

**UNDERSTANDING DISHONEST BEHAVIOR:
WHAT MOTIVATES IT, HOW TO PREVENT IT, AND HOW PEOPLE
RESPOND TO IT**

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Fangtingyu Hu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Avner Ben-Ner, advisor

September 2022

© Copyright by Fangtingyu Hu 2022

All Rights Reserved

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Avner, for your endless support, guidance, and understanding during the most difficult time of my doctoral study. You are the best adviser one could ever have.

To my parents and grandparents, thank you for supporting me, encouraging me, and always be there for me. Your unconditional love is the strongest backbone accompanying and enlightening me in every step of my life.

To Lei, thank you for your unconditional faith in me. You have been my motivation and inspiration. Without your support, I would not achieve where I am.

I would like to also extend my sincere gratitude to my committee, Alan, Marc, and Betty, Michael, to my peers, Julie, Wei, Yuening, Bori, Liz, Sima, Doug, and to all work staff and social science lab staff, Pernu, Jennifer, Mindy, Shelley. It has been a wonderful time studying and working with all of you.

DEDICATION

To my beloved family.

ABSTRACT

Do people lie less in repeated interactions with the same partner than in a series of one-shot interactions with strangers? We find that under asymmetric information, senders lie substantially less if paired with the same receiver than when randomly re-matched with different receivers. However, the lying gap diminishes if the receiver is allowed to offer feedback to the sender

We investigate the effects of feedback on the decision to lie in a sender-receiver deception game with imperfect lie detection. We find evidence of feedback effects through two channels. First, the mere expectation of receiving feedback, including the anticipation of positive feedback and the threat of negative feedback, reduces lying. Second, actually-received feedback affects the subsequent decision to lie, but only in one situation: honest-type people who are being falsely punished with negative feedback become three times as likely to lie as those who are correctly rewarded with positive feedback. Our results indicate that the anticipation effect is the primary deterrent of lying, rather than the experience of receiving negative or positive feedback. Feedback may backfire and should be used with caution: honest-type individuals who are condemned as liars are surprised and react with moral indignation.

Much of the research on lying focuses on the lie-teller, whereas the lie-receiver remains understudied. We provide an overarching view of how people respond to possible lies and articulate the motivations behind those responses. We find evidence that receivers are less likely to give negative feedback to a partner with whom they will interact repeatedly in the future than to a stranger in a one-shot game. The difference is driven by strategic responses to dishonest reports in expectation of future reciprocity.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	I
ABSTRACT	III
LIST OF TABLES	V
LIST OF FIGURES	VI
CHAPTER 1: INTRODUCTION	1
1.1 INTRODUCTION.....	1
1.2 LYING IN REPEATED GAME VS. ONE-SHOT GAME.....	2
1.3 THE EFFECT OF FEEDBACK ON LYING	3
1.4 RESPONSES TO LYING	5
CHAPTER 2: METHOD	8
2.1 EXPERIMENTAL DESIGN.....	8
2.2 EXPERIMENTAL PROCEDURE.....	11
CHAPTER 3: LYING IN FINITELY REPEATED GAME VS. ONE-SHOT GAME	12
3.1 LITERATURE REVIEW AND HYPOTHESIS.....	12
3.2 RESULTS	13
3.3 CONCLUSIONS.....	16
CHAPTER 4: THE EFFECTS OF FEEDBACK ON LYING	18
4.1 LITERATURE REVIEW & HYPOTHESIS	18
4.2 RESULTS.....	21
4.2.1 <i>Overall Treatment Effect</i>	22
4.2.2 <i>Feedback Expectation and Feedback Experience</i>	23
4.2.3 <i>Robustness check: using self-report general honesty scores (DType) instead of LType</i>	28
4.3 CONCLUSIONS	30
CHAPTER 5: RESPONSES TO LYING	39
5.1 LYING SIGNALS - REPORT TYPES.....	39
5.2 RELATED LITERATURE AND HYPOTHESES.....	41
5.3 RESULTS.....	48
5.4 CONCLUSIONS & LIMITATIONS.....	56
REFERENCES	58
APPENDIX A	62
EXPERIMENT INSTRUCTIONS	62
APPENDIX B	64
APPENDIX C	66

LIST OF TABLES

Table	Page
TABLE 3-1. DESCRIPTIVE STATISTICS – MEAN (SD), BY CONDITION.	14
TABLE 3-2. DETERMINANTS OF LYING (LOGIT).....	14
TABLE 4-1. VARIABLE DEFINITION AND DESCRIPTIVE STATISTICS	32
TABLE 4-2. TREATMENT EFFECTS OF FEEDBACK ON LYING (OLS)	33
TABLE 4-3. THE EFFECT OF FEEDBACK EXPERIENCE ON LYING (OLS)	34
TABLE 4-3A. AVERAGE PREDICTIVE PROPORTION OF LYING BY FEEDBACK EXPERIENCE AND LYING TYPE	35
TABLE 4-3B. AVERAGE PREDICTIVE PROPORTION OF LYING USING SELF-REPORT HONESTY SCORES (<i>DTYPE</i>)	35
TABLE 5-5. THE EFFECT OF REPORT TYPE HISTORY ON GIVING NEGATIVE FEEDBACK, BY CONDITION (OLS).....	53
TABLE 5-6. FEEDBACK IN T GIVEN HISTORY OF MESSAGE TYPES IN T-1	55
TABLE C1. FREQUENCY COUNTS OF REPORTED NUMBERS AT EACH ROLLED NUMBER – FB TREATMENT	66
TABLE C2. FREQUENCY COUNTS OF REPORTED NUMBERS AT EACH ROLLED NUMBER - NO FB TREATMENT.....	66

LIST OF FIGURES

Figure	Page
FIGURE 3-2. DYNAMICS OF MEAN OF LYING INCIDENCE OVER TIME, BY CONDITION	15
FIGURE 4-1. MEAN PROPORTION OF LYING AT EACH ROLLED NUMBER WITH 95% CONFIDENCE INTERVAL.....	36
FIGURE 4-2. MEAN PROPORTION OF LYING OVER TIME BY CONDITION	37
FIGURE 4-3A. AVERAGE PREDICTIVE PROPORTION OF LYING BY FEEDBACK EXPERIENCE AND LYING TYPES.....	38
FIGURE 4-3B. AVERAGE PREDICTIVE PROPORTION OF LYING USING SELF-REPORT HONESTY SCORES (<i>DTYPE</i>).....	38
FIGURE 5-1. PREDICTED PROBABILITY OF GIVING NEGATIVE FEEDBACK, BY REPORT TYPE HISTORY	55
FIGURE 1A. STFB & FIGURE 1B. PTFB.....	55
FIGURE 1C. ST & FIGURE 1D. PT.....	55
EXHIBIT 1. A CHOOSES THE NUMBER TO REPORT TO B.	64
EXHIBIT 2. B RECEIVES THE REPORTED NUMBER FROM A.....	64
EXHIBIT 3. B CHOOSES WHICH FEEDBACK TO SEND TO A.	65
EXHIBIT 4. A RECEIVES B'S FEEDBACK	65

CHAPTER 1: INTRODUCTION

1.1 Introduction

Lies are common in the workplace. An employee may lie about the progress of a project to obtain a better performance evaluation from a supervisor, or may overstate his or her abilities to receive a job offer. Employers may conceal financial difficulties to prevent workers from leaving the company, or may portray a financial situation worse than it is to discourage wage demands. Lying is at the heart of the principal-agent problem. Both moral hazard and adverse selection are caused by lying, as one party fails to act in good faith or takes advantage of information asymmetry to conceal the truth, at a cost to the other party.

Among various types of lies - white, selfish, altruistic, Pareto, and punitive lies - we concentrate on selfish lies that benefit liars but harm recipients. Selfish lies could erode trust by fueling suspicion and creating uncertainty, reduce efficiency, and harm collaboration. It is therefore important to understand what motivates them, what fuels responses to them, and identify effective ways to reduce their incidence.

In Chapter 2, we introduce an experimental design that studies lying in a variant of the sender-receiver deception game. In our game, the sender's payoff increases with the reported number, whereas the recipient's payoff is probabilistic, higher when the sender tells the truth than when the sender lies; hence lies cannot be perfectly detected by receivers.

Chapter 3 and Chapter 4 concentrate on senders' lying behavior. Chapter 3 compares lying propensity under repeated games versus one-shot games, with and without feedback. Chapter 4 discusses the effect of feedback on lying.

