Giving Feedback to Clients

Teck-Hua Ho and Catherine Yeung[1]

April 27, 2013

[1] Authors are listed in alphabetical order. We thank the editor, the associate editor, and 3 anonymous reviewers for their helpful comments and suggestions. Taizan Chan developed the web-based software for data collection and Xing Zhang provided superb research assistance. Direct correspondence to any of the authors: Ho: University of California at Berkeley and National University of Singapore. Yeung: National University of Singapore. Email: Ho: hoteck@haas.berkeley.edu; Yeung: cyeung@nus.edu.sg.

Giving Feedback to Clients

Abstract

Client-agent interactions are prevalent in our daily lives. In this research, we study one such interaction through an economic game that captures the essence of clients' and agents' dealings with one another: the fact that agents, whose pay is frequently determined by client happiness, have the incentive to maximize client happiness by giving distorted feedback. In this game, a client undertakes a nontrivial task and her performance is disclosed to an agent. The agent must then give the client feedback to inform her of her performance. Upon receiving the agent's feedback and before knowing her true performance, the client reports her happiness level, which in turn determines the agent's monetary payoff.

In a series of eight studies involving a total of 844 subjects, we systematically vary the way the agent's cash earnings depend on the client's happiness, but we always compensate the client based on her true task performance. In Study 1, we make the agent's cash earnings proportional to the client's happiness. We find that the agent inflates feedback, and the client reports a higher level of happiness than that reported in a control condition where the agent always provides honest feedback. As a result, the agent ends up making more money (i.e. inflated feedback pays). In Study 2, we show that the client reports the same higher level of happiness even if she reports her happiness and her preferred payment for the agent on two separate measurement scales. Studies 3 and 4 show that neither the agent nor the client behaves altruistically in their reporting. Study 5 shows that the client reports a higher level of happiness because she over-estimates her own task performance and mistakenly believes that the agent's inflated message is genuine, leading to higher cash earnings for the agent. Study 6 shows that the agent stops inflating feedback when his cash earnings are made proportional to the client's ex-post happiness (i.e., happiness *after* the client finds out her true performance), suggesting that the agent is opportunistic. Studies 7 and 8 show that the main findings extend to client-agent interactions where agent feedback is categorical (e.g., good versus bad) and where agent feedback is consequential in that it can directly affect the client's payoff in a subsequent task. Together, the eight studies provide strong evidence that the agent behaves opportunistically and this strategy works because the client is overconfident of her own performance.

Keywords: opportunism, altruism, overconfidence, feedback giving game, client and agent interaction, behavioral economics

1. Introduction

Giving and receiving feedback are central to social interactions. People regularly give and receive feedback. For example, as professors, we give feedback to our students about their class performance and we receive feedback from them about our teaching. Giving and receiving feedback are also very important forms of communication because they can have a significant emotional impact on the recipients who receive feedback. Positive feedback can affirm the worth of the recipients and make them feel happy and competent. Negative feedback frequently produces negative emotions and may even generate a sense of incompetence among certain individuals. In this research, we develop a simple feedback giving game to study how people strategically manage happiness in social interactions.

Specifically, we investigate client-agent interactions in which the agent is in a position to provide feedback to the client based on some private information the former has about the latter's performance[2]. Since the client cares about her task performance, the agent's feedback can significantly affect the client's happiness. The agent wants to make the client happy because his cash earnings will increase with the client's happiness.[3] There are many real-world examples of such social interactions.[4] Consider, for example, a subordinate who must give his boss feedback about the latter's performance in an inaugural speech. The subordinate cares about the boss's happiness because she is empowered to determine his payoff through performance evaluation. By putting himself in her shoes, he can imagine that she will experience unpleasant feelings upon receiving negative feedback. To avoid dealing with her negative emotions, he may instead choose to provide inflated feedback and compliment her. Clearly, the boss is also keenly aware of his incentive to sugarcoat feedback. As such, she may not believe his compliment

[2] For consistency and brevity, we use "she" to refer to the client and "he" to refer to the agent.

[3] Andrade and Ho (2007, 2009) show that an individual whose payoff is dependent on her partner's decision often believes she will receive better treatment from a happier partner. This clearly explains why individuals are motivated to manage their partners' happiness when opportunities arise. When such opportunities involve giving and receiving feedback, the feedback message will naturally serve as a platform for the strategic management of happiness.

[4] Prior research has documented such client-agent interactions in a wide range of organizational settings (see Gordon 1996 for a review). People not only strategically please others in a high-status supervisory position by providing pleasant feedback, they also do the same to those in a low-status subordinate position, especially when they are dependent on the latter's specialized expertise. A similar kind of phenomena occurs in market settings too. For example, sales people please their customers to get the latter to buy products from them. In these examples, agents interact with their clients either only once or infrequently so that the specific feedback-giving episode is like a one-shot interaction (e.g., a real estate agent selling a house to a client). Our research is particularly relevant to interactions of this type, which involves one-shot feedback giving and receiving interaction.

and this potential loss of credibility may stop him from inflating feedback. Hence, it is unclear, ex ante, whether the subordinate will inflate feedback, and whether the boss will believe in the provided feedback. We conceptualize such communication as a client-agent interaction and investigate whether feedback inflation occurs; if so, whether it pays; and if it pays, why it does.

We develop a new *feedback giving game* to capture the above client-agent interaction. We simultaneously examine whether the agent inflates feedback and how the client responds to the agent's message, when it is common knowledge that the client's reported happiness can directly influence the agent's cash earnings. In this game, the client undertakes a nontrivial task (solving 10 challenging math problems) which earns her $1 for each question answered correctly. The client's performance is privately made known to the agent. The agent must send a feedback message ("You have got [ ] correct answers.") to the client. Upon receiving the message and before knowing her true performance, the client must indicate her level of happiness on an 11-point scale that ranges from 0 (not happy) to 10 (very happy). The agent earns 50 cents for each incremental point reported on this scale, so that the more happiness the client reports, the more money the agent makes. This game captures two important characteristics of client-agent interactions. First, it is clear to the agent that his payoff is determined by the client's happiness. Second, it is clear to the client that the agent may have an incentive to inflate feedback and, as a consequence, she may have to calibrate her emotional reaction accordingly. In a series of eight studies, we systematically manipulate the various features of this feedback giving game in order to study the agent's strategic attempt at managing the client's happiness, the client's anticipation of that strategic attempt, and how the agent's attempt and client's anticipation interact to influence the agent's feedback and the client's happiness.

In Study 1, we run the experimental condition with the agent's cash earnings increase linearly with the client's reported happiness and the control condition where the agent is incentivized to always provide honest feedback. The study generates two empirical regularities: 1) the agent inflates feedback and 2) the client reports a higher level of happiness when the agent's payoff is tied to the client's happiness. As a consequence, the agent earns a higher monetary payoff. That is, inflated feedback pays. Before we proceed to explain the two empirical regularities, we conduct a new study to ensure that the tying of the agent's payoff to the client's happiness does not lead to any distortion of the latter. In Study 2, the client is either asked to report a single combined measure of happiness (and payment for the agent) as in Study 1 or to report two separate measures (one about her happiness and one about the agent's

payoff). Since the client reports the same level of happiness in both cases, we conclude that combining the two measures does not distort the way the client reports her happiness.[5]

There are at least two competing explanations for the two empirical regularities identified in Study 1 and 2. First, the *altruism hypothesis* posits that the agent inflates feedback and tells a white lie to make the client happy. The client also reports a higher happiness level simply to help the agent make more money (especially when it does not cost her materially to do so). Second, the *opportunism-overconfidence hypothesis* states that the agent inflates feedback in order to make more money. The client is overconfident in estimating her own performance and, as a consequence, mistakenly believes that the (inflated) agent feedback is genuine and reports a higher level of happiness. We investigate these two hypotheses using the same feedback giving game but change the way the agent is compensated.

Study 3 tests the *altruism hypothesis* by first examining the client's altruistic motive in two experimental conditions. In the first condition, the agent's cash earnings increase linearly with the client's reported happiness; in the second condition, the agent's cash earnings are also proportional to the client's happiness but this payment rule applies only if the agent provides truthful feedback. If the client is purely altruistic, she will report the same level of happiness in both experimental conditions (since her report is mainly to make the agent money). On the other hand, if she is simply responding to the agent's feedback, she will report a lower level of happiness in the second condition, because the feedback she receives from the agent will be less positive. Since we observe a lower level of reported happiness in the second condition, we conclude that the client's reported happiness reflects her genuine feeling towards the feedback and her reporting is not driven by an altruistic motive to make the agent more money. Study 3 also examines whether the agent inflates feedback because he wants to tell a while lie to make the client happy. In a third experimental condition, we pay the agent a fixed fee for providing feedback (i.e., he receives the same payment regardless of whether his feedback is truthful or not and the level of payment does not depend on the client's happiness). If the agent is altruistic and wants to tell a white lie to make the client happy, he should continue to do so in this experimental condition. We find that the agent does not inflate feedback in this experimental condition. Overall, Study 3 rules out the altruism hypothesis.

To further examine the altruism hypothesis, Study 4 rules out the *conditional altruism* hypothesis that the client reports a higher level of happiness to make the agent more money (i.e., be altruistic) only if she is ahead of the agent in cash earnings. We run an experimental condition where the client is paid only \$0.5 for each correctly answered math question but the agent is paid \$1 for each happiness scale point. Under this revised payment scheme, the client is behind in terms of cash earnings. Since the client reports

[5] We thank an anonymous review for suggesting Study 2. Findings of Study 2 also imply that the empirical regularities are relevant as long as the agents believe that making the clients happy will increase their current or future payoffs.

the same higher level of happiness as before, we rule out the explanation that the client reports a higher level of happiness only if she is ahead of the agent.

