mu_i(f) = 0$, $\mu_i(\emptyset) = \eta p$. Average beliefs conditional on state are $\mu_i(\theta_i = s) = \mu_i(s)$ and

$$\mu_1(\theta_1 = f) = \left(\frac{p(fs)}{p(fs) + p(ff)} q_1(s|fs) + \frac{p(ff)}{p(fs) + p(ff)} q_1(s|ff) \right) \mu_1(s) \\ + \left(\frac{p(fs)}{p(fs) + p(ff)} q_1(\emptyset|fs) + \frac{p(ff)}{p(fs) + p(ff)} q_1(\emptyset|ff) \right) \eta p \\ \mu_2(\theta_2 = f) = \left(\frac{p(sf)}{p(sf) + p(ff)} q_2(s|sf) + \frac{p(ff)}{p(sf) + p(ff)} q_2(s|ff) \right) \mu_2(s) \\ + \left(\frac{p(sf)}{p(sf) + p(ff)} q_2(\emptyset|sf) + \frac{p(ff)}{p(sf) + p(ff)} q_2(\emptyset|ff) \right) \eta p$$

Note that $\mu_i^S(s) > p$ and hence $\mu_i(s) > p$.

Proof of Proposition 8. Suppose, for contradiction, that there exists $r \neq ss$ such that $q(r|ss) > 0$. Then $v(a(\mu(r)), ss) > v(a(\mu(ss)), ss)$, which implies that $v(a(\mu(r)), \theta) > v(a(\mu(ss)), \theta)$ for all θ . But then no type would find it optimal to send ss . Since there are some ss types who send ss because of high lying costs, this implies that $\mu_i(ss) = 1$ for $i = 1, 2$. This, however, shows that ss induces the highest beliefs and actions and hence all types ss should send ss , a contradiction. For small costs of lying this also

implies that some other type must also send ss . Finally, there exist principals with high enough lying costs who always tell the truth. \square

Beliefs in public-unverifiable feedback with lying cost and naive agents: The strategy is given by

$$\begin{aligned}
q(ss|ss) &= 1 \\
q(ss|sf) \in (0, 1), \quad q(s\emptyset|sf) \in (0, 1), \quad q(sf|sf) &= 1 - q(ss|sf) - q(s\emptyset|sf) > 0 \\
q(ss|fs) \in (0, 1), \quad q(\emptyset s|fs) \in (0, 1), \quad q(fs|fs) &= 1 - q(ss|fs) - q(\emptyset s|fs) > 0 \\
q(ss|ff) \in (0, 1), \quad q(\emptyset\emptyset|ff) \in (0, 1), \quad q(ff|ff) &= 1 - q(ss|ff) - q(\emptyset\emptyset|ff) > 0
\end{aligned}$$

Bayes' rule implies that beliefs are given by

$$\begin{aligned}
\mu_1^S(ss) &= \frac{p(ss) + p(sf)q(ss|sf)}{p(ss) + p(sf)q(ss|sf) + p(fs)q(ss|fs) + p(ff)q(ss|ff)} \\
\mu_2^S(ss) &= \frac{p(ss) + p(fs)q(ss|fs)}{p(ss) + p(fs)q(ss|fs) + p(sf)q(ss|sf) + p(ff)q(ss|ff)} \\
\mu_1^S(sf) &= 1, \mu_2^S(fs) = 1, \mu_1^S(fs) = 0, \mu_2^S(sf) = 0 \\
\mu_1^S(s\emptyset) &= 1, \mu_2^S(\emptyset s) = 1, \mu_1^S(\emptyset s) = 0, \mu_2^S(s\emptyset) = 0 \\
\mu_i^S(\emptyset\emptyset) &= 0, \mu_i^S(ff) = 0, \mu_i^S(\emptyset f) = \text{free}, \mu_i^S(f\emptyset) = \text{free}
\end{aligned}$$