Chapter 5 focuses on receivers' feedback provision. We explore how do they respond with feedback under different conditions, and the possible motivation behind their responses.

1.2 Lying in repeated game vs. one-shot game

In Chapter 3, we investigate under controlled lab conditions how lying behavior varies between finitely-repeated and one-shot games under asymmetric information, with and without evaluative feedback. Our motivation is to understand whether institutions that contribute to build reputation and moral image may protect against dishonesty. For example, an individual reports her work progress to others who do not observe it directly may choose to tell the truth or misrepresent her progress for personal gains depending on whether this situation occurs just once or repeatedly (e.g., one-off team or long-term assignment), and whether others may comment on her report (e.g., feedback).

By comparing the lying propensity across the four conditions, we find that repeated interaction deters lying in the absence of feedback and such deterrence significantly diminishes if feedback is present. We discuss the possible explanations, including the moral image concerns, and disruptions of reputation building.

1.3 The effect of feedback on lying

Chapter 4 takes a closer look at the effects of feedback on lying behavior. Previous research has explored extensively how different types and sizes of monetary gains and losses influence lying (Erat & Gneezy, 2012; Faravelli, Friesen, & Gangadharan, 2015; Gneezy, 2005; Gneezy, Meier, & Rey-Biel, 2011). However, only a few studies have investigated how social factors influence lying behaviors (see Gneezy, Kajakaite, & Sobel, 2018). In this study, we focus on one type of social factor – feedback. We examine the role of affirming (positive) feedback and critical-accusatory (negative) feedback after an individual makes a statement that may represent a lie or truth.

Feedback is defined as evaluative comments and responses people make about others' behavior. Respondents may provide feedback to deter lying, or just express a moral sentiment. We introduce both positive and negative feedback, where positive feedback ("Thank you for being an honest person!") expresses appreciation and recognition of honesty and negative feedback ("It is not good to be a liar!") expresses suspicion of dishonesty and hurls an accusation.

Generally, we ask whether feedback affects people's decision to lie, and if so, how. There are two principal channels through which feedback can affect individuals' decision to lie: the expectation of feedback, and the effect of received feedback. The desire to receive positive feedback and the fear of receiving negative feedback may affect the decision to lie in the first place, before any feedback is provided. The lying decision may also be influenced by actual feedback through activation of various emotions and moral sentiments. We offer evidence on both channels.

In our experiments, receivers send feedback to senders in the Feedback treatment, with the feedback channel shut down in the No Feedback treatment (control condition). This allows us to assess the group-level effect of feedback on subjects' lying propensity. Within the Feedback treatment, because lies cannot be perfectly detected, there are four possible feedback experiences: honest senders being recognized and rewarded with positive feedback, dishonest senders being punished with negative feedback, honest senders being punished with negative feedback, and dishonest senders being rewarded with positive feedback. The first two scenarios are legitimate situations where feedback is correctly given, whereas in the latter two scenarios feedback is falsely given. We explore the effect of different feedback experiences on senders' subsequent decision to lie and find that honest people may respond differently, depending on whether the feedback is legitimate or erroneous.

We find that people are less likely to lie in the Feedback treatment than in the NO Feedback treatment. Moreover, we find evidence of feedback effects through the two channels noted above. First, the mere expectation of receiving feedback appears to reduce lying: the incidence of lying in the first round is significantly lower in the Feedback treatment than in the NO Feedback treatment. Second, feedback affects the subsequent decision to lie, but only in one situation: honest-type people who are being falsely punished with negative feedback become more likely to lie than those who are correctly rewarded with positive feedback. This finding is similar to that of Houser, Vetter and Winter (2012), where individuals who believe that they were unfairly treated are subsequently more likely to cheat. However, neither positive nor negative feedback affects significantly liars' decisions, who may not care about how others think of them or

may have already adjusted their behavior to expectations of receiving disapprobation for lying. Therefore, it appears that the expectation of feedback is the primary deterrent of lying, rather than the experience of negative emotions upon receipt of negative feedback, or positive feedback that encourages subsequent truthful behavior. The paper's principal contributions consist of providing the first examination of the role of feedback in lying behavior, and the introduction of the probabilistic rewards for potential lie-receivers, which provides scope for the provision of incorrect feedback.

1.4 Responses to lying

Chapter 5 addresses the behavior of potential lie-recipients and examines how they respond to lying or truth-telling in forms of feedback.

Consider a dyadic interaction whereby an individual S (sender) can take an action that benefits him at a cost to his counterpart, R (receiver), but because there are additional factors that affect trusting's payoff in addition to S's action, R cannot infer with certainty S's action. This situation is extremely common. In the canonical agency model, the principal (R) observes her outcome, which is the result of the agent's actions and random factors she cannot observe so she can make only probabilistic attributions to the agent's action. Such interactions may occur just once between particular sender and receivers, or may be repeated.

A large and diverse literature examines various aspects of interactions that reflect such situations. Agency and game theory studies how receivers may affect senders' behavior

to better serve principals' interests through various incentive schemes and monitoring. The dominant assumption in this literature is that both senders and receivers are self-interested, rational and have no concerns about honesty. The behavioral and experimental economics literature examines broader sender motivations, including truth-telling preferences, concern for moral image and more. This literature finds that, due to such considerations, many senders do not universally exploit their informational advantage to make financial gains. A small literature investigates how receiver responses, whether evaluative "cheap talk" feedback or financial punishments, affect sender behavior. Receiver behavior has been the subject of a more diffuse literature, evaluating receiver characteristics that affect the ability to decipher sender behavior, receiver emotional reactions to sender non-cooperation (in public goods, prisoner dilemma and similar games). In the human resources management literature, emphasis is laid on feedback provided by receivers-managers to senders-employees.

Much is known about senders from experiments with sender-receiver games, but it appears that the behavior of receivers in such games has not been studied. We seek to contribute to this literature by studying experimentally the following questions. (1) Do receivers give evaluative feedback differently in one-shot and repeated interactions? (2) If so, do receivers behave strategically to improve their payoffs? 3) If so, what kind of strategy is adopted by receivers?

We find that receivers become significantly less likely to give negative feedback in repeated game (partners) than in one-shot game (strangers). We also observe the same

gap among partners between real feedback and hypothetical feedback, which indicates there exists a decrease in strategic play of feedback in hypothetical conditions. Further investigation suggests that the strategy adopted by partners in real feedback condition is to “conditionally reciprocate” those who show an honest potential in the past but are dishonest in the current round, in expectation of inducing future honest behavior.

CHAPTER 2: METHOD

2.1 Experimental Design

Our experiment builds upon the sender-receiver deception game introduced by Gneezy et al. (2013). Subjects are randomly assigned fixed roles, senders or receivers. Senders roll a digital die carried out by the computer that generates a random number s between 1 and 6. Only senders know the number s . Senders report a number r to receivers; r is constrained to be between 1 and 6 but does not have to be the actual rolled number s (see Appendix B Exhibit 1). The criteria for payoffs to senders and receivers are known in advance to both. The payoffs for senders and receivers are as follows:¹

$$\pi_{sender} = 30 \text{ ECUs} + 10 \times r$$

$$\pi_{receiver} = \begin{cases} (40 \text{ ECUs with probability } 0.75 \text{ and } 0 \text{ ECU with probability } 0.25) & \text{if } A \text{ told the truth} \\ (40 \text{ ECUs with probability } 0.25 \text{ and } 0 \text{ ECU with probability } 0.75) & \text{if } A \text{ told a lie} \end{cases}$$

where ECU=Experimental Currency Unit, 1 ECU = 0.1 USD

The receiver receives a message from the sender about the rolled number r (see Appendix B Exhibit 2) and enters the lottery. The computer reveals the lottery payoff to the receiver privately (the sender does not know the receiver's payoff). Due to the probabilistic element in our design, receivers cannot know with certainty whether senders

¹ The payoff parameters were carefully designed so that 1) senders are sufficiently motivated to lie as each unit of misreport is worth 10 ECUs. Senders' payoff from lying ranges from 10 ECUs to 50 ECUs, on top of the fixed gain of 30ECUs; and 2) sender's minimum payoff is 40 ECUs, which is the same as the maximum payoff of the receiver, to ensure that lying is not caused by sender's sense of unfairness or difference aversion.

lie or not, but can make an inference based on the outcome (a standard asymmetric information situation). This game is repeated for 18 rounds.