Studies 5 and 6 test the *opportunism-overconfidence* hypothesis. Study 5 examines why the client feels genuinely happier, despite her keen awareness of the agent's incentive to make her happy. In the study, we measure the client's estimate of her own performance after she has taken the task but before she finds out her true performance. We find that the client overestimates her own performance. In addition, the agent's feedback and the client's own performance estimate are remarkably close. These results explain why feedback inflation leads to a genuinely happier client and a higher payoff for the agent—that is, because the overconfident client mistakenly believes that the agent's inflated feedback is truthful.

Study 6 provides direct evidence that the agent is opportunistic when he provides inflated feedback. Specifically it tests whether the agent will stop inflating feedback when the opportunity to take advantage of the overconfident client is eliminated. In the first five studies, the agent can benefit from the client's overconfidence because he is compensated based on the latter's happiness reported *before* she finds out her true performance. In study 6, we make the agent's cash earnings proportional to the client's ex-post happiness (her report of happiness *after* she has found out her true performance). Here, the agent stops inflating feedback. Hence, we conclude that the agent is indeed opportunistic in providing feedback.

Studies 7 and 8 generalize the result by showing that the agent inflates feedback in two other prevalent client-agent interactions[6]. In study 7, we examine a setting where agent feedback must be categorical (e.g., "bad" versus "good" performance) such that feedback inflation entails sending a message that has a literally opposite meaning. In study 8, we examine a setting where agent feedback is consequential in that it can influence the client's decision of how she wishes to be compensated (i.e. a fixed fee independent of performance versus a fee proportional to performance) in an identical task performed right after receiving agent feedback. If the agent inflates feedback, the over-confident client might mistakenly choose to be paid based on performance, which could mean a lower payoff for the client. Both settings make it more difficult (morally) for the agent to justify feedback inflation. Nevertheless, we continue to observe feedback inflation in both settings.

The organization of the rest of the paper is as follows. Section 2 describes the feedback giving game. Section 3 describes Studies 1 and 2 and summarizes the main empirical regularities. Section 4 develops the altruism and opportunism-overconfidence hypotheses and discusses how they can be tested within our experimental paradigm. Section 5 describes Studies 3 and 4, which serve to rule out the

[6] We thank a reviewer for making this excellent suggestion. These studies make our result relevant to many more practical client-agent interactions.

altruism hypothesis. Section 6 describes Studies 5 and 6, which test and support the opportunism-overconfidence hypothesis. Section 7 reports Studies 7 and 8, which show that the agent inflates feedback in two other prevalent client-agent social interactions. Section 8 discusses the implications of the results, highlights the research's limitations, and suggests future research directions.

2. The Feedback Giving Game

The feedback giving game consists of 3 separate stages, namely a task performance stage, the feedback communication stage, and an emotional reaction stage. In the task performance stage, each participant is asked to solve 10 SAT-type mathematics problems. The task is consequential because every correct answer earns the participant $1 so that he or she has a potential to earn up to $10 from the mathematics quiz. The participants will only know their true performance and receive payment after the emotional reaction stage is over.[7]

In the feedback communication stage, participants are randomly matched in pairs and are either assigned the role of the agent or the client. The agent in the dyad first receives information about the client's actual performance on the mathematics quiz. The agent will then send feedback to the client anonymously and electronically. The possible feedback messages are "You have correctly answered [y] questions," where $y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Let the client's actual number of correct answers be x. The agent's feedback is deflated if $x > y$, truthful if $x = y$, and inflated if $x < y$.

In the emotional reaction stage, the client responds to the feedback message by reporting her happiness on an 11-point scale ranging from 0 (not happy at all) to 10 (very happy). Note that the feedback message sent by the agent is the only information the client receives before she reports her happiness. She does not know the actual number of correct answers, x, until the experiment ends. We repeat the feedback communication and emotional reaction stages but match each subject with a different partner. Our experimental design ensures that each participant assumes the role of the agent and the client once and that the order of role assignment is determined randomly. Figure 1 depicts the 3-stage experimental paradigm.

[Insert Figure 1]

[7] This is the case for all the studies, except for Study 6, in which clients are informed of their actual performance before they report their happiness.

Note that the same feedback giving game paradigm is utilized in all of the eight studies reported in this research. Table 1 provides an overview of these studies and their corresponding payment schemes. By simply employing different incentive schemes to compensate the agent or making simple modifications to the procedure, we are able to establish empirical regularities of feedback giving and receiving (Studies 1 and 2), test competing explanations (Studies 3-6), and extend our main findings to two other agent-client interactions (Studies 7 and 8). Thus, our feedback giving game provides a simple yet rich experimental paradigm for studying strategic interactions between the agent and the client. For further discussion of why simple games are particularly useful for studying social interactions in experimental settings, see Camerer (2003).

[Insert Table 1]

3. Establishing the Empirical Regularities

3.1 Study 1

One hundred and four undergraduates participated in this experiment. They were paid a show-up fee plus additional cash earnings contingent on their performance in the experiment[8]. Participants on average spent about one hour and earned $13. They were randomly assigned to either one of the two experimental conditions: the proportional condition (in which the agent was paid based on the client's happiness) or the truth-telling condition (in which the agent was paid $3 whenever he reported the truth). For brevity, we refer the former as the "prop condition" and the latter, the "truth condition". Both conditions followed an identical procedure except for the way the agent was paid. The appendix provides the experimental instructions for the prop condition.

Procedure

Participants were randomly seated at one of the cubicles equipped with a computer. They were informed that they would engage in a mathematics quiz and a feedback communication task, and their payment would depend on their actual performance on these two tasks. Participants were told explicitly that all communication during the experiment would be made through a web-based software program, such that interactions between participants could only occur electronically and anonymously.

Participants first engaged in a mathematics quiz. They were given 15 minutes to solve 10 SAT-type math problems. The difficulty of the problems was calibrated such that subjects on average answered

[8] Cash payments were paid in Singapore dollars (1SGD = 0.8USD) in all studies reported in this paper.

about five problems correctly. This design provided room for the agent to inflate or deflate the client's performance in feedback communication. Participants were informed that every correct answer would earn them $1. Their answers were graded electronically by a software program, but participants had no access to their actual performance until the end of the experiment.

Participants were randomly matched in pairs and engaged in feedback communication about each other's quiz performance. They were informed that they would engage in two rounds of feedback communication and would be paired with a different participant in each round. In one round, they would act as the agent by providing feedback, and in the other round they would act as the client by receiving feedback and reporting their own level of happiness. The order of their role assignment was determined randomly so that half the subjects acted as the agent first and the other half as the client first. The agent then received a report showing his client's performance in the mathematics quiz. He was required to send feedback to his client by completing the statement "You have correctly answered [ ] questions." Upon receiving the message, the client indicated how happy she was on an 11-point scale ranging from 0 (not happy at all) to 10 (very happy). This ended the first round of feedback giving. The second round then took place with each participant matched with a different partner and assumed a different role. When the second round ended, participants were informed of their total payment from the mathematics quiz and the feedback communication task. Participants were paid in cash before they left the experiment.

Payment scheme. In the prop condition, agents were paid based on their clients' reported happiness; each higher scale point yielded an additional 50 cents. In the truth condition, agents received $3.0 for truthful reporting and received nothing ($0) for untruthful reporting.

To ensure that participants had fully understood the payment scheme and the experimental instructions, we administered a short test and required participants to answer all the questions correctly before they could proceed to the feedback giving game. All participants passed the test.

Results and Discussion

Manipulation and confound checks. We pooled the data from the two rounds of feedback communication in our analysis of the agent's feedback[9] and the client's happiness[10].The top panel of

[9] We conducted the following analyses to verify whether the agent's feedback differed as a function of the order of role assignment (i.e., whether one played the agent role first or the client role first). For each study, we ran an ANOVA with the agent's feedback as the dependent variable and the order of role assignment and the experimental conditions as the independent variables. The p-values for the main effect of role assignment order ranged from .27 to .93 (average = .57) across all the studies.

Table 2 summarizes the main results. Participants in the two experimental conditions did not differ in their actual quiz performance (M_{truth} = 4.88 and M_{prop} = 5.25; $F(1, 102) < 1$). All agents in the truth condition reported truthfully. Therefore, we could use the truth condition as a no-inflation benchmark for comparison.

Feedback Inflation. In the prop condition, 59.4% of the agents inflated their feedback and the remaining agents honestly reported. The average feedback in the prop condition was 6.97, which was 2.09 units higher than that in the truth condition (M_{prop} = 6.97 vs. M_{truth} = 4.88; $F(1, 102) = 25.0$, $p < .001$). A scatter plot of the agent's feedback against the client's actual performance in the prop condition is shown in Figure 2. If the agents report truthfully, the data points will fall on the 45^0 dotted line in Figure 2. As shown, a majority of the points are above the line, suggesting that the agents inflated their feedback.

[Insert Figure 2 here.]

Client's Reported Happiness. Clients in the prop condition reported a higher level of happiness than their truth condition counterparts (M_{prop} = 6.25 vs. M_{truth} = 4.18; $F(1, 102) = 11.60$, $p = .001$). This higher reported happiness resulted in an average payoff of $3.13. This result implies that feedback inflation pays, because if the agent had truthfully reported, the client would have indicated a lower happiness rating of 4.18 (as in the truth condition). This would have resulted in a payoff of $2.09, 33% lower than the actual payoff of $3.13.

Study 1 generates two empirical regularities: 1) the agent inflates feedback and 2) the client reports a higher level of happiness when the agent's payoff is tied to the client's happiness. Before we proceed to examine the explanations for these empirical regularities, we must first verify whether the tying of the agent's payoff to the client's reported happiness has changed the way the client reports her happiness. Specifically, does the client report her happiness the way she normally does when she knows

---

The p-values for the interaction effect ranged from .13 to .98 (average = .66) in all the studies. Hence, we pooled the data from the two rounds of feedback communication in all our analysis of the agent's feedback.