$$\begin{aligned}
\mu_1(ss) &= \eta + (1 - \eta)\mu_1^S(ss) \\
\mu_2(ss) &= \eta + (1 - \eta)\mu_2^S(ss) \\
\mu_1(sf) &= 1, \mu_2(fs) = 1, \mu_1(fs) = 0, \mu_2(sf) = 0 \\
\mu_1(s\emptyset) &= 1, \mu_2(\emptyset s) = 1, \mu_1(\emptyset s) = \eta p, \mu_2(s\emptyset) = \eta p \\
\mu_i(\emptyset\emptyset) &= \eta p, \mu_i(ff) = 0, \\
\mu_1(\emptyset f) &= \eta p + (1 - \eta)\mu_1^S(\emptyset f), \mu_2(\emptyset f) = (1 - \eta)\mu_2^S(\emptyset f) \\
\mu_1(f\emptyset) &= (1 - \eta)\mu_1^S(f\emptyset), \mu_2(f\emptyset) = \eta p + (1 - \eta)\mu_2^S(f\emptyset)
\end{aligned}$$

Beliefs as a function of own feedback are given by

$$\begin{aligned}
\mu_1^S(s) &= \frac{p(ss) + p(sf)[q(ss|sf) + q(sf|sf)]}{p(ss) + p(sf)[q(ss|sf) + q(sf|sf)] + p(fs)q(ss|fs) + p(ff)q(ss|ff)} \\
\mu_1^S(f) &= \frac{p(ff)}{p(fs) + p(ff)}\mu_1^S(ff) + \frac{p(fs)}{p(fs) + p(ff)}\mu_1^S(fs) = 0 \\
\mu_1^S(\emptyset) &= \frac{p(\emptyset\emptyset)}{p(\emptyset\emptyset) + p(\emptyset s)}\mu_1^S(\emptyset\emptyset) + \frac{p(\emptyset s)}{p(\emptyset\emptyset) + p(\emptyset s)}\mu_1^S(\emptyset s) = 0
\end{aligned}$$

$$\begin{aligned}
\mu_1(s) &= \eta + (1 - \eta)\mu_1^S(s) \\
\mu_1(f) &= 0 \\
\mu_1(\emptyset) &= \eta p
\end{aligned}$$

Average beliefs conditional on state are

$$\begin{aligned}
\mu_1(\theta = s) &= \frac{p(ss)}{p(ss) + p(sf)} \mu_1(ss) \\
&+ \frac{p(sf)}{p(ss) + p(sf)} [q(ss|sf) \mu_1(ss) + q(s\emptyset|sf) \mu_1(s\emptyset) + q(sf|sf) \mu_1(sf)] \\
&= \frac{p(ss)}{p(ss) + p(sf)} \mu_1(ss) + \frac{p(sf)}{p(ss) + p(sf)} [q(ss|sf) \mu_1(ss) + 1 - q(ss|sf)] \\
\mu_1(\theta = f) &= \frac{p(fs)}{p(fs) + p(ff)} [q(ss|fs) \mu_1(ss) + q(\emptyset s|fs) \mu_1(\emptyset s) + q(fs|fs) \mu_1(fs)] \\
&+ \frac{p(ff)}{p(fs) + p(ff)} [q(ss|ff) \mu_1(ss) + q(\emptyset \emptyset|ff) \mu_1(\emptyset \emptyset) + q(ff|ff) \mu_1(ff)] \\
&= \frac{p(fs)}{p(fs) + p(ff)} [q(ss|fs) \mu_1(ss) + q(\emptyset s|fs) \eta p] \\
&+ \frac{p(ff)}{p(fs) + p(ff)} [q(ss|ff) \mu_1(ss) + q(\emptyset \emptyset|ff) \eta p]
\end{aligned}$$

B Instructions, Quiz, and Survey (For online publication only)

B.1 Instructions⁵¹

Welcome. Thank you for participating in our study, which is about economic decision making. During the study, you will be asked to solve a set of questions and make decisions regarding your performance. You will earn 10 TL for your participation. Besides this show-up fee, your earnings in the experiment will depend on your performance, your decisions and chance. There is no misleading or deception in this study. The rules that we will state are completely correct and your payment will be determined accordingly. Payments will be made privately and in cash at the end of the experiment. Your decisions during the experiment will be recorded under an anonymous subject number and will never be matched with your identifying information.

The experiment consists of 5 periods and within each period there are 2 rounds. At the end, one round out of the 10 rounds will be randomly selected and your earnings will be determined according to your payoff in that chosen round. During the experiment, all monetary earnings will be denoted in ECU (“experimental currency unit”), where 1 ECU equals 0.06 TL.