There are four experimental conditions. Each subject participates in a single session, a between-subjects design. In the Stranger condition (ST), senders are randomly matched with a new receiver at the start of each of the 18 rounds. In the Partner (PT) condition, the match remains fixed for all 18 rounds. In the other two conditions, we require receivers to respond to the sender's reports after observing r and their own payoff by sending a positive or negative feedback message to their senders: "Thank you for being an honest person!" or "It is not good to be a liar!" The feedback is added to the Stranger condition (ST) to create the Stranger-Feedback condition (STFB) and to the Partner condition (PT) to generate the Partner-Feedback condition (PTFB). Subjects are paid for one randomly-chosen round at the end of the experiment. Full instructions for the STFB condition are in Appendix A.

The novel aspects of our design are that 1) we create a context of imperfect lie-detection – people cannot tell with certainty whether others are lying or not, but can only have suspicions, thus allowing for erroneous feedback, and 2) we created conditions with and without feedback to examine feedback effect on lying, and conditions with repeated games or with one-shot to examine repetition effect on lying.

Stranger - Feedback Treatment (STFB)

Participants are randomly re-matched to play the role of sender and receiver in each round of the game, for a total of 18 rounds. After receivers' payoffs are revealed,

they are mandated to give costless feedback to senders. Feedback happens at the end of each round (see Appendix B Exhibit 4).

Stranger- No Feedback Treatment (ST)

This entails the same procedure as STFB but eliminates the feedback option.² Senders and receivers proceed directly to the next round after seeing their own payment in each round. The receiver is asked to answer hypothetical questions, of which the sender is not aware.

Partner – Feedback Treatment (PTFB)

This entails the same procedure as STFB except that participants are matched as sender and receivers in the 1st round and remain in the same pairs for all 18 rounds of the game.

Partner – No Feedback Treatment (PT)

This entails the same procedure as PTFB but eliminates the feedback option. Receivers will answer hypothetical questions about feedback, as in STFB.

² The receiver is asked to answer some hypothetical questions after receiving the reported number from the sender. We ask whether the receiver thinks the sender lied or not and what message the receiver would have sent to the sender if given the chance to do so. The sender is not aware of these questions. Since the receiver's behavior is not analyzed in this paper, we do not present the receiver's answers here.

2.2 Experimental Procedure

We recruited 218 pairs of subjects at the University of Minnesota through posters, announcements, and Carlson School Paid Subject Pool. There were 60 pairs of subjects in the STFB, 46 pairs of subjects in the ST, 54 pairs in PTFB, and 58 pairs in PT. We ran 25 sessions with 18 to 26 subjects per session. Subjects received a show-up fee of \$10, plus earnings from the online registration stage and the lab experiment. The average earning (including the show-up fee) was \$21.

A 15-minute online registration survey was completed at least 24 hours before participation in the 30-minute lab session at the Social and Behavioral Lab at the University of Minnesota. The survey asks about demographic information, a personality questionnaire, and opinions about various issues. The lab session was operated through oTree (Chen, Schonger, & Wickens, 2016), an open-source platform.

Upon arrival at the lab, subjects sat at standard private computer stations and were instructed to read through a printed copy of the instructions. An electronic copy of instructions was also provided. Subjects' understanding of the instructions was tested before proceeding to the task. Each subject was randomly assigned to play either as a sender or a receiver for 18 rounds. At the beginning of each round, subjects were reminded that they were paired with a different counterpart. After the completion of the 18-rounds sender-receiver game, subjects saw a summary of their earnings for each round and one round was randomly chosen to be paid. Subjects completed a short post-lab survey and were paid in cash and then dismissed.

CHAPTER 3: LYING IN FINITELY REPEATED GAME VS. ONE-SHOT GAME

3.1 Literature Review and Hypothesis

In finitely-repeated interactions, backward induction results in non-cooperation just like in single shot interactions. However, greater cooperation may arise in finitely-repeated interactions from: building a reputation that can be exploited later (Andreoni, 1988), the warm-glow of being nice towards others (Andreoni, 1990), conditional cooperation (Fischbacher et al. 2001) or a punishment strategy under asymmetric information (Kreps et al., 1982). Reduced cooperation may result from confusion-induced errors (Andreoni, 1995; Palfrey & Prisbrey, 1997). The literature revealed mixed findings on free-riding in finitely-repeated interactions, i.e., some found free-riding decreased (Andreoni & Croson, 2008; Croson, 1996; Sonnemans et al., 1999), some found it increased (Andreoni, 1988; Burlando & Hey, 1997), and some are inconclusive (Brandts & Schram, 2001). The experimental literature on evaluative feedback in public goods games found that it promotes cooperation (e.g., Masclet et al. 2003) - on average, people act in ways that earn them approval and avoid disapproval.

Hypothesis: People lie less in finitely-repeated interactions than in one-shot interactions.

3.2 Results

The experiment was conducted at the University of Minnesota with 218 sender-receiver pairs, with an average of 16 participants per session. Subjects are students (93.3%) and staff. Subjects earned on average \$21, including a \$10 show-up fee. The present paper focuses on senders' lying behavior. Table 1 presents descriptive statistics of senders' individual characteristics, by condition; chi² tests indicate that subjects are comparable across conditions.

The lying incidence in Table 3-1 is the percentage of times a sender lied over 18 rounds, averaged over all senders. The highest incidence of lying is in ST (0.59), the lowest in PTFB (0.41), with PT and STFB in between (0.46 and 0.47). This suggests that, absent feedback, turning one-shot into repeated interactions reduces lying significantly (ST vs PT: $p=0.05$, two-sided t-test); while with feedback, repeated interaction does not significantly reduce lying (STFB vs. PTFB: $p=0.35$). Adding feedback to ST reduces lying marginally (ST vs. STFB: $p=0.09$). There is no significant difference among the condition with repetition only, with feedback only, and with both (PT vs. STFB: $p=0.82$, PT vs. PTFB: $p=0.49$). These results are further confirmed in Table 2, presents the odds ratios of lying in repeated versus one-shot interactions, with and without feedback, from a logit regression controlling for rolled numbers, round and individual characteristics. Compared to ST, the probability of lying is significantly lower in PT (column 1). Whereas in the presence of feedback, the probability of lying is lower in PTFB versus STFB but is not statistically significant (column 2).

Consider next the dynamics of lying over time in Figure 3-1. We observe that lying remains constantly high in ST, while the lying incidence increases in STFB, in

PTFB and (weakly) in PT, which implies that subjects may engage in early reputation building and later advantage-taking of reputation. The difference in lying between ST and PT is prominent but diminishes sharply during the last few rounds, indicating a persistently strong reputation effect except at the end. When we compare STFB and PTFB, we find that the additional reputation effect appears smaller and does not sustain over time. Those results echo findings from Table 3-1 & 3-2. We discuss implications in next section.

Table 3-1. Descriptive statistics – Mean (SD), by condition.

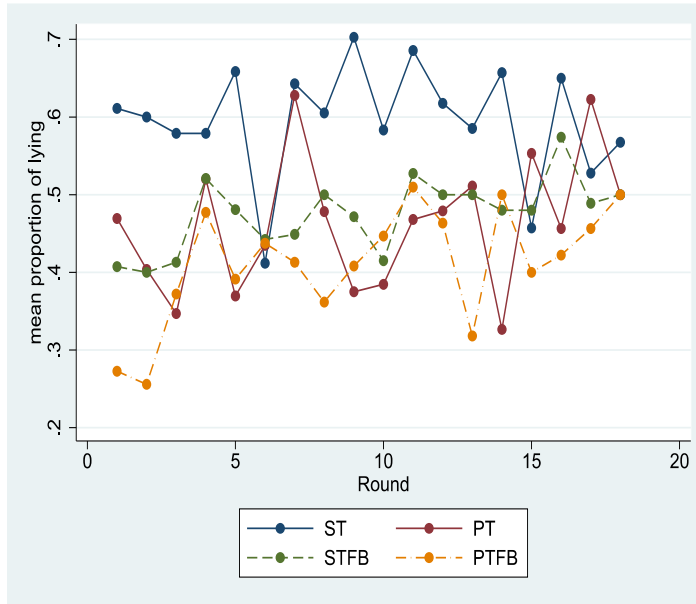
	<i>Number of senders</i>	<i>Proportion female</i>	<i>Age</i>	<i>Altruism</i>	<i>Agreeableness</i>	<i>Lying incidence</i>
Stranger no feedback ST	46	0.57	21.8 (4.68)	4.00 (3.43)	4.00 (0.83)	0.59 (0.33)
Partner no feedback PT	58	0.60	22.0 (6.63)	4.18 (3.54)	3.97 (0.78)	0.46 (0.37)
Stranger + feedback STFB	60	0.53	21.1 (3.54)	3.93 (3.47)	3.96 (0.76)	0.47 (0.37)
Partner + feedback PTFB	54	0.63	21.4 (4.71)	4.02 (3.44)	4.03 (0.85)	0.41 (0.31)
Notes. <i>Lying incidence</i> is the mean of the average individual proportions of lies over 18 rounds. <i>Altruism</i> is the amount donated to one of seven charitable organizations of subjects' choice out of an endowment of 10 ECUs. <i>Agreeableness</i> is measured on a 1-5 scale of the mini-IPIP.						