[10] Similarly, for each experiment, we ran an ANOVA with the client's reported happiness level as the dependent variable and the order of role assignment and the experimental conditions as the independent variables. The p-values for the main effect of role assignment order ranged from .25 to .99 (average = .71) across all the studies. The p-values for the interaction effect ranged from .13 to .93 (average = .54) in all studies except for Study 2, where a marginally significant interaction effect was observed (p = .06). To check whether the order of role assignment affected the main results in Study 2, we also conducted statistical tests with role assignment order added to the statistical models. These additional analyses yielded the same results as those reported in the paper. For the sake of consistency and brevity, we pooled the data from the two rounds of feedback communications in all our analyses of the client's happiness.

that the agent's payoff is pegged to her response? The answer to this question would facilitate a more precise interpretation of subsequent findings on client happiness.

3.2 Study 2

A total of 136 participants participated in this study and were randomly assigned to one of the following three experimental conditions: the prop condition, the truth condition, and a new "decoupled condition". The decoupled condition was the same as the prop condition, except that the client was asked to report two separate measures (one about her own happiness and another about the agent's payoff) instead of a single combined measure. Specifically, the client first reported her happiness on an 11-point scale (0 = not happy at all; 10 = very happy) and then indicated how much the agent should be compensated on a separate scale ranging from 0 to 10, knowing that each scale point would earn the agent 50 cents. As the client's reporting of happiness is decoupled from her determination of the agent's payoff, her reporting of happiness level should be free from consideration of the agent's payoff. A comparison between this "decoupled" happiness level and that observed in the prop condition should reveal whether the tying of the agent's payoff to the client's happiness rating has led to any distortion of the client's reported happiness.

Results

The second panel of Table 2 shows the experimental results. Again, the Math quiz performance did not differ across the three conditions ($F < 1$). Before comparing the client's happiness rating in the decoupled and prop conditions, we first had to ensure that clients in the two experimental conditions had received the same level of feedback from their agents and that the average feedback in the two conditions was higher than that in the control condition. Indeed, agent's feedback in the prop and decoupled conditions was statistically not different ($M_{prop} = 6.00$ and $M_{decoupled} = 6.35$; $F < 1$, $p = .43$). Moreover, the average feedback in the two conditions was more positive than that in the truth condition (6.18 vs. 5.27; $F(1, 133) = 5.01$, $p < .05$). The reported happiness did not differ between the decoupled and the prop conditions ($M_{prop} = 6.14$ and $M_{decoupled} = 6.35$; $F < 1$, $p > .50$) and clients in these two conditions reported being happier than their counterparts in the truth condition (6.25 vs. 4.63; $F(1, 133) = 9.29$, $p < .05$). In combination, these findings suggest that tying the client's happiness to the agent's payoff did not distort the client's report of happiness.

When the clients in the decoupled condition were asked to determine how much their agents should be compensated, their average rating was 7.73, earning the agents a higher payoff than those in the

prop condition ($M_{decoupled}$ = \$3.87 vs. M_{prop} = \$3.07; $F(1, 94) = 7.11$, $p < .01$). Note that the clients in the prop condition could have made their agents the same amount of money simply by reporting a higher level of happiness. However, these agents in the prop condition did not do so.

In summary, the findings so far suggest that when the agent's payoff is proportional to the client's happiness, the agent inflates feedback and this feedback inflation results in a happier client and higher earnings for the agent. This result raises two fundamental questions. First, for the agent, why did he inflate feedback? Second, for the client, why did she react favorably towards the feedback even though she was keenly aware that the agent had an incentive to inflate? To answer these questions, we review past research that bears relevance to the strategic interaction involved in the feedback giving game. We then develop alternative hypotheses to be tested in Studies 3-6.

4. Understanding Feedback Giving and Receiving

There are two possible theories underlying the empirical regularities established in Studies 1 and 2. The first theory assumes that individuals are *opportunistic* and their behaviors are guided primarily by self-interests. As a consequence, they choose actions that maximize their own payoffs. The second theory assumes that people are *altruistic* and care about others' welfare. They derive utility when others' welfare improves and hence are willing to choose actions in order to increase it (Andreoni 1990; Sugden 1982). Below we elaborate on how these two theories could account for the main empirical regularities. Note that these theories can lead to distinct predictions when different payment schemes are used for the agent. Indeed, we systematically vary the payment scheme in our subsequent experiments in order to separate the two stories and determine which theory is better at explaining the main empirical regularities.

4.1 Opportunism-Overconfidence hypothesis

The agent may have acted opportunistically in inflating the feedback. If the agent anticipates that the client will report a higher level of happiness when she thinks she performs well in the math quiz, it is in the agent's self-interest to enhance the feedback in order to please the client and earn more money. However, the client is keenly aware of the agent's incentive to please her. As a result, if she receives favorable feedback, she may not necessarily believe in it and feel happy about it. Thus, whether or not the client believes in the inflated feedback determines whether the agent's feedback inflation will work.

Research in psychology suggests that people generally have a desire for favorable feedback because such feedback enhances their self-conception and makes them feel good about themselves (Sedikides 1993; Taylor and Brown 1988). They over-weight positive feedback relative to negative even in situations where they are incentivized to make an accurate judgment of the self (Eil and Rao, 2011; Mobius, Niederle, Niehaus, and Rosenblat, 2011). People's positive self-conception also leads them to be overconfident about their performance in challenging tasks (Moore and Healy 2008). For example, one may believe one has answered five questions correctly when in fact only four are answered correctly[11]. In the feedback giving game, if the client who receives favorable feedback has also over-estimated her performance, she may come to believe that the feedback is truthful, despite her knowing that the agent may be opportunistic. Thus, not only does the client desire favorable feedback, but also she may truly believe that such feedback is genuine. Moreover, the closer the degree of feedback inflation matches with the level of overestimation, the more the client will believe in the feedback and feel genuinely happy about the (inflated) performance.

The agent's opportunism and the client's overconfidence together can explain why the agent inflates feedback and why inflation works to increase the client's happiness. Specifically, the agent opportunistically inflates his feedback message in order to make the client happy and earn more money for himself. This strategy works because the client is overconfident and mistakenly believes the feedback message is genuine and hence reports a higher level of happiness. We refer to this theory as the *opportunism-overconfident* hypothesis.

4.2 Altruism Hypothesis

The agent and the client might have acted altruistically in the feedback giving game. The agent might have inflated feedback because he wanted to make the client happy and, similarly, the client might have reported a higher level of happiness because she wanted the agent to make more money. An extensive literature on dictator games provides compelling evidence in support of this altruism hypothesis. In a standard dictator game, a dictator determines an allocation of a fixed pot of money between herself and another person (an anonymous responder). Since the responder cannot reject the decision made by the dictator, the dictator has the absolute power in dividing the pot of money in any way she pleases. If the dictator is purely self-interested, she will maximize her own payoff by making the

[11] Overestimation of own performance has been distinguished from other forms of overconfidence, such as overplacement (one believes one is better than others and overplace one's own performance relative to those of others) and overprecision (one is overly certain of one's own performance) (Moore and Healy 2008). People may also exhibit other types of errors and biases in estimating one's performance; see Kruger and Burrus, 2004; Moore and Small 2007, and Moore et al. 1999.

smallest possible offer to the responder (i.e., $0). Research however shows that the dictator frequently gives about 20% of the pot to the responder, which in effect reduces her own payoff by 20% (Forsythe et al. 1994). Indeed, many dictator game experiments report positive dictator offers (see Camerer 2003 for a comprehensive review). This result suggests that people care about others' payoff and are willing to give others money even at the expense of their own payoff.

We reason that if altruism takes place even if people have to bear a direct material cost (as in the dictator game), it should occur more readily when people do not have to bear any material cost to be altruistic. This is potentially the case in the feedback giving game on the part of the client. When the client knows that she can earn the agent more money by reporting a higher level of happiness, at no financial cost to herself, she may simply do so to benefit the agent.

It is also possible that the client has this other-regarding preference *only* when she is ahead of the agent. Specifically, when the client thinks that she is going to make more money than the agent (which is likely in our feedback giving game), she may choose to be altruistic by reporting a higher level of happiness in order to benefit the agent[12]. We refer to this altruistic preference as *conditional altruism*, because the client's altruism is conditional on the premise that she is ahead of the agent (i.e., she makes more money than the agent). This conditional altruism is different from the *unconditional altruism* discussed above in that the latter form of altruism can occur without the condition on the client's advantageous payoff position over the agent.

On the side of the agent, his inflation of feedback can also be driven by an altruistic motive. If the agent anticipates that telling the client her true performance may cause her to be unhappy, he may wish to sugarcoat the feedback in order to make the client happy. However, this inflated feedback can be costly because the client may take the supposedly good intention wrongly as the agent's self-interested strategy to make himself more money. As a consequence, this altruistic act may cause the agent to receive a lower payoff from a suspicious client. Nevertheless, research findings show that people sometimes are willing to tell an altruistic white lie that improves others' welfare even if their white lie may lower their own payoff slightly (Erat and Gneezy 2012; see also Gneezy 2005). Also, the robust finding of dictator games discussed above suggests that people are willing to make a sacrifice if they believe their action will benefit others. Hence, the agent's decision to inflate feedback might also be driven by an altruistic motive to make the client happier, even if this altruistic act may be costly to implement.

[12] See Bolton and Ockenfels (2000) for a review of related works on equity and distribution preference.

Both the altruism and the opportunism-overconfidence hypotheses offer compelling explanations for the agent's decision to inflate feedback and the client's favorable response to such feedback. In Studies 3 to 6, we systematically vary the agent's payment scheme in order to determine which hypothesis best explains the observed empirical regularities.

5. Testing the Altruism Hypothesis

5.1 Study 3

Study 3, designed to test the altruism hypothesis, consisted of 3 experimental conditions. We have the usual prop condition like before. To test whether the agent inflated his feedback because he simply wanted the client to be happier, we ran a new experimental condition in which the agent was paid a flat fee of $3 for any feedback he provided. In the flat-fee condition, the agent could choose to give honest feedback, inflated feedback, or deflated feedback, without having to worry about payoff consequences. In other words, the agent had the complete freedom to influence the client's happiness through the feedback he provided, while keeping his payoff constant. If feedback inflation was indeed driven by an altruistic motive to make the client happier, the agent should at least be as, if not more, likely to inflate his feedback in this "flat-fee" condition as in the prop condition.