For each of the 5 periods, there will be different rules. We will now explain the decision task which you will encounter in all rounds. Specific instructions that pertain to each of the 5 periods will be explained just before that period starts.

General Rules

At the beginning of the experiment, participants will be randomly assigned to one of the roles of “Principal” and “Agent”. The roles will be fixed throughout the experiment. Now, we are going to explain the rules that will be implemented for agents.

Agents:

If you are in the role of agent, you will perform a task. There are 2 types of tasks:

1. Addition task
2. Verbal task

One of these will be assigned to you randomly. In the addition task, you are asked to solve questions involving the addition of 4 or 5 two-digit numbers in 120 sec. (e.g. $11+48+96+24=?$). In the verbal task, you are asked to solve in 120 sec. the following type of questions:

- General knowledge questions (geography, literature, sports...). For example:
 - Which country has the capital city of *Ankara*?
 - Who is the writer of the novel *Crime and Punishment*?
- Verbal classification questions. For example:
 - Which word does not belong to the group?

⁵¹Original instructions were in Turkish and are available upon request. Note that verbal instructions were supplemented with graphical slides to ease understanding. These are also available upon request. The instructions given here are for the following treatment order configuration: Truthful, Private-Verifiable, Public-Verifiable, Private-Unverifiable and Public-Unverifiable feedback.

- Which of the following has the relationship between *Tea: Sugar*?
- Number-word matching questions. Here we assign each letter to a number and ask the following type of question:
 - ALI= 123
 - ILE= 324
 - AILE=? (where the answer is AILE= 1324.)

For each period, a “target score” will be determined for each type of task (addition and verbal), independently and separately. You will work on only one type of task in the two rounds of a period. The outcome will be determined as follows: if the number of correct answers is:

- higher than or equal to the target score, *Success*,
- lower than the target score, *Failure*.

The target score will be kept fixed throughout the two rounds of the same period. That is, your number of correct answers in both rounds within the same period will be compared with the same target score. You will not be informed about the target score.

The payoffs to success and failure in the 1st round are as follows:

If you are successful (you scored at least as high as the target score), you will earn 300 ECU. If you have failed (you scored less than the target score), you will earn 100 ECU.

After the first round, you have two options for the second round. You can either choose to base your 2nd round payoff on your second-round performance, or leave it to a chance mechanism. As an agent:

- If you perform in the second round,
 - You will perform the same type of task as in the first round.
 - If you are successful in the second round (meet/pass the target score) you will get 300 ECU,
 - If you fail in the second round (fall below the target score) you will get 100 ECU.
- If the chance mechanism is implemented in the second round,
 - You will not perform in the 2nd round.
 - You will earn 300 ECU with $X\%$ chance, and 100 ECU with $(100 - X)\%$.

You will make this performance/chance mechanism decision by answering the following question:

What is the minimum % chance of winning in the chance mechanism (X), that will make you willing to leave your second-round payoff to the chance mechanism rather than your own performance?

In other words, you will state for what winning chances you will choose to base your payoff on your own performance and for what winning chances you prefer the chance mechanism. How you make this decision is related to how confident you are about your success in the second round. Consider the following example:

Suppose that someone thinks there is a 80% chance of succeeding in the 2nd performance task. In this case, this person would need the chance mechanism to give at least an 80% chance of winning, to be convinced to leave his/her payoff to chance. This is because he/she has higher expected earnings from performing in the 2nd round (as opposed to picking the chance mechanism), if there is a less than 80% chance of winning in the chance mechanism.

The probability of winning in the chance mechanism (X) will be determined randomly by the computer. If you choose the performance mechanism at the stated chance probability, then you will perform in the second round. Otherwise, your payoff will depend on the chance mechanism (you will earn 300 ECU with probability $X\%$ and 100 ECU with probability $(100 - X)\%$).

Groups:

In each period, 3-person groups that consist of one principal and two agents will be formed. At the beginning of each period, groups will be reshuffled and formed again. The two agents in the same group will always perform different type of tasks (if one does addition the other does verbal) and will be evaluated according to different target scores.

Principals:

Participants in the role of principals will not perform any task. However, they will be able to see whether each agent in their group has been successful or not on their screen.