Table 3-2. Determinants of lying (logit)

<i>Dependent Variable: Lie = 1 if lied, = 0 if told the truth</i>		
	<i>No Feedback PT vs. ST</i>	<i>With Feedback PTFB vs. STFB</i>
	(1)	(2)
Repeated Interaction (Partner) vs. One-Shot (Stranger)	0.537**	0.683
	(0.167)	(0.194)
N	1872	2052
R ² (pseudo)	0.243	0.195
Notes. Odds ratio estimates. Standard errors clustered at the individual level in parentheses.		

All analyses control for rolled number fixed effects, round fixed effects, gender, altruism, and agreeableness.
The omitted condition dummy is ST in column (1) and STFB in column (2).
* p<0.1 **p<0.05 ***p<0.01.

Figure 3-2. Dynamics of mean of lying incidence over time, by condition



3.3 Conclusions

Why does repeated interaction deter lying in the absence of feedback? A promising explanation is reputation. Those who care about their financial returns but are also concerned with their moral image in the eyes of others - their reputation – will refrain from lying in the early stage of repeated interactions to be able to earn more by lying later without losing their moral standing in the eyes of their receiver, exploiting the reputation acquired earlier. A complementary explanation is that people are more altruistic towards partners than strangers, so senders will curb lying to limit financial harm to the same receiver.

Why does the repetition effect diminish greatly when feedback is given? We surmise that repeated interaction and feedback may influence people's lying behavior through the same underlying mechanism – moral image concerns. In repeated interactions with a partner, the moral image may be damaged by consistently reporting a high number, as this goes against the simple odds. In one-shot interactions, senders may hide behind the veil of the odds in each interaction with a stranger. Feedback is an expression of how a receiver perceives the sender's moral image. In interactions with strangers, feedback forces senders to balance financial returns and moral image considerations by telling the truth or lying in a similar fashion to what reputation does in repeated interactions. Thus, the application of either institution exhausts most of the moral image effect. Future research should explore more the concept of moral image and mechanisms that activate it. Another possibility is that the presence of feedback disrupts the reputation-building process in repeated interactions; for instance, a positive evaluative feedback may give people a false sense of security that a good reputation (or moral

image) is already installed for depletion, or a negative feedback may deprive prematurely one's attempt to build good reputation. Future research should explore such disruption process.

The effects of reputation and feedback that we found in anonymous interactions in the lab are likely to be stronger in real-life face-to-face interactions. Our study has implications for the design of institutions aimed at reducing dishonest behavior in economic and social situations. The creation of stable ongoing relationships where concerns for moral image can be expressed is crucial; this could be implemented through long-term and stable affiliation with individuals, teams, or social organizations. In addition, provision of responsible feedback, even in short-term interactions, could be another important institution that protects society from dishonest behavior. However, combined use of different institutions may not bring substantial extra benefits. Finally, as these institutions rely on the concern for moral image, it is important that, where honesty is crucial to performance, selection of members and especially of leaders exclude those devoid of moral image concerns.

CHAPTER 4: THE EFFECTS OF FEEDBACK ON LYING

4.1 Literature Review & Hypothesis

Dishonest behavior happens in a social context, when individuals interact with each other (Jacobsen, Fosgaard, & Pascual-Ezama, 2018). A body of literature explores social factors – social identity, social distance and social consequences - that may affect people’s moral decisions (Ben-Ner & Putterman, 2009; Conrads, Irlenbusch, Rilke, & Walkowitz, 2013; Gino & Galinsky, 2012; Gino, Gu, & Zhong, 2009; Gross, Leib, Offerman, & Shalvi, 2018; Hermann & Ostermaier, 2018; Lundquist, Ellingsen, Gribbe, & Johannesson, 2009).³ For instance, Gneezy, Kajakaite, and Sobel (2018) in a die-rolling experiment find that the decision to lie is determined by both a self-assessment of social identity and the perception of others. Participants reported more partial lies than full lies when experimenters could observe the rolled number, which implies that people care about others’ perceptions. A prominent social mechanism that is known to influence a range of social behaviors consists of social sanctions and rewards, such as the provision of feedback. Feedback in the form of an evaluative message, a gesture or even a look on one’s face, represents people’s perception of an individual’s identity, which may affect the behavior of those who care about how others perceive them. Examining the role of

³ There is also literature in psychology that investigated the influence of moral cues on dishonest behaviors (Jacobsen, Fosgaard, & Pascual-Ezama, 2018). For example, having participants recall the 10 commandments (Mazar, Amir & Ariely, 2008), pledging to an honor code, or displaying a moral reminder such as “please do not be a cheater” (Bryan, Adams, & Monin, 2013) can effectively reduce dishonesty. In addition to explicit moral cues, some implicit moral cues are found to affect dishonest behaviors, such as the presence of a mirror (Vincent, Kyle, & Goncalo, 2013), placement of signatures (Shu et al., 2012), and a reminder of former transgressions (Barkan et al., 2012).

feedback would fill in a missing essential piece in understanding lying behavior in social contexts.

Research has demonstrated that social sanctions and rewards are effective in promoting cooperation in public goods games (Dugar, 2013; Masclet, Noussair, Tucker, & Villeval, 2003; Peeters & Vorszatz, 2013), reducing free-riding in teamwork (Ben-Ner, Putterman, & Wang, 2018), promoting fair economic exchange (Xiao & Houser, 2009), and increasing donations (Ellingsen & Johannesson, 2008). The reason for these findings is presumably that people refrain from unethical behavior in order to gain rewards and praise and avoid punishment and sanctions. Monetary sanctions appear to be effective in limiting lying (Khalmetski, Rockenbach, & Werner, 2017; Sánchez-Pagés & Vorszatz, 2009). However, to our knowledge, there is no evidence on how non-monetary sanctions and rewards affect lying. In our study, we focus on non-costly feedback that does not carry monetary consequences, but only acts as an indication of others' perception of one's honesty. The literature reviewed above suggests that that feedback will reduce overall lying.

Hypothesis 1: Lying decreases when feedback becomes available.

The information context in which lying behavior is studied varies. Some studies ensure complete privacy where no one will be able to detect lying.⁴ Other studies ensure

⁴ For example, Fischbacher and Föllmi-Heusi (2013) develop a die-rolling game where participants privately roll their die and self-report the number to the experimenter, which ensures that lying cannot be detected by anyone at the individual level. Likewise, Gneezy et al. (2018) let participants draw a card privately and report the card number to the experimenter.

complete information where lying is revealed by the outcome.⁵ An important contribution of our study is that we study lying under imperfect detection. This is a much more realistic representation of situations in which people lie, and replicates the informational assumption of agency theory. With complete and accurate information, there is no room for errors: dishonest people expect and are most likely to receive negative feedback and honest people expect and are most likely to receive positive feedback (assuming that people follow social norms that reward honesty and condemn dishonesty, and feedback is obligatory). However, when people cannot detect perfectly whether someone is lying but suspect that this may be the case, then they may falsely accuse honest behavior or falsely reward dishonest behavior.

Erroneous feedback violates social norms, and may induce various emotional responses. Suppose that, in contrast with the common norm of rewarding honest behavior with praise, one experiences condemnation. Houser, Vetter, and Winter (2012) demonstrate that the experience of violation of the fairness norm makes one more likely to violate another norm, the no-cheating norm.⁶ Erroneous feedback may also trigger various emotions (Naqvi, Shiv, & Bechara, 2006). Falsely punishing honesty may induce negative emotions, such as anger, outrage, or irritation associated with being unfairly treated (Xiao and Houser, 2005).

⁵ For example, in the sender-receiver game by Gneezy, Rockenback & Serra-Garcia (2013), senders are exposed to full monitoring pressure from receivers who will know with certainty whether senders lie or not by observing the final payoff.

⁶ This may be the result of volitional belief distortion to facilitate profitable lying (Bicchieri, Dimant, and Sonderreger, 2019).

Hypothesis 2: Falsely punishing honest behavior rather than rewarding it may encourage lying.

Falsely rewarding dishonesty may lead to diverse feelings: some may feel guilt and shame for undeserved rewards (Clay-Warner et al., 2016), while others may feel relief and joy for not being caught and punished (Berridge & Kringelbach, 2008). Therefore, the emotional effect on subsequent behavior cannot be determined due to possible heterogeneity in responses, and we therefore offer two alternative hypotheses.

Hypothesis 3a: Falsely rewarding dishonest behavior rather than punishing it may encourage lying.

Hypothesis 3b: Falsely rewarding dishonest behavior rather than punishing it may discourage lying.

4.2 Results

We first present summary statistics followed by an examination of the treatment effect of feedback on senders' propensity to lie, and distinguish between the expectation of feedback and the experience of feedback. We continue with a detailed examination of the distinct effects of different feedback experiences on senders' subsequent lying propensity. Finally, we present a robustness analysis.

Table 4-1 presents descriptive statistics for the two treatments. Individual characteristics, such as gender, age, and self-report honesty,⁷ are comparable across the two subsamples (Wilcoxon-Mann-Whitney test). Subjects on average lie 41% of the time in FB and 49% of the time in NO FB ($p = 0.08$, two-tailed t-test). In the first round, the proportion of lying is 37% in FB and 48% in NO FB ($p = 0.12$). In the final round, the proportion of lying is 43% in FB and 46% in NO FB ($p = 0.81$). Appendix C Tables c1 and c2 provide information about what numbers subjects reported for each rolled number in the two treatments.

4.2.1 Overall Treatment Effect

Figure 4-1 presents the mean proportion of lying and the 95% confidence interval at each rolled number for the two treatments. There are fewer lies in FB than in NO FB when the rolled numbers are 1, 2, 3 and 4. For instance, when the rolled number is 1, people lie 78% of the time in NO FB but only 62% of the time in FB. The differences in the mean proportion of lying range from 11% (when 4 is rolled) to 17% (when 1 is rolled), and are statistically significant ($p=0.001$ at 1, $p=0.036$ at 2, $p=0.023$ at 3, $p=0.006$ at 4, two-tailed test). For rolled numbers 5 and 6, the differences are not statistically significant ($p>0.1$). Lying at rolled number 6 is rare, most likely is a result of input errors. The lack of difference in the incidence of lying at rolled number 5 between NO FB and

⁷ In the registration survey, we ask “How would you describe yourself: “1 = very dishonest, 2 = dishonest, 3 = moderate, 4 = honest, 5 = very honest”. We also ask about subjects’ honesty in some specific contexts, the results are similar but not presented here since the general item has consistently higher correlations with lying behavior in the experiment than other items.

FB, which is about 26% in both, maybe because those who lie at 5 to make a small gain (10 ECU, as they report 6 instead of 5) are “hardcore” liars, who are not sensitive to criticism and are not deterred by the possibility of negative feedback. Feedback may be most effective at reducing the incidence of lies for people who are relatively honest by nature, and who balance the pursuit of monetary self-interest and the desire of being honest. The number of those who succumb to self-interest increases with the size of the reward, but is smaller in FB than in NO FB.

Columns (1) – (3) of Table 4-2 present the overall treatment effect on lying, controlling for the rolled number and round fixed effects, in an OLS regression (similar results obtain from a logistic regression). Results indicate that feedback reduces overall lying by 9.5% (marginally significant at 10%). The coefficient estimates of rolled-number dummies are highly significant and conform with what we have seen in Figure 4-1, where lying decreases as the rolled number increases.

In sum, we find support for hypothesis 1, that lying is lower in the Feedback treatment than in the NO Feedback treatment. More lying is observed at small numbers than at large ones.

4.2.2 Feedback Expectation and Feedback Experience

How does feedback influence lying decisions? As noted earlier, we consider two potential channels. The first channel is the mere expectation of receiving feedback, which may act as a threat of condemnation or promise of praise, which changes people’s behaviors. Individuals may refrain from dishonest behavior to avoid negative feedback

or to earn positive feedback. The second channel captures the effect of *received* feedback on behavior. In particular, in a context of uncertainty, erroneous feedback could overturn people's reward-punishment expectations and prompt emotional reactions, which influence their subsequent behavior. Next, we explore in depth the two channels.

Effects of Feedback Expectation

To show the influence of the expectation of feedback, we examine lying behavior only in the first round, when subjects in FB expect feedback but have not yet received any, while subjects in NO FB have no feedback expectations. According to column (4) in Table 4-2, lying decreases by 18.1% under feedback expectation ($p = 0.039$).

The fact that the overall treatment effect is weak while the expectation effect of feedback in the first round is much stronger presents the possibility that the expectation of feedback reduces lying, whereas some feedback experience may induce lying. Before turning to a detailed exploration of the effects of received feedback, consider Figure 4-2, which plots the mean proportion of lying in each round, by treatment. The gap between the FB and NO FB treatments is largest in the first round, then declines gradually over the remaining rounds. Lying increases significantly over time only in FB – the slope of the linear fitted line for FB is positive and statistically significant ($p=0.044$) while the slope for NO FB does not differ from zero ($p=0.716$). Moreover, column (5) of Table 4-2 shows that lying is similar across treatments in the final round. The dynamic pattern supports our conjecture that the expectation of feedback is powerful in deterring lying while some feedback experience may provoke lying.

The next section explores the role of different feedback experiences (in the FB treatment) and shows how erroneous feedback may induce lying.

Effects of Feedback Experience

How does received feedback influence people's decision to lie? Without feedback, the decision to lie depends only on the rolled number.⁸ With feedback, the decision to lie depends additionally on the feedback experience. We explore this matter by analyzing data from the FB treatment. The most recent feedback, at $t-1$, is most influential on behavior at t ; we obtain small and non-significant effects of feedback experience at $t-2$, hence we omit it from the analysis reported below.

We define four kinds of *feedback experience*, based on lying behavior and the type of feedback received: having been honest and received positive feedback (correctly rewarded - CR_{t-1}), having been honest and received negative feedback (falsely punished - FP_{t-1}), having lied and received positive feedback (falsely rewarded - FR_{t-1}), and having lied and received negative feedback (correctly punished - CP_{t-1}). The feedback experience is thus represented by four dummy variables.⁹

Next, we introduce *LType*, a variable that reflects one's lying type (lying tendency) when he or she rolled the number N for the first time in the experiment; *LType* is a dummy variable that equals 1 if the subject lied and 0 if the subject was honest. As a result, each subject has a "lying type" at each number, honest or dishonest.¹⁰ People may

⁸ An important feature of the experimental design is that subjects are only paid for one randomly-chosen round, which allows them to treat each round independently.

⁹ The feedback experience can also be represented as an interaction term between a lie/honest dummy and a negative/positive feedback dummy, with the same results but a less intuitive interpretation of estimates.

¹⁰ For example, if a person lied at 3 when he or she first encountered it, then this person is a dishonest type at number 3; the same person may be an honest type at another number.

react differently to feedback experience, depending on their lying type at the number at which they had one of the four feedback experiences. This is captured by an interaction term between *LType* and feedback experience. The effect of experienced feedback on lying, conditional on a person's lying type, can be expressed as:

$$Lie_t = \beta_0 + \beta_1 FP_{t-1} + \beta_2 CP_{t-1} + \beta_3 FR_{t-1} + \beta_4 LType + \beta_5 X \\ + \gamma_1 FP_{t-1} * LType + \gamma_2 CP_{t-1} * LType + \gamma_3 FR_{t-1} * LType \quad (1)$$

Table 4-3 presents OLS results with “lies/tells the truth at t” as the dependent variable (standard errors clustered at the individual level for 18 decisions); similar results are obtained from a logistic regression. The first two models differ only in interaction terms, whereas the third is a robustness check, discussed later. The first column presents the effect of the feedback experience in t-1 compared to the omitted category – being correctly rewarded (CR_{T-1}), controlling for the effect of the lying type and the rolled number. Unsurprisingly, rolling a higher number induces a higher probability of lying, and *LType* is large, positive and statistically highly significant, indicating that dishonest types are much likely to lie than honest types.

Regarding the effect of the feedback experience, the estimate on falsely punished (0.102) is positive and significant, suggesting that those who were falsely punished become more likely to lie than those who were correctly rewarded, given they both were honest at t-1. The estimates on correctly punished (0.202) and falsely rewarded (0.175) are also positive and significant, but are likely to be spurious and endogenous: in both

cases, the subject lied in t-1 and is likely to lie in t, more so than the subject who told the truth in t and was correctly rewarded – the omitted category. Nevertheless, we can derive a robust comparison between the estimates of correctly punished and falsely rewarded, in which both lied at t-1 but received different feedback. The two estimates (0.202 and 0.175, respectively) are statistically indistinguishable.

Column (2) adds interaction terms to explore possible divergent effects of lying experience for different lying types, as shown in equation (1). Using the estimates from column (2), we derive the average predictive proportion of lying (predictive margin) for each feedback experience by lying type – honest and dishonest. These are shown in Table 3a, and visualized in Figure 3a, which shows the 95% confidence interval. The average predictive proportion of lying differs across feedback experiences among honest-type subjects. Compared to correctly rewarding honesty, falsely punishing honest behavior increases lying by 9.7% ($p=0.038$): the average proportion of lying at t would have been 4.3% if all honest-type subjects were correctly rewarded at t-1, whereas the average proportion of lying at t would have been 14.0% if all honest-type subjects were falsely punished at t-1. This result suggests an *indignation effect*: honest-type people who are being falsely punished with negative feedback become three times as likely to lie as those who are correctly rewarded with positive feedback. This supports Hypothesis 2. In contrast, for the dishonest type, the average proportion of lying is consistently high (76.1% to 87.8%), the difference between being correctly rewarded (76.1%) and being falsely punished (87.0%) is not statistically significant ($p>0.1$; see also the overlapping confidence intervals in Figure 4-3a). We fail to find a significant indignation effect among the dishonest type.

For the honest type, compared to those who were correctly punished (32.5%), there is a slight decrease in lying among those who were falsely rewarded (25.4%), but the difference is statistically insignificant ($p > 0.1$, see the overlapping confidence intervals in Figure 3a). Likewise, for the dishonest type, the difference between being correctly punished (87.8%) and being falsely rewarded (87.7%) is statistically insignificant ($p > 0.1$). We interpret these results as evidence of heterogeneity in emotional responses as discussed in Hypotheses 3a and 3b, with perhaps a larger proportion among subjects of honest-type people who feel guilt and shame upon over-rewards. However, this effect is insignificant in our sample.

Why does the feedback experience not affect the behavior of the dishonest type? We propose two possible explanations. First, people who have determined to lie will not be affected by negative feedback because they have already expected one upon lying. Therefore, we do not observe a lying-deterring effect when lying has been correctly punished. Second, it is possible that those who have lied are less sensitive to feedback in general. The fact that the feedback effect vanishes at number 5 in Figure 4-1 is evidence that the strong liars who lie for small gains (who represent about a quarter of our sample) do not care about feedback.

4.2.3 Robustness check: using self-report general honesty scores (DType) instead of LType

One major threat to the internal validity of our results concerns the measure of lying types. The lying decision at the initial occurrence of a number is endogenous:

unless the initial number occurred in the first round, the lying behavior is subject to the influence from the feedback experience in the previous round. To deal with this issue, we use the subjects' self-report of honesty ratings provided during the registration stage as an alternative measure of lying type. We construct a dummy variable *DType* by dividing individuals into two types – honest and dishonest. The distribution of honesty scores is skewed to the right; hence we classify those with honesty scores larger of 4 (“Honest”) and 5 (“Very Honest”) as the honest type (N=48), and those with scores less than 4 (“Moderate,” “Dishonest,” and “Very Dishonest”) as the dishonest type (N=12).

The model in column (3) of Table 4-3 is identical to the one in column (2), replacing the variable *LType* with *DType*. Table 3b produces the average predictive proportion of lying under the four feedback experiences, and Figure 4-3b provides visualization and 95% confidence intervals. Unlike *LType*, *DType* is a constant for each person, hence it cannot capture the more nuanced differences among people who vary in their willingness to lie at different numbers. Results in Table 4-3b demonstrate that the indignation effect is manifested among the Honest type: honest-type subjects who are falsely punished are more likely to lie than those who are correctly rewarded - the proportion of lying increases by 9.1% ($p = 0.061$). As with the previous findings, feedback experience does not influence the propensity to lie among dishonest-type subjects.

4.3 Conclusions

In *The Theory of Moral Sentiments* (1759), Adam Smith argued that moral behavior is greatly affected by people's desire to gain praise and approbation and to avoid contempt and condemnation from others. In this paper, we explore how lying and truth-telling are influenced by approbation and condemnation. We conduct lab experiments with 18 rounds of a sender-receiver game whereby the sender rolls a six-sided die and reports a number to a receiver. The sender's payoff increases in the reported number, which provides a clear financial incentive to lie. The receiver's expected payoff depends on the truthfulness of the sender's report and a random factor. We explore senders' choice of lying relative to positive and negative feedback sent by receivers who do not know with certainty whether the sender's report is truthful or not, but must infer that probabilistically on the basis of the reported number and their payoff. We compare the incidence of lying in a treatment with obligatory negative or positive feedback with the incidence of lying in a no-feedback treatment, and evaluate the overall effect of the availability and the use of feedback on lying. In addition, exploiting the fact that some inferences made by receivers are erroneous, we thus explore the effect of different feedback experiences: correct condemnation of lying and praise of truthfulness, as well as erroneous condemnation of truthfulness and praise of lying.

Overall, lying is lower in the treatment with feedback. But it turns out that, in our experiment and sample, correctly given punishment has no impact on subsequent lying, whereas incorrectly given condemnation elicits lying (in comparison to correctly given praise). How can we reconcile the two sets of findings, the overall lower incidence of lying in the treatment with feedback as compared to the treatment without feedback, and

the sole, negative effect of feedback when it causes what we call moral indignation for being falsely accused of lying? We suggest that dreading condemnation and seeking approbation as suggested by Adam Smith and a large literature in the social and behavioral sciences has a disciplining effect in and of itself. This expectation effect deters some lying, and it is separate and distinct from the effect of received feedback. We confirm the existence of this difference when we find that in the first round of the feedback treatment lying is substantially scarcer than lying in the same round in the no-feedback treatment, despite the fact that there was no previously received feedback. It is indeed the threat or expectation of feedback that reduces lying and enhances truth-telling. Realized (actual) feedback comes as a surprise particularly to individuals who have been truthful but receive condemnation, and they react to it with indignation. Our results imply that feedback may be effective in reducing dishonesty but should be used with caution in the context of imperfect information, because mistakenly punishing honest behavior may induce more lying.

Our study is subject to certain limitations. First, with mandatory feedback in the experiment, we are only able to compare the effect of correct feedback with the effect of erroneous feedback, but cannot compare the effect of receiving (in)correct feedback with the effect of not getting any feedback. Second, we only allow subjects to send two types of standardized feedback messages, whereas in real life feedback takes various forms. Lastly, emotions seem to play important roles in people's decision to lie; our study was not designed to capture emotional responses. Future research could extend our work to allow free communication, measure emotional responses, and seek to explore other non-monetary factors that drive lying aversion.

Table 4-1. Variable Definition and Descriptive Statistics

		Feedback Treatment				No Feedback Treatment			
	Definition	Mean	SD	Min	Max	Mean	SD	Min	Max
Female	Dummy variable, 1 if female, 0 if male.	0.53	-	0	1	0.57	-	0	1
Age	Age of respondents.	21.1	3.54	18	38	21.8	4.68	17	38
Honesty	How do you describe yourself: 1=very dishonest, 2=dishonest, 3=moderate, 4=honest, 5=very honest.	4.03	0.71	2	5	4.04	0.76	2	5
Proportion of lying	Proportion of lies in all rounds.	0.41	-	0	1	0.49	-	0	1
Proportion of lying first round only	Proportion of lies in the first round.	0.37	-	0	1	0.48	-	0	1
Proportion of lying final round only	Proportion of lies in the final round.	0.43	-	0	1	0.46	-	0	1
Subjects	Number of subjects play as senders	60				46			
Observations	Number of subjects times 18 rounds	1080				828			

Table 4-2. Treatment Effects of Feedback on Lying (OLS)

Dependent variable: Lie = 1 if lied, 0 if told the truth

	All 18 rounds			First round only	Last round only
	(1)	(2)	(3)	(4)	(5)
FB condition	-0.0955* (0.0564)	-0.0954* (0.0567)	-0.0956* (0.0569)	-0.181** (0.0865)	-0.0638 (0.0893)
Rolled Number=1	0.414*** (0.0521)	0.412*** (0.0522)	0.421*** (0.0454)	0.449*** (0.138)	0.169 (0.137)
Rolled Number=2	0.428*** (0.0483)	0.424*** (0.0484)	0.401*** (0.0451)	0.419*** (0.133)	0.247* (0.141)
Rolled Number=3	0.317*** (0.0496)	0.314*** (0.0494)	0.315*** (0.0448)	0.0138 (0.145)	0.601*** (0.154)
Rolled Number=4	0.121*** (0.0439)	0.118*** (0.0439)	0.111*** (0.0390)	0.0935 (0.135)	0.207 (0.138)
Rolled Number=6	-0.247*** (0.0393)	-0.245*** (0.0390)	-0.227*** (0.0380)	-0.346** (0.138)	-0.333** (0.139)
Constant	0.326*** (0.0501)	0.312*** (0.0561)	0.313*** (0.0544)	0.414*** (0.0969)	0.363*** (0.107)
<i>Round Fixed Effects</i>	<i>No</i>	<i>Yes</i>	<i>Yes</i>	.	.
<i>Individual Random Effects</i>	<i>No</i>	<i>No</i>	<i>Yes</i>	.	.
N	1908	1908	1908	106	106
R ²	0.248	0.251	0.251	0.292	0.269

Notes: Standard errors in parentheses. Standard errors in columns (1) – (3) are corrected for clustering at the individual level. The results are robust to clustering of standard errors also at the session level.

FB treatment is a dummy equaling 1 if a subject was assigned to the treatment with feedback.

Rolled Number = 5 is the omitted dummy variable.

* p<0.1 ** p<0.05 *** p<0.01

Table 4-3. The Effect of Feedback Experience on Lying (OLS)

Dependent variable: Lie = 1 if lied, =0 if told the truth

	<i>LType</i> (1)	<i>LType</i> (2)	<i>DType</i> (3)
Falsely punished $t-1$ (FP_{t-1})	0.102*** (0.0372)	0.0966** (0.0454)	0.0916* (0.0502)
Correctly punished $t-1$ (CP_{t-1})	0.202*** (0.0527)	0.281*** (0.0826)	0.435*** (0.0616)
Falsely rewarded $t-1$ (FR_{t-1})	0.175*** (0.0465)	0.210*** (0.0689)	0.439*** (0.0637)
Lying type (<i>LType</i>)	0.576*** (0.0647)	0.638*** (0.0607)	0.0434 (0.100)
$FP_{t-1} \times LType$		0.0126 (0.0853)	-0.0831 (0.0811)
$CP_{t-1} \times LType$		-0.164* (0.0846)	-0.0776 (0.132)
$FR_{t-1} \times LType$		-0.0940 (0.0868)	-0.145 (0.155)
Rolled Number = 1	0.193*** (0.0662)	0.195*** (0.0650)	0.345*** (0.0637)
Rolled Number = 2	0.152** (0.0622)	0.151** (0.0615)	0.400*** (0.0638)
Rolled Number = 3	0.164** (0.0642)	0.168** (0.0648)	0.311*** (0.0613)
Rolled Number = 4	0.102 (0.0703)	0.103 (0.0697)	0.0713 (0.0476)
Rolled Number = 6	-0.0806* (0.0469)	-0.0887* (0.0499)	-0.230*** (0.0570)
Round	0.00520* (0.00274)	0.00459* (0.00273)	0.00225 (0.00176)
Constant	-0.0609 (0.0377)	-0.0677* (0.0403)	0.0485 (0.0434)
N	729	729	1020
R ²	0.574	0.579	0.361

Notes: Standard errors in parentheses are corrected for clustering at the individual level. The results are robust to clustering of standard errors also at the session level.

Rolled Number = 5 is the omitted dummy variable.

CR_{t-1} is the omitted category among the four possible feedback experiences.

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 4-3a. Average predictive proportion of lying by feedback experience and lying type

	Honest _{t-1}		Lie _{t-1}	
	Positive Feedback _{t-1}	Negative Feedback _{t-1}	Negative Feedback _{t-1}	Positive Feedback _{t-1}
	Correctly Rewarded _{t-1}	Falsely Punished _{t-1}	Correctly Punished _{t-1}	Falsely Rewarded _{t-1}
Honest Type	0.043	0.140	0.325	0.254
Dishonest Type	0.761	0.870	0.878	0.877

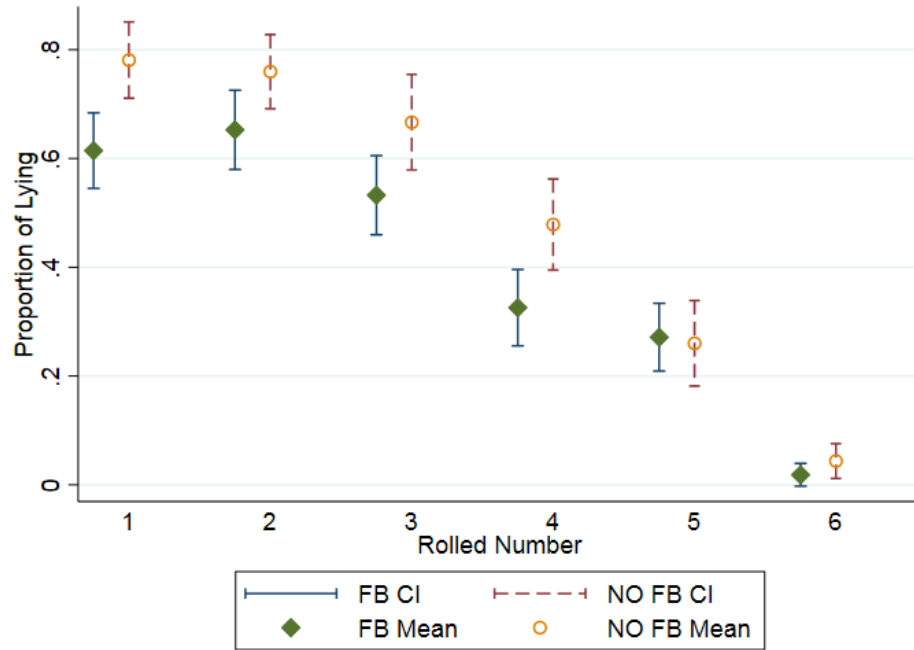
Notes: The average predictive proportion of lying is derived from column (2) of Table 3. Honest type is defined as one who tells the truth when encountering that number for the first time. Dishonest type is defined as one who lies when encountering that same number for the first time.

Table 4-3b. Average predictive proportion of lying using self-report honesty scores (*DType*)

	Honest _{t-1}		Lie _{t-1}	
	Positive Feedback _{t-1}	Negative Feedback _{t-1}	Negative Feedback _{t-1}	Positive Feedback _{t-1}
	Correctly Rewarded _{t-1}	Falsely Punished _{t-1}	Correctly Punished _{t-1}	Falsely Rewarded _{t-1}
Honest Type	0.222	0.313	0.656	0.660
Dishonest Type	0.270	0.278	0.627	0.563

Notes: The average predictive proportion of lying is derived from column (3) of Table 3, replacing *LType* with *DType*, which is a dummy taking the value of 1 if subjects rate themselves as “honest” or “very honest” (i.e., Honest Type), and 0 if subjects rate themselves as “moderate,” “dishonest,” or “very dishonest” (i.e., Dishonest Type).

Figure 4-1. Mean proportion of lying at each rolled number with 95% confidence interval.



Notes: The figure reports the mean proportion of lying at each rolled number, with the corresponding 95% confidence interval. The mean proportions of lying at rolled numbers 1 to 6 are 61.5%, 65.3%, 53.3%, 32.6%, 27.1%, and 1.8%, respectively, in the FB treatment and 78.1%, 76.0%, 66.7%, 47.9%, 26.0%, and 4.4% in the NO FB treatment.

Figure 4-2. Mean proportion of lying over time by condition

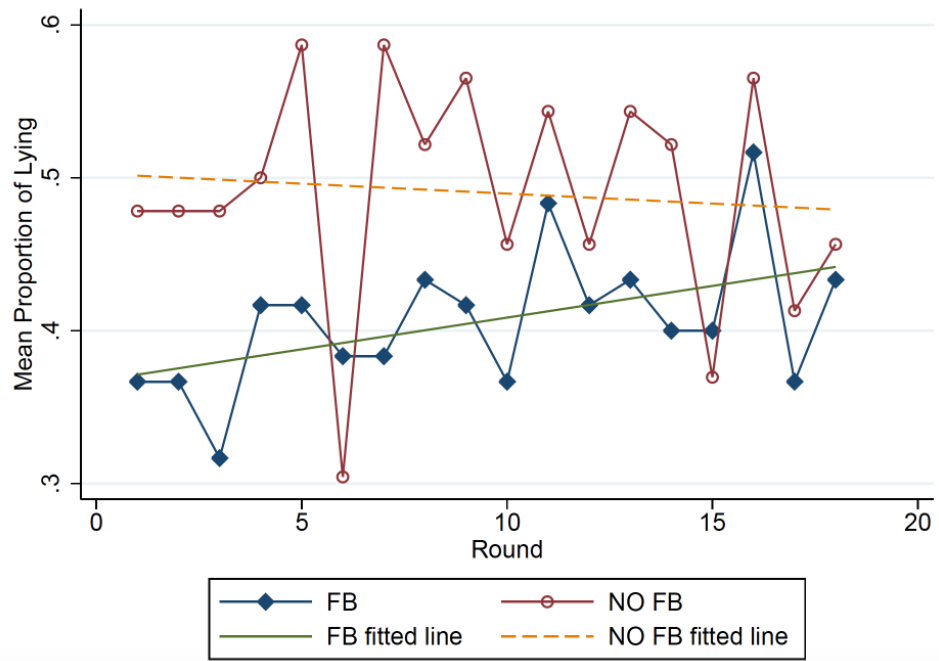
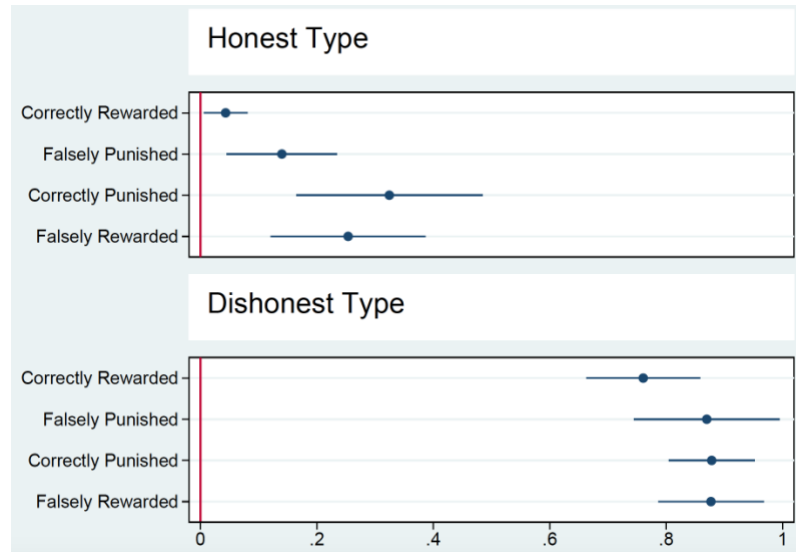
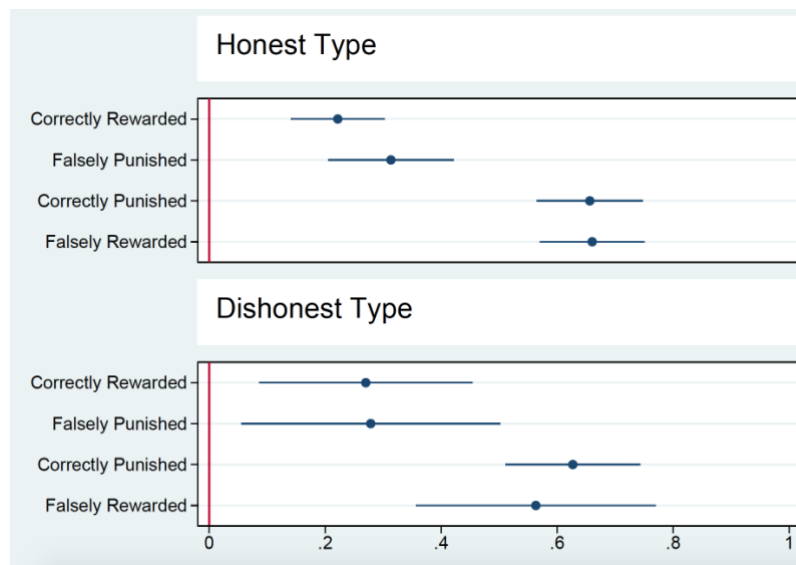


Figure 4-3a. Average predictive proportion of lying by feedback experience and lying types.



Note: Derived from column (2) of Table 3; the exact predictive proportions are reported in Table 3a.

Figure 4-3b. Average predictive proportion of lying using self-report honesty scores (*DType*)



Note: Derived from column (3) of Table 3; the exact predictive proportions are reported in Table 3b.

CHAPTER 5: RESPONSES TO LYING

5.1 Lying signals - report types

The reported number and receiver's payoff provide signals about the truthfulness to the sender. However, these do not convey a certain message of honesty or dishonesty. For example, a reported number 6 and zero payoff to the receiver may occur with probability 0.25 if the sender is honest; the same combination of 6 and zero payoff will occur with probability $0.75 \cdot (5/6)$ if the sender is dishonest and always reports 6 (he will be necessarily truthful $1/6$ of the time). Receivers may use a heuristic to simplify the calculation of conditional probabilities when facing unknown senders who may be honest, dishonest, or partially honest-dishonest who report (for strategic or other reasons) higher numbers than the rolled number but not 6 (partial lies). Hence the receiver has to employ some heuristic to evaluate the truthfulness of reported numbers.

An intuitive heuristic would be based on both signals. Low reported numbers and payoff 40 signify likely honesty and high numbers and zero payoff indicating dishonesty. A combination of low reported number and zero payoff may reflect a strategy by a sender who rolled a relatively high number but wishes to appear honest to get positive feedback and perhaps also gain the receiver's trust who will later accept high reported numbers as truthful and will again give positive feedbacks. This is not a likely scenario, and it is more likely that such combinations reflect truthful reports and realization of the 25%

probability of zero payoff, but they may nonetheless be regarded as ambiguous by some (naïve) receivers. High reported numbers and payoff 40 are also ambiguous and may lead some receivers to suspect dishonest report combined with the 25% probability of payoff 40. The heuristic classification of reports is summarized in Table 5-2.

	<i>Payoff 0</i>	<i>Payoff 40</i>
<i>Low reported number</i>	Ambiguous (A)	Honest-appearing (H)
<i>High reported number</i>	Dishonest-appearing (D)	Ambiguous (A)

Table 5-2. Heuristic classification of the message implied by different combination of reported number and receiver payoff

It is evident that this heuristic – or even a precise calculation of conditional probabilities of honest and dishonest reports – will lead to some incorrect conclusions. Receivers know that their judgment may be erroneous for any message, especially for ambiguous ones.

When choosing what feedback to send in response to a given message they have to balance the costs and benefits of the possibilities of giving erroneous positive feedback or erroneous negative feedback.

5.2 Related Literature and Hypotheses

The passive-feedback paradigm has been used extensively in experiments, where the feedback ranges from predetermined approval or disapproval options (Masclot et al., 2003) to comments in a chat room (Duffy and Feltovich, 2002; Blume and Ortmann, 2007;) to financial rewards and punishments (see survey: Feher and Schmidt, 2006). However, most research on passive feedback focuses on how it influences behavior of feedback-recipient, while less focus is put on feedback-sender. We attempt to fill this gap by exploring the feedback-sender's behavior: how they decide which feedback to send at what situation?

The stranger-partner comparison was introduced by Andreoni (1988), and has