We also studied whether the client reported a higher level of happiness because she was genuinely happier or because she wanted to make the agent more money. To do so, we ran a "truth-telling, proportional" condition ("truth-p condition" hereafter) in which the agent was paid based on the client's reported happiness but he would receive this payoff only if he truthfully reported. This design eliminated the agent's incentive to inflate feedback, while allowing the client to be altruistic by reporting a higher happiness level and hence making the agent more money. If the client was genuinely happier as a result of receiving inflated feedback in the prop condition, the level of reported happiness would decrease in the truth-p condition when the feedback became less favorable. We would not expect such a decrease in reported happiness if the reporting was solely driven by an altruistic motive to make the agent more money. In fact, in this latter case, one could even argue that a higher reported happiness could arise in the truth-p condition because the client was more certain that the agent was truthfully reporting.

One hundred and forty participants were randomly assigned to one of the three conditions: 1) the prop condition, 2) the flat-fee condition, and 3) the truth-p condition. The procedure was similar to that employed in Study 1 except that participants in the two latter conditions were told differently about the ways they would be paid.

Results

Findings of Study 3 are shown in Table 2. As expected, the Math quiz performance did not differ across the three conditions ($F < 1$). Next, we reported two sets of statistical comparisons that tested the altruism hypothesis. We first compared the prop and the flat-fee conditions to examine altruistic behavior in the agent, and then compared the prop and the truth-p conditions to investigate altruistic behavior in the client.

Are Agents Altruistic in Giving Feedback? In the prop condition, 48.3% of agents inflated their feedback and the remaining 51.7% truthfully reported. In the flat-fee condition, 83.3% of agents truthfully reported, 8.3% of agents inflated feedback, and remaining 8.3% of agents deflated feedback. Feedback was less favorable in the flat-fee condition than in the prop condition ($M_{\text{prop}} = 6.47$ vs. $M_{\text{flat-rate}} = 4.83$, $F(1, 82) = 10.65$, $p < .001$) and, in fact, the feedback provided in the flat-fee condition was not statistically different from the client's actual quiz performance ($M_{\text{feedback}} = 4.83$; $M_{\text{quiz}} = 4.75$; $t(23) = .62$; $p > .50$). Thus, the agents stopped inflating feedback when their monetary payoff was not tied to the clients' happiness. Consequently, we conclude that the agents' decision to inflate feedback was not driven by an altruistic motive to make the clients happy.

Are Clients Altruistic in Responding to Agent Feedback? As expected, no agent inflated in the truth-p condition. The average feedback was 4.93 in the truth-p condition, which was statistically lower than that in the prop condition of 6.47 ($F(1, 114) = 17.97$, $p < .001$). Correspondingly, the client reported being less happy in the truth-p condition than in the prop condition (4.91 vs. 6.33; $F(1, 114) = 6.01$, $p = .01$). If the clients' happiness reporting had been driven by their preference to make the agents more money and not by how they genuinely felt, the reported happiness level would not have decreased in the truth-p condition. However, this was not the case. The client reported a lower happiness level when the feedback she received was less favorable (yet truthful), despite her knowing that a lower level of reported happiness would mean a lower payoff for the agent. Hence, we rule out the explanation that the client has an unconditional desire to improve the agent's welfare.

The client's preference to improve the agent's welfare may not be unconditional, but it may depend on whether she is ahead of the agent in terms of cash earnings. We next report Study 4 that examines this conditional altruism hypothesis.

5.2 Study 4

The conditional altruism hypothesis states that the client is altruistic only if she is ahead of the agent in terms of cash earnings. When she is ahead, she may want to increase the agent's payoff by reporting a higher happiness rating so that the agent will not be in a relatively disadvantaged position in terms of payoff. We tested the conditional altruism hypothesis by putting the client in a relatively disadvantaged position in terms of payoff and by examining how the client's happiness reporting would change as a result. Study 4 consisted of two experimental conditions: 1) the prop condition like before and 2) the new "swap payoff" condition, in which the client was paid only \$0.5 for each correctly answered math question but the agent was paid \$1 for each incremental happiness point reported by the client. If the client indeed cares about how her payoff compares to the agent's, as the conditional altruism hypothesis would suggest, we would observe a decrease in the client's happiness rating in the swap payoff condition.[13] However, if the relative payoff is not considered in her happiness reporting, the client's happiness rating would not vary between the two conditions.

Ninety two participants were randomly assigned to the two experimental conditions. The experimental procedure was identical in the two conditions (and as the previous experiments), except that the client in the swap payoff condition was paid only \$0.5 for each correctly answered question and the agent was paid \$1 for each point of happiness reported by the client.

Result

Findings are presented in Table 2. Clients' math quiz performances did not differ statistically between the two conditions (M_{prop} = 4.33 and M_{swap} = 5.05; $F(1, 90) = 2.47$; $p = .12$). Neither agent's feedback (M_{prop} = 6.02 and M_{swap} = 6.27; $F < 1$; $p > .50$) nor client's reported happiness (M_{prop} = 7.08 and M_{swap} = 6.43; $F < 1$; $p = .34$) differed between the two conditions. Although the payment scheme for the swap payoff condition was in favor of the agent, the client reported the same level of happiness as in the prop condition, suggesting that relative payoff was not a main consideration when the client reported her happiness. Hence we rule out the conditional altruism hypothesis.

Our findings thus far suggest that, when the agent's payoff is tied to the client's happiness, the agent inflates feedback and the feedback inflation results in genuinely happier clients and higher payoffs for the agents. Studies 3 and 4 specifically show that the agent does not tell a white lie to make the client

[13] In fact, the client might even deliberately report an exceptionally low happiness rating because, by doing so, she could make the agent less money and reduce the disparity between her own payoff and the agent's payoff.

happier and the client does not deliberately report a higher happiness rating to make the agent more money. We next report Study 5 and 6 that test the opportunism-overconfidence hypothesis.

6. Testing the Opportunism-Overconfidence Hypothesis

In this section, we report Study 5 and 6 that test the opportunism-overconfidence hypothesis. In Study 5, we explore why feedback inflation pays, despite that the client is keenly aware of the agent's incentive to manage her happiness. In Study 6, we provide direct evidence to validate the hypothesis that the agent is opportunistic and inflates feedback to make more money.

6.1 Study 5

The opportunism-overconfidence hypothesis posits that the agent is opportunistic and inflates feedback in order to increase his cash earnings. The objective of study 5 is to provide insights into why this strategy works. We posit that the client believes in the agent's (inflated) feedback and reports a higher happiness because she overestimates her own performance and this overestimation is roughly the same as the degree of feedback inflation.

Study 5 consisted of two experimental conditions: 1) the prop condition and 2) the truth-telling condition (i.e., the same set of conditions as in Study 1). The experimental procedure and payment schemes for these conditions were identical to those employed in Study 1, except that participants in the prop condition were asked to forecast their actual performance immediately after they had taken the quiz. The prediction task was consequential in that an accurate prediction would earn participants a bonus payment of 50 cents. Hence participants were motivated to forecast accurately.

Results

In the prop condition, 40% of agents inflated their feedback and the remaining 60% reported truthful feedback. Also, as expected, no feedback inflation was observed in the truth condition (i.e., all agents reported truthfully). The average feedback was 6.08 in the prop condition, which was higher than that in the truth condition of 5.07 ($F(1, 78) = 4.07$, $p = .05$). The client's reported happiness was higher in the prop condition ($M_{prop} = 6.28$ vs. $M_{truth} = 4.25$; $F(1, 78) = 8.68$, $p = .004$), resulting in a higher payoff for the agent. These differences could not be attributed to difference in task performance because participants in the two conditions did not differ in their quiz performance ($M_{truth} = 5.07$ and $M_{prop} = 4.93$; $F(1, 78) < 1$).

Performance Estimation and Feedback Inflation. Consistent with the opportunism-overconfidence hypothesis, clients overestimated their math quiz performance. Their estimated number of correct answers was 6.23, which was statistically higher than their actual number of correct answers (M = 4.93; $t(39) = 5.18$; $p < .001$).

We next assessed how close the agent's (inflated) feedback was to the client's forecast of her own performance. At the aggregate level, the agent's average feedback (M = 6.08) was remarkably close to the client's average performance forecast (M = 6.23). To examine whether the agent's feedback matched the client's forecast at the individual level, we computed the difference between the two values for each client-agent interaction. The average difference was .15, which was statistically not different from zero ($t(39) = .42$, $p > .50$). In addition, we fitted a regression line using the client's forecast as the x-variable and the agent's feedback as the y-variable. The best fitted line that passes through the origin (represented by the solid line in figure 3) has a slope of .92 ($t(39) = 17.84$, $p < .001$, $R^2 = .89$).[14] Figure 3 shows a scatter plot of these two variables with the solid line representing the best fitted line and the dotted line representing the 45^0 line. Overall, these findings suggest that the agent's feedback matches almost perfectly with the client's performance forecast, which, in turn, explains why the feedback is considered credible by the client.

[Insert Figure 3 here.]

The opportunism-overconfidence hypothesis suggests that the opportunistic agent inflates feedback only if he thinks that he can benefit by doing so (i.e., the client reports her happiness *before* finding out her actual performance). This implies that the agent will not inflate feedback if he knows that the client reports her happiness *after* finding out her actual performance (i.e., ex-post happiness). Hence, a simple way to confirm whether the agent is indeed opportunistic is to check whether the agent will indeed not inflate feedback if his earnings are proportional to the client's ex-post happiness. In this case, the opportunistic agent's goal is to choose a feedback message that is most likely to lead to the highest level of ex-post happiness. Clearly, the agent should not inflate his feedback if he believes by doing so will

[14] We also examined clients' reactions when agents inflated feedback *beyond* their own forecasts. We ran the following regression with two dummy variables:

$$\text{Client's happiness} = b_0 + b_1 * I_1\,(\text{client's forecast} - \text{feedback}) + b_2 * I_2\,(\text{feedback} - \text{client's forecast})$$

where: $I_1 = 1$ when (client's forecast – feedback) > 0 and $I_1 = 0$ otherwise
$I_2 = 1$ when (feedback – client's forecast) > 0 and $I_2 = 0$ otherwise

The regression results show that when feedback was below expectation, the client's happiness dropped ($b_1 = -.96$; $t(37) = -2.33$, $p < .05$) and that when feedback was above expectation, the client's happiness remained unchanged ($b_2 = .28$; $t(37) = .73$, $p = .47$).

make the client unhappy and report a lower level of ex-post happiness. The agent should truthfully report if he believes the client will appreciate a truthful message. Alternatively, the agent may wish to under-report if he believes that a pleasant surprise will help to increase the client's ex-post happiness. In Study 6, we provide a further test of the opportunism-overconfidence hypothesis by examining whether the opportunistic agent will stop inflating feedback when his earnings are proportional to the client's ex-post happiness.

6.2 Study 6

This study had the two experimental conditions as in Study 1 (prop condition and truth condition), except that the client in the prop condition was asked to report her happiness a second time — once as in the prior experiments and also after she had found out her actual quiz performance — and that the agent was paid based on the ex-post happiness with each happiness scale point earning him 50 cents. We refer to this new condition as the "prop-ex-post" condition. The payment scheme for the truth condition was identical to that in Study 1; that is, the agent received $3 for truthful reporting and received nothing ($0) for untruthful reporting. Ninety six undergraduates participated in this experiment. They were randomly assigned to the prop-ex-post condition and the truth condition.

Results

As shown in Table 2, math quiz results were not statistically different between the two conditions ($M_{\text{prop-ex-post}}$ = 5.66 vs. M_{truth} = 5.03; $F(1, 94) = 1.67$, $p = .20$). All agents in the truth condition honestly reported. In the prop-ex-post condition, a large majority of the agents (67.2%) honestly reported and only a small fraction (9.4%) inflated their feedback. Interestingly, a significant minority (23.4%) deflated (i.e., under-reported). The mean feedback value did not differ between the two conditions ($M_{\text{prop-ex-post}}$ = 5.39 vs. M_{truth} = 5.03; $F(1, 94) < 1$). These results suggest that the agent is indeed opportunistic and choose not to inflate feedback when his earnings are proportional to the client's ex-post happiness.

Client's Reported Happiness. Clients' happiness before she found out her true performance in the two conditions were not statistically different ($M_{\text{prop-ex-post}}$ = 5.09 vs. M_{truth} = 4.34; $F(1, 94) = 1.22$, $p = .27$). Interestingly, clients' ex-post happiness was higher in the prop-ex-post condition than that in the truth condition ($M_{\text{prop-ex-post}}$ = 5.73 vs. M_{truth} = 4.22; $F(1, 94) = 4.83$, $p < .05$). Recall that 23.4% of the agents deflated their feedback. To determine whether feedback deflation increased ex-post happiness and hence the agent's earnings, we regressed ex-post happiness against feedback deflation. There was no evidence that feedback deflation had affected ex-post happiness (beta = .31; $t(62) = 1.04$, $p = .30$). In fact, when the

client's actual math quiz performance was added to the regression as another independent variable, the result clearly showed that it was the actual performance (beta = .56, $t(61) = 3.15$, $p < .01$) and not feedback deflation (beta = .28, $t(61) = 1.03$, $p = .31$) that determined ex-post happiness. In sum, feedback deflation does not lead to higher ex-post happiness.

Discussion

The results of Studies 5 and 6 suggest that feedback inflation is driven by the agents' opportunistic motive: they inflate feedback only if feedback inflation is found to be financially rewarding. They receive a higher payoff when they inflate feedback because their clients mistakenly believe their (inflated) feedback is genuine and hence report a higher level of happiness. This inflated feedback appears credible because the agents' amount of inflation closely matches the degree of clients' overestimation of their own performance.

7. Two Other Client-Agent Interactions

In this section, we report two new studies to show that the main findings generalize to two other client-agent interactions. In these studies, we make slight modifications of the feedback-giving game to make it morally more difficult for the agent to inflate feedback. We examine whether the agent will still inflate feedback for opportunistic gains.

Past research has examined opportunistic acts in self-reporting, such as the over-reporting of one's performance in a task or the over-claiming of the cost of a repair bill in order to earn a higher financial reward (Mazar, Amir, and Ariely, 2008; Schwartzer and Hsee, 2002). People's propensity to misreport is shown to depend not only on the size of the financial reward brought about by misreporting, but also on the ease of morally justifying it. Small magnitude of misreporting falls within the threshold of acceptable dishonesty and is morally justifiable (Mazar et al., 2008). Similarly, inflating information by a few units out of a wide range of possible outcomes can be justified as "stretching the truth" instead of plain dishonesty (Schwartzer and Hsee, 2002). By the same argument, agents in our feedback-giving game might find it justifiable to inflate feedback because they could treat a few units of inflation as truth-stretching rather than dishonest reporting. In study 7, we seek to replicate our findings in situations where there are only two categories of performance—"bad" and "good"—such that the inflation of feedback from "bad" to "good" is outright dishonesty.

The moral cost of feedback inflation may also increase with the extent of negative consequences it can potentially bring to the client. In Study 8, we examine feedback giving in a setting where agent feedback is consequential in that it can influence the client's decision of how she will be compensated (i.e. a fixed fee independent of task performance versus a fee proportional to task performance) in an identical task performed right after the agent's feedback. Specifically, after the client has received feedback from the agent about her performance in the first quiz, she must work on a second math quiz. She will only be compensated for the second math quiz and she must choose between two compensation options: 1) $5 fixed fee regardless of the number of her correct answers and 2) $1 for each correct answer. An agent's honest feedback will therefore help the client choose the option that maximizes her payoff. Inflated feedback, on the other hand, can mislead the overconfident client to choose to be paid based on performance, which might result in a lower payoff for her. In sum, the experimental settings in Studies 7 and 8 make it morally more difficult for the agent to justify feedback inflation. We examine whether the empirical regularities will still hold in these new settings.

7.1 Study 7

116 participants were randomly assigned to one of the two experimental conditions: 1) the prop condition and 2) the truth condition (the same set of conditions as in Study 1). The experimental procedure was similar to that of Study 1 except that the client's math quiz performance was converted into qualitative grades as "good" or "bad" by a web-based software program. This binary categorization applied both to the information the agent received about the client's performance ("Your paired member's performance is ____.") and the feedback he conveyed to the client ("Your performance is ___."). It was common knowledge that performance was coded as "bad" when the number of correct answers ranged from 0 to 5 and that performance was coded as "good" when the number of correct answers ranged from 6 to 10. This information was shown both in the experimental instructions and on the page where the agent read about the client's performance. As in Study 1, the agent was either paid based on the client's happiness (prop condition) or based on whether he reported the truth (truth condition). The client was paid $1 for each math question she correctly answered.

Results

As shown in Table 2, participants' performance in the math quiz did not differ between the two conditions (M_{prop} = 5.33, M_{truth} = 5.79; $F(1, 114) = 1.39$, $p = .24$). All agents reported truthfully in the truth condition but only 67% did so in the prop condition. Of particular interest is the number of agents who

inflated their feedback from "bad" to "good". While none of the agents did so in the truth condition, half of them (18 out of 35) told their clients that their performance was "good" when in fact it was "bad". As a consequence, the client reported being marginally happier in the prop condition than in the truth condition (M_{prop} = 7.22 vs. M_{truth} = 6.14; $F(1, 114) = 3.17$, $p = .08$).

7.2 Study 8

In this study, the client had to perform two math quizzes that were identical in terms of format and level of difficulty (i.e., to finish 10 SAT-type math questions within 15 minutes). The client was compensated only for the second quiz and she had to choose between being compensated (a) $1 for each correctly answered question or (b) fixed fee of $5 for simply undertaking the second quiz. The payment choice was made after the client had finished and received feedback from the agent about her performance in the first math quiz but before the second quiz began.

Eighty participants participated in this study. There was only one experimental condition in this study. All agents were paid based on the level of happiness the clients reported after receiving feedback on the first quiz. Each scale point earned the agent 50 cents.

Results

We first validated the premise that the client used the agent's feedback on the first math quiz as a basis for her choice of payment option in the second math quiz. We ran a logistic regression using the agent's feedback as the independent variable and the client's choice of payment scheme as the dependent variable. The result showed that the probability of choosing the performance-based payment scheme (i.e., $1 for each correct answer) increased with the favorability of the feedback ($\beta = .31$; $\chi^2(1) = 7.27$, $p < .01$). Thus, it was evident that the client indeed used the agent's feedback as a way to decide how she would get paid in the second math quiz.

We next examined whether the agent inflated feedback despite the potential negative consequence it could have on the client's payoff. Out of the 80 agents, 61% of them honestly reported, 38% inflated, and 1% (one agent) deflated. Clients answered an average of 5.51 questions correctly in the first quiz and the feedback they received was inflated by 1.18 units (significantly different from 0; $t(79) = 5.53$, $p < .01$). We are particularly interested in cases where clients answered less than 5 questions correctly. For this group of clients, inflated feedback that went beyond 5 would have been detrimental

because it would have pointed them to the performance-based scheme instead of the more profitable fixed-fee scheme. Our data show that 41 clients answered fewer than five questions correctly in the first quiz, of which 15 of them (37%) received feedback that had been inflated beyond 5. This result clearly shows that the agents continued to inflate feedback in this two-stage game, despite its detrimental effect on the clients' cash earnings. (Other findings on clients' happiness were reported in Table 2; they were consistent with those reported in other studies and therefore are not discussed here.)

8. General Discussion

This paper investigates a prevalent form of client-agent interaction using a feedback-giving game. In this game, the client performs a task and her performance is disclosed to the agent. The agent must formulate feedback to inform the client of her performance. Upon receiving feedback and before knowing her true performance, the client reports her level of happiness, which in turn determines the agent's monetary payoff—the happier the client, the higher the agent's payoff. We investigate whether feedback inflation occurs in this game; if so, whether it pays; and if it pays, why it does.

In eight tightly linked studies, we systematically vary the relationship between the agent's monetary payoff and the level of the client's happiness. We show that agents inflate feedback when their cash earnings increase linearly with their clients' reported happiness. The agents receive a higher payoff when they inflate feedback because their clients mistakenly believe their (inflated) feedback is genuine and hence report a higher level of happiness. This inflated feedback appears credible because agents' amount of inflation closely matches the degree of clients' overestimation of their own performance.

We rule out the altruism hypothesis both on the parts of the client and the agent. The agent does not inflate his message and tells a white lie to make the client happier. The client does not report a higher level of happiness in order to help the agent make more money. Also, we reject the conditional altruism hypothesis that the client is altruistic only if she is ahead of the agent in terms of payoff. Finally, we observe that the agent stops inflating feedback when we make his cash earnings proportional to the client's level of happiness reported *after* she has found out her actual performance. This finding confirms that the agent is indeed opportunistic and will inflate feedback only if he can benefit by doing so (i.e., before the client finds out her true performance). The main empirical regularities are replicated in two

other prevalent client-agent interactions, in which feedback inflation is more difficult (morally) for the agent to justify. Thus, the main empirical regularities are quite robust.[15]

8.1 Does Context Matter?

Our feedback-giving game is specifically designed to understand how people strategically manage others' happiness. Hence, the term "happiness" was explicitly mentioned to the participants in the experimental instruction when we explained the experimental procedure and was also used to label one of the key dependent variables (i.e., report happiness on a scale ranged from 0 = not happy at all to 10 = very happy).[16] It is however worthwhile to check whether the agent would still inflate feedback if happiness is not explicitly mentioned and brought to the participants' attention. To test whether the happiness label matters, we conducted an experiment with the usual prop condition and a new "label-free" condition. In the "label-free" condition, the term "happiness" was removed entirely from the experimental instruction. Thus, the client was not asked to report her happiness but was instead asked to choose a number on a scale that ranged from 0 to 10, knowing that each scale point would earn the agent 50 cents.[17] The agent also knew the client would determine his payoff by reporting a value on a scale, but the scale was not labeled as one that measured the client's happiness. As such, it was left to the agent's own interpretation of whether or not his feedback would affect the client's determination of his payoff and, if yes, how.

Eighty-four subjects participated in this experiment and were randomly assigned to the two experimental conditions (prop and label-free). Our findings showed no difference between the two conditions in terms of math quiz performance (M_{prop} = 4.68, $M_{label\text{-}free}$ = 4.30; F<1). Of the 40 agents in the prop condition, 42.5% inflated feedback and the other 57.5% reported honestly. Of the 44 agents in the label-free condition, 29.5% inflated, 68.2% reported honestly, and one agent underreported. The difference in the proportions of feedback inflation was not statistically significant (42.5% vs. 29.5%; z = 1.24, p = .21). The average feedback was statistically higher than the actual quiz performance in both the

[15] Positive and negative feedback may result in learning and subsequent behavioral changes. Our Study 8 clearly suggests that feedback giving results in behavioral change (i.e., payment choice) on the part of the client. For further discussion on the topic, see Bandura (1971).

[16] Experimental economists prefer to use abstract instructions in order to avoid additional "induced value" associated with the labeling. We have intentionally used the happiness label because we are interested in people's strategic management of others' happiness and we believe this labeling helps to increase the externality validity of our empirical regularities.

[17] We thank a reviewer for pushing us to conduct this study. Note that our feedback giving game is a signaling game. In this game, the agent receives some private information and then sends a message to the client. The client then responds by choosing an action that can affect the agent's monetary payoff. One can potentially derive standard equilibrium prediction in the resulting signaling game.

prop condition (6.43 versus 4.68; $t(39) = 4.45$, $p < 0.001$) and the label-free condition (5.23 versus 4.30; $t(43) = 3.07$, $p < .001$). The average amount of feedback inflation was marginally higher in the prop than in the label-free condition ($M_{prop} = 1.75$ versus $M_{label\text{-}free} = .93$; $F(1, 82) = 2.76$, $p = .10$). These findings reveal that, removing the explicit mention of "happiness" from the experimental instruction does not change the main finding that the agent strategically manages the client's happiness by inflating feedback. Apparently, the client's happiness is so integral to the client-agent interaction that the agent naturally takes it into consideration in formulating his feedback strategy.

Interestingly, we also found removing "happiness" label from the experimental instruction led to a higher payoff for the agent. Specifically, the client reported a higher rating when she was asked to determine the agent's payoff ($M_{label\text{-}free} = 7.19$) than to report her own happiness ($M_{prop} = 5.42$; $F(1, 82) = 7.56$, $p < .01$), despite the less favorable feedback the client received in the label-free condition. It appears that removing the "happiness" context made the client more generous towards the agent (with the experimenter's money). Put differently, the "happiness" label actually makes the client less generous (because the client wants to truthfully report her level of happiness).

8.2 Feedback Giving and Receiving in Real-world Settings

Our research findings are obtained under an experimental setting in which the client's performance is objectively defined and measured, and the true performance is revealed to the client within the time frame of an experimental session. This design makes the incentive mechanism for the client transparent and her task consequential. Interestingly, the design also sets a stringent condition for the agent to distort his feedback. Both the existence of well-defined objective truth and the short elapsed time for this truth to be revealed may make distorting client's performance morally difficult and causing ultimate client disappointment emotionally hard. Our experimental setting therefore provides a conservative examination of feedback inflation. Given that the agent inflates feedback under this setting, one would expect him to be even more likely to do so in real-world settings where there is no well-defined objective truth or when it takes a long time for the truth to materialize, if ever.

When truth is vaguely defined, the vague definition leaves both the agent and the client in a gray area to determine the exact truthfulness of a feedback message. It makes it difficult for the client to verify whether the agent has been honest or dishonest with her. For example, when a salesperson is asked to evaluate the stylistic fit of a product by a potential customer, the customer can never judge whether the

salesperson has been truthful in giving his opinion because there is no objective truth in this regard. Similarly, there is no way for a boss to tell what her subordinate truly thinks when the latter is asked to provide feedback on her personal traits. In both cases, the non-existence of well-defined objective truth reduces negative consequences of feedback inflation on the part of the agent and makes the agent more likely to take advantage of the situation by inflating.

When the truth takes a long time to materialize, ex-ante happiness (i.e., before the truth is known) is more salient and important than ex-post happiness in client-agent interactions. For example, a junior colleague is likely to inflate feedback in assessing a senior colleague's research paper because the true quality of the paper can only be imperfectly determined by a lengthy publication review process. The senior colleague may find out the true quality eventually, but the junior colleague will, meanwhile, enjoy the support of the senior colleague in a forthcoming promotion case. Here, the ex-ante happiness of his senior colleague drives strategic behavior on the part of the junior colleague.

In a related vein, the above examples also illustrate the practical relevance of the incentive scheme of tying the agent's payoff to the client's ex-ante happiness. In the above examples, the ex-post happiness is either non-existent (when there is no well-defined objective truth) or is dominated by the ex-ante happiness in determining the agent's payoff (when there is a long elapsed time for the truth to be revealed). Thus, the explicit linkage between the agent's payoff and the client's ex-ante happiness is highly representative of the manner agents are compensated in real life situations.[18]

8.3 Limitations and Future Research

Our research has several limitations and allows room for possible extensions. First, our research assumes a one-shot client-agent interaction. If the feedback communication stage is repeated several times between the same agent and client, and actual performance is disclosed to the client each time, then one would expect the client to learn about the agent's tendency to inflate feedback and hence will discount it accordingly. As a consequence, feedback inflation might not pay in this repeated setting where there is room for learning and reputation-building. Future research can examine this issue explicitly. Second, the agent in our feedback-giving game has a perfect knowledge of the client's performance. In some client-agent interaction, the agent might only have an imperfect knowledge of the client's actual performance. Thus, it is interesting to examine whether this imperfect knowledge on the part of the agent will increase

[18] This explicit linkage is common knowledge in all our experiments. In situations where the agent's incentive is not common knowledge (i.e. when he knows how he is paid but the client does not know the agent's incentive), the client is likely to believe in the agent's feedback even more.

or decrease the extent of his feedback inflation. Our experimental paradigm can easily be extended to shed light on this practical setting. Specifically, we can inform the agent on how well the client does in only a subset of the math quiz problems (e.g., the number of correct answers in the first 5 of the 10 questions). Third, our agent is allowed to provide only a precise feedback (*x* out of the 10 questions are answered correctly). In practice, feedback can be imprecise or ordinal (e.g., poor, average, good). It will be worthwhile to study how the degree of inflation changes as a result of ordinal feedback. Fourth, it will be useful to generalize the feedback giving game to a setting where there is more than one agent. Here, individual agents who provide feedback to a client must consider how their feedback interacts with others' feedback and how the client would report her happiness towards them when their feedback is either more positive or more negative than those of their peers.

References

Andrade E., T-H Ho. 2007. How is the boss's mood today? I want a raise. *Psychological Science* **18**(8) 668-671.

Andrade E., T-H Ho. 2009. Gaming Emotions in Social Interactions. *Journal of Consumer Research* **36** 539-552.

Andreoni, J. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal* **100** 464-477.

Bandura, A. 1971. *Social learning theory*. New York: General Learning Press.

Bolton, G., A. Ockenfels. 2000. ERC: A Theory of equity, reciprocity, and competition. *American Economic Review* **90**(1) 166-193.

Camerer, C. F. 2003. *Behavioral game theory: Experiments in strategic interaction.* Princeton: Princeton University Press.

Erat, S., U. Gneezy. 2012. White Lies. *Management Science* **58**(4) 723-733.

Eil, D., J. M. Rao. 2011. The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, **3**(2): 114-38.

Forsythe, R., J. L. Horowitz, N. E. Savin, M. Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* **6** 347-369.

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

Gordon, R. A. 1996. Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology* **71**(1) 54-70.

Kruger, J., J. Burrus. 2004. Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology* **40**(3) 332-340.

Moore, D. A., P. J. Healy. 2008. The trouble with overconfidence. *Psychological Review* **115**(2) 502-517.

Moore, D. A., T. R. Kurtzberg, C. R. Fox, M. H. Bazerman. 1999. Positive illusions and forecasting errors in mutual fund investment decisions. *Organizational Behavior and Human Decision Processes* **79**(2) 95-114.

Moore, D. A., D. A. Small. 2007. Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology* **92**(6) 972-989.

Mobius, M. M., M. Niederle, P. Niehaus, T. S. Rosenblat. 2011. Managing self-confidence: Theory and Experimental evidence. Working paper 17014, National Bureau of Economic Research,.

Sedikides, C. 1993. Assessment, enhance, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology*, **65**(2) 317-338.

Schweitzer, M. E., C. K. Hsee. 2002. Stretching the truth: Elastic justification and motivated communication of uncertain information. *The Journal of Risk and Uncertainty* **25**(2) 185-201.

Sugden, R. 1982. On the economics of philanthropy. *Economic Journal*, **92** 341-350.

Taylor, S. E., J. D. Brown. 1988. Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, **103**, 193-210.

Table 1: Experimental Objectives, Experimental Conditions, and Corresponding Payment Schemes across the Eight Studies

Experiments and Objectives	Experimental Conditions and Corresponding Payment Schemes		
Establishing Main Empirical Regularities and Validating Happiness Measure			
Study 1 Establishing Main Empirical Regularities	Proportional: $.50 × client's reported happiness	Truth-telling: $0 if feedback is not truthful $3 if feedback is truthful	
Study 2 Validating Combined Measure of Client Happiness and Agent Payoff	Proportional: $.50 × client's reported happiness	Truth-telling: $0 if feedback is not truthful $3 if feedback is truthful	Decoupled: $.50 × client's reported value on pay-scale
Testing the Altruism Hypothesis			
Study 3 Testing Altruism	Proportional: $.50 × client's reported happiness	Truth-telling, proportional: $0 if feedback is not truthful, $.50 × client's reported happiness if feedback is truthful	Flat-fee: $3 upon completion of the feedback giving task
Study 4 Testing Client's Conditional Altruism	Proportional: $.50 × client's reported happiness	Swap Incentive: $1.0 x client's reported happiness	
Testing the Opportunism-Overconfidence Hypothesis			
Study 5 Investigating Why Feedback Inflation Pays	Proportional: $.50 × client's reported happiness	Truth-telling: $0 if feedback is not truthful; $3 if feedback is truthful	
Study 6 Testing Agent's Opportunism	Proportional, ex-post: $.50 × client's ex-post happiness (i.e., reported after knowing actual performance)	Truth-telling: $0 if feedback is not truthful $3 if feedback is truthful	
Generalizing Results to Two Other Client-Agent Interactions			
Study 7 Giving Categorical Feedback	Proportional: $.50 × client's reported happiness	Truth-telling: $0 if feedback is not truthful $3 if feedback is truthful	
Study 8 Giving Consequential Feedback	Proportional: $.50 × client's reported happiness		

Table 2: Summary of Results (Study 1- Study 8)

	Math Quiz Performance	Agent's Feedback	Client's Reported Happiness
Study 1 (n=104):			
Proportional (n=64)	5.25 (2.22)	6.97 (2.17)	6.25 (3.25)
Truth-telling (n=40)	4.88 (1.92)	4.88 (1.92)	4.18 (2.61)
Study 2 (n=136)			
Proportional (n=44)	4.84 (1.94)	6.00 (2.04)	6.14 (2.65)
Decoupled (n=52)	5.15 (2.35)	6.35 (2.41)	6.35 (2.99)
Truth-telling (n=40)	5.23 (1.99)	5.27 (1.83)	4.63 (2.76)
Study 3 (n=140):			
Proportional (n=60)	4.92 (1.72)	6.47 (2.17)	6.33 (3.26)
Flat-fee (n=24)	4.75 (1.75)	4.83 (1.81)	4.37 (2.67)
Truth-telling, proportional (n=56)	4.93 (1.69)	4.93 (1.69)	4.91 (2.97)
Study 4 (n=92):			
Proportional (n=48)	4.33 (2.14)	6.02 (2.38)	7.08 (3.07)
Swap Payoff (n=44)	5.05 (2.21)	6.27 (2.46)	6.43 (3.44)
Study 5 (n=80):			
Proportional (n=40)	4.93* (2.45)	6.08 (2.43)	6.28 (3.27)
Truth-telling (n=40)	5.07 (1.98)	5.07 (1.98)	4.25 (2.86)
Study 6 (n=96) :			First / Second
Proportional, ex-post (n=64)	5.66 (2.09)	5.39 (2.46)	5.09 (3.17) / 5.73(3.15)
Truth-telling (n=32)	5.03 (2.51)	5.03 (2.51)	4.34 (3.07) / 4.22 (3.25)
Study 7 (n=116):			
Proportional (n=60)	5.33 (1.96)		7.22 (3.34)
Truth-telling (n=56)	5.79 (2.16)		6.14 (3.15)
Study 8 (n=80) :			
Proportional (n=80)	Quiz 1: 5.51 (2.17) Quiz 2: 6.81 (2.56)	6.69 (2.49)	6.64 (3.56)

Proportional (n=60) — Feedback

Actual:	Bad	Good
Bad	17	18
Good	2	23

Truth-telling (n=56) — Feedback

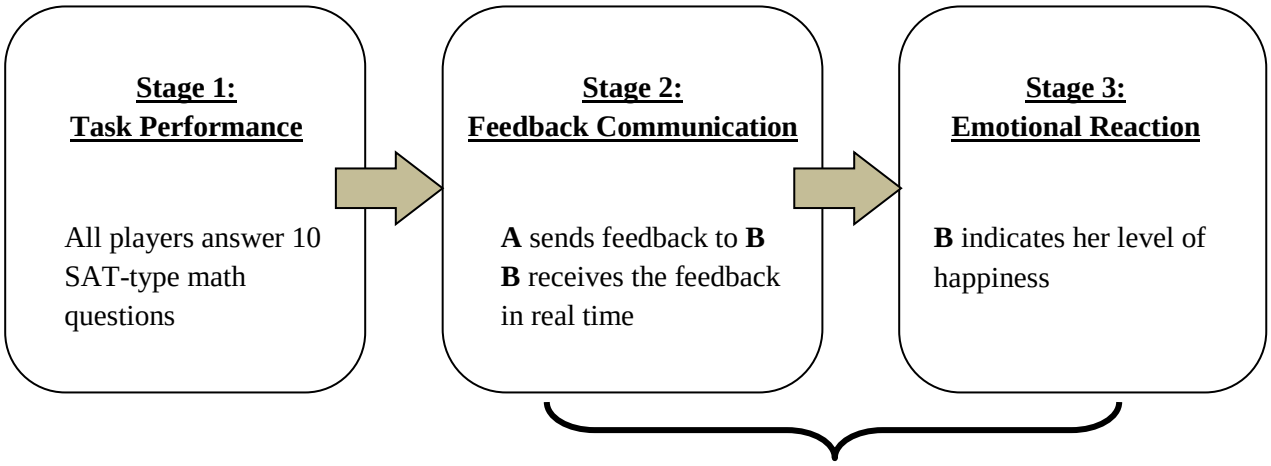
Actual:	Bad	Good
Bad	23	0
Good	0	33

*Client's prediction of her own performance was 6.23 (SD = 2.53)

Note. The number in each cell shows the mean response and the number in the parenthesis shows the standard deviation

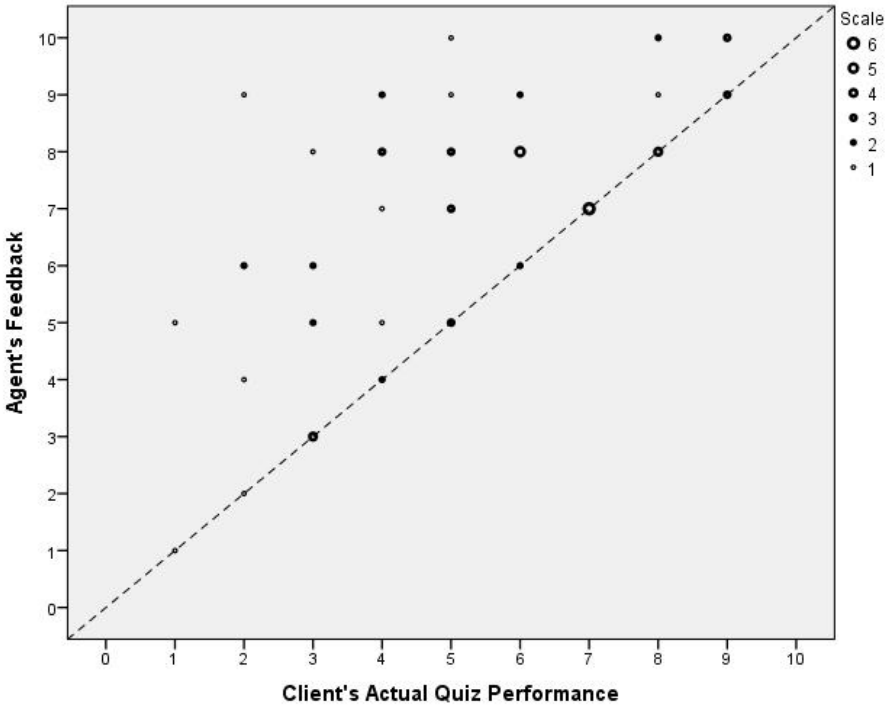
Figure 1. The 3-Stage Experimental Paradigm for the Feedback-Giving Game

This figure illustrates the way four players (A, B, C, and D) participate in the feedback-giving game. In an actual experiment, the number of players can be any multiples of four.



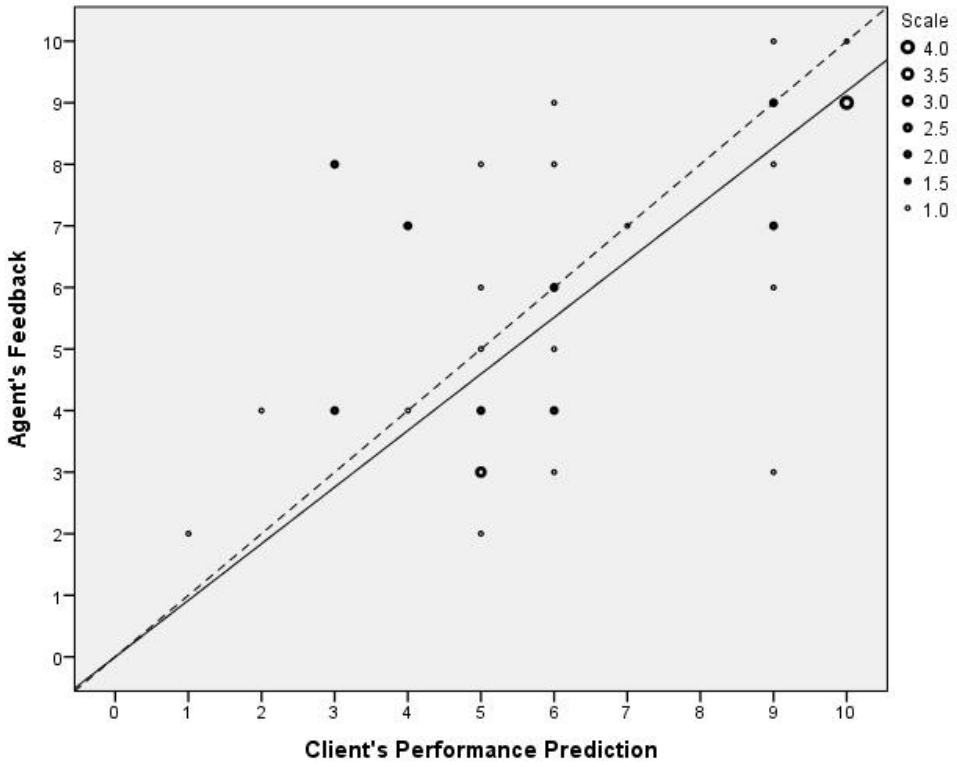
Stage 2 and Stage 3 are repeated once. In the repetition, player **C** is paired up with **A** and sends feedback to **A**, whereas player **D** is paired up with **B** and receives feedback from **B**. Each player's role assignment is randomly determined.

Figure 2: Scatter Plot of Agent's Feedback Against Client's Actual Quiz Performance in the Prop Condition in Study 1



The dotted line represents the 45^0 line: Feedback = Actual Quiz Performance

Figure 3: Scatter Plot of Agent's Feedback Against Client's Performance Forecast in the Prop Condition in Study 5



The solid line represents the best fitted line: Feedback = 0.919*Performance Prediction

The dotted line represents the 45^0 line: Feedback = Performance Prediction

Appendix: Experimental Instructions for Study 1 (Prop-Condition)

This is an experiment on decision making. The instructions are simple and if you follow them carefully, there is a chance that you could earn a considerable amount of money which will be paid to you in cash before you leave today.

The experiment will consist of 2 parts. In the first part, all subjects will undertake a math quiz. Subjects will be compensated based on their performance on the math quiz. In the second part, subjects will be paired in groups of two twice. The matching protocol ensures that each subject will be paired with a different subject in each of the two pairings (i.e., you will not be matched with the same subject twice). Subjects will communicate to their paired members about the latter's performance in the math quiz in one pairing and will receive feedback from their paired members about their own performance in the math quiz in another pairing. That is, subjects will provide feedback about others' performance in one pairing and receive feedback from others about their own performance in another pairing. The order of these two communication episodes (i.e., sending feedback and receiving feedback) is determined by a computer-simulated coin toss and depends solely on chance. Subjects will also be compensated when they provide feedback and the details of the compensation will be discussed in the following pages.

Any communications between subjects in each pair must be made through a Web-based computer program. It is important that you do not look at each other, and that you do not talk, laugh, or exclaim aloud during the experiment. You will be warned if you violate this rule the first time. If you violate this rule twice, you will be asked to leave, and you will not be paid. That is, your earnings will be $0.

Experimental Procedure

Again, the experiment consists of two tasks: 1) a Math Quiz and 2) a Social Communication Task. Your total cash payment will be the sum of the money you earn in each of these tasks.

Math Quiz: Every subject will undertake a math quiz in the first part of the experiment. You will answer 10 standard SAT Math questions (SAT stands for "Scholastic Aptitude Test", a standardized test that is often used to assess a student's readiness for college. In other words, all college students should be reasonably capable of solving these questions). Your answers will be graded by the computer program. You will receive $1 as a reward for each correct answer. In other words, you can earn up to $10 based on the Math Quiz and the money will be paid to you by cash before you leave the lab today.

Social Communication Task: As mentioned, your answers in the Math Quiz will be graded by the computer program. However, you will not find out your performance directly from the program. Your performance report will be sent to a subject who is randomly paired with you, and this subject will give you feedback on your performance. Likewise, you will receive the performance report of a randomly paired subject and you will give feedback to this paired member. The order of these two communication episodes is determined by a computer-simulated coin toss and depends solely on chance. The exact steps involved in this task and the payoff scheme are outlined next.

You give feedback to your paired member.

After the system has finished grading all the tests, you will receive a performance report from the program, indicating your paired member's performance in the math quiz. The report should look like the one below:

QUESTION	ANSWER
1	Correct
2	Incorrect
3	Incorrect
4	Correct
5	Correct
6	Correct
7	Correct
8	Correct
9	Incorrect
10	Incorrect

Your job is to give feedback to your paired member ("You have correctly answered _____ questions."). Your paired member will receive this feedback and indicate how happy he or she is on the following scale:

0	1	2	3	4	5	6	7	8	9	10
Not happy at all										Very happy

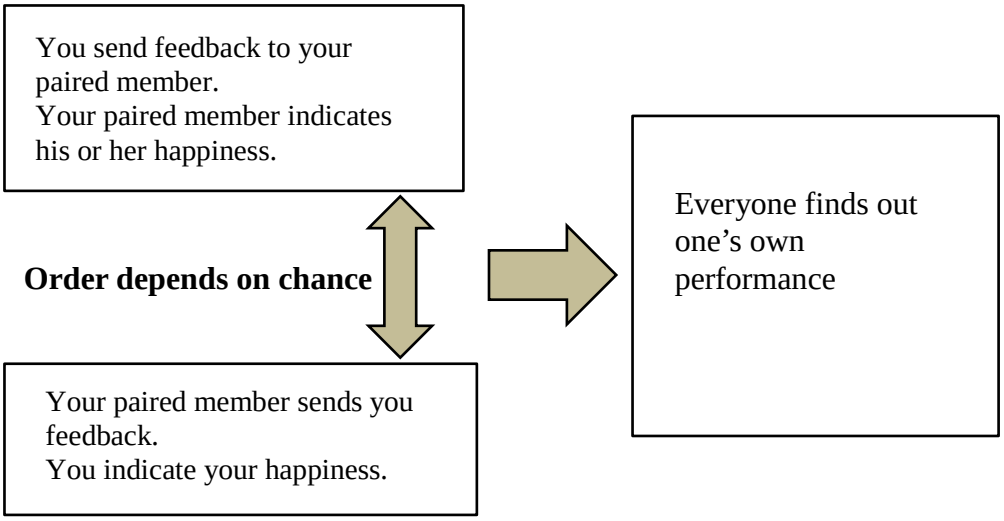
You receive feedback from a different paired member. The procedure is identical to the above except that each subject will be paired up with a different paired member and that the roles are switched (i.e., you will receive feedback from your paired member). More specifically,

- Your paired member will receive your performance from the computer program. His or Her job is to provide feedback to you ("You have correctly answered _____ questions.").

- You will then read this feedback and indicate how happy you are on the scale (0-10) mentioned earlier.

All subjects retrieve their own performance report. All subjects will be provided with their own actual performance report from the computer program.

Below is a flow chart summarizing the procedure of the second part of the experiment:



PAYOFF

As indicated above, you will be compensated for both parts of the experiment. Your payment for the Math Quiz will be based on your actual performance in the Quiz, and that you will receive $1 for each correct answer in the Quiz.

In the Social Communication Task, you will be paid by providing feedback and your payoff will depend on the paired subject's happiness right after reading your feedback. As mentioned before, your paired subject will indicate his or her happiness on a scale from 0 (not at all happy) to 10 (very happy). Each scale point will yield a payoff of $0.5. For example, if your paired member indicates a happiness rating of "10", you will receive $5 (10 x 0.5) for the Social Communication Task. If your paired member indicates "0", you will receive nothing for the Social Communication Task. Similarly, a happiness rating of 5 will give you a payoff of $2.5. In other words, the higher your paired member's reported happiness after reading your feedback, the higher will the payment be.

Your total payoff for today's experiment is the total amount of money you earn from the Math Quiz and the Social Communication Task. You will receive the payment in cash when you leave the experiment.