The payoffs of principals are as follows (valid for all periods):

1st round of each period: A fixed payoff of 100 ECU

2nd round of each period: The payoff will depend on agents' decisions and performances.

Return of agent i to principal in the second round ($return_i$):

- If the agent performed in the second round: $return_i =$ Number of correct answers in the second round * 20 ECU
- If the agent did not perform (if the chance mechanism was implemented): $return_i = 0$ ECU

Payoff function of the principal for the second round = $50\text{ECU (constant payoff)} + \min(return_1, return_2)$, where $return_1$ and $return_2$ refer to the returns from agent 1 and agent 2, respectively.

[Examples were graphically shown on slides. They are available upon request.]

Therefore if at least one of the agents chooses the chance mechanism, the principal will not earn a payoff from the agents' performance. So in order to receive a payoff from agents' performance, both of the agents must perform in the second round.

Now, we have a short quiz about the rules we have just explained. Your answers will not affect your earnings. Rather, they are aimed to check your understanding of the rules. Please do not hesitate to ask questions if there is any part you do not understand.

[Quiz: see Section B.2]

Part I (Truthful Feedback):

As explained before, 3-person groups, consisting of 1 principal and 2 agents, will be formed.

Agents will observe their performance outcome (Success/Failure) before making their second (post-feedback) performance/chance decision. This feedback will be sent by the computer and it will certainly be correct information. Agents will make the performance/chance mechanism decision twice: before and after they receive the feedback on their performance outcome (Success or Failure). One of the two (pre-feedback and post-feedback) stated minimum probabilities will be randomly chosen and used to determine whether the agent will perform or not in the 2nd stage.

The timeline for the agents will be as follows *[shown graphically on slides]*

1. (1st round) Task will be performed.
2. The decision about performance/chance mechanism for the 2nd round will be made.
3. Feedback about the 1st round outcome will be given (Success/Failure).
4. The pre-feedback performance/chance decision can be changed or kept the same.
5. One of the pre- and post-feedback decisions will be randomly chosen. The stated minimum will be compared to the (randomly chosen) chance probability X and either the performance or the chance mechanism will be applied in the 2nd round accordingly.
6. (2nd round) If the performance mechanism is implemented: The same type of task as in the first round will be performed and your earnings will depend on whether you succeed or not (300 ECU/100 ECU). If the chance mechanism is implemented: There will be no performance. 300 ECU with probability X and 100 ECU with probability $(100 - X)$ will be earned.

As agents, the feedback provides you with information about your performance. It can help you evaluate your performance with respect to the target score. Note that the difficulty of the questions might be different between rounds.

Principals will see the 1st-round performance outcomes of the two agents in their group. Principals will also be asked to guess, for each agent, the required minimum winning chance (%) to use the chance mechanism instead of performing in the 2nd stage, as stated by that agent. That is, the principals will guess at which winning probabilities in the chance mechanism each agent will enter/stay out. This guess concerns the post-feedback entry decisions of each agent. If the principal's guess is within ± 5 percentage points of the agent's actual stated minimum winning chance, he/she will earn 10 ECU extra (for each successful guess). At the end of the experiment, these extra earnings in the chosen round will be added to the payoff from that round.

Do you have any questions?

Part II (Private Verifiable Feedback):

In this part, agents may receive feedback about their performance outcome from the principal. Any information that is sent by the principal has to be true. However, the principal also has the option to withhold the information, i.e. not send any message.

The messages from principal will be privately sent. In other words, principals can choose different messages (either the true outcome or no information) for different agents and agents will only see the message the message sent to them, not to the other agent.

As in the previous part, agents make a pre-feedback and post-feedback decision, and one is randomly chosen to be implemented. Principals are asked to guess the required minimum winning probabilities in the chance mechanism (%), stated by the agents post-feedback. The same rules apply for bonus calculation.

[Screenshots of the principal and the agent decision screens were shown on the slides. They are available upon request.]

Do you have any questions?

Part III (Public Verifiable Feedback):

In this part, agents may receive feedback about their performance outcome from the principal. Any information that is sent by the principal has to be true. However, the principal also has the option to withhold information.

The messages from the principal will be publicly sent, meaning that both agents can see the set of messages. The principal can choose to send either truthful information to both of the agents or no information to either.

If the principal chooses to send the information, the agents will see the message sent to the other agent along with their own. Note that the other agent performs another type of task and is evaluated with a different target score. Therefore, the other agent's outcome is independent of yours. However, the message sent to the other agent might provide you with information about the principal's strategy.

As in the previous part, agents make a pre-feedback and post-feedback decision, and one is randomly chosen to be implemented. Principals are asked to guess the required minimum winning probabilities in the chance mechanism (%), stated by the agents post-feedback. The same rules apply for bonus calculation.

[Screenshots of the principal and the agent decision screens were shown on the slides. They are available upon request.]

Do you have any questions?

Part IV (Private Unverifiable Feedback):

In this part, agents may receive feedback about their performance outcome from the principal. Information sent by the principal does not need to be true. The principal also has the option to withhold information.

The principal will observe the actual performance of the two agents and after that, will choose to send one of the following messages: "Success", or "Failure" or "No Information".

The messages from the principal will be privately sent. In other words, principals can choose different messages for agents and agents will only see the message about their own performance.

As in the previous part, agents make a pre-feedback and post-feedback decision, and one is randomly chosen to be implemented. Principals are asked to guess the required minimum winning probabilities in the chance mechanism (%), stated by the agents post-feedback. The same rules apply for bonus calculation.

[Screenshots of the principal and the agent decision screens were shown on the slides. They are available upon request.]

Do you have any questions?

Part V (Public Unverifiable Feedback):

In this part, agents will receive feedback about their performance outcome from the principal.

Information sent by the principal does not need to be true. The principal also has the option to withhold information.

The principal will observe the actual performance of the two agents and after that, will choose to send one of the following messages: "Success", or "Failure" or "No Information".

The messages from the principal will be publicly sent, meaning that agents will see the message sent to the other agent along with their own. Note that the other agent performs another type of task and is evaluated with a different target score. Therefore, other agent's outcome is independent of yours. However, the message sent to other agent might provide you with information about principal's strategy.

As in the previous part, agents make a pre-feedback and post-feedback decision, and one is randomly chosen to be implemented. Principals are asked to guess the required minimum winning probabilities in the chance mechanism (%), stated by the agents post-feedback. The same rules apply for bonus calculation.

[Screenshots of both the principal and the agent screen were shown on the slides. They are available upon request.]

Do you have any questions?

Survey:

Now you will be asked to answer several questions that will come up on your screen. Thank you again for your participation.

B.2 Quiz

1. Assume that you state a 60% chance of being successful in the 2nd stage, and the computer picks the winning probability in the chance mechanism as 55%. Which of the following is true?
 - (a) My probability of earning 300 ECU is 55% and my payoff is independent of my performance.
 - (b) My payoff is 300 ECU if I am successful and 100 ECU otherwise.
2. Assume that you state a 30% chance of being successful in the 2nd stage, and the computer picks the winning probability in the chance mechanism as 55%. Which of the following is true?
 - (a) My probability of earning 300 ECU is 55% and my payoff is independent of my performance.
 - (b) My payoff is 300 ECU if I am successful and 100 ECU otherwise.
3. Suppose Person A stated 80% as the minimum winning probability that makes her leave her payoff to chance, while Person B stated 45% for the same decision. Which of those people is more self-confident about their success in the 2nd period?
 - (a) A
 - (b) B

B.3 Post-experiment Survey Questions

1. How old are you?
2. What is your gender?
3. Which year of your degree program are you in?
4. What is your major/faculty?
5. What is your current GPA?
6. Were the rules of the experiment clear and understandable? Please answer on a scale of 1 to 10: 1= not understandable at all, 10= extremely understandable.
7. How difficult was the addition task? Please answer on a scale of 1 to 5: 1= not difficult at all, 5= extremely difficult.
8. How difficult was the verbal task? Please answer on a scale of 1 to 5: 1= not difficult at all, 5= extremely difficult.
9. As agents, when you were making the decision about performance/chance mechanism, was the thought of “if I don’t do the performance task in the 2nd period, I will get bored” a factor in your decision? (Yes/No)
10. Disregarding the monetary payoff it brings, how important was it personally for you to be “Successful” (to score higher than the target score)? Please answer on a scale of 1 to 5: 1=“Not important at all”, 2= “Not important”, 3= “Neither important nor unimportant”, 4=“Important”, 5= “Extremely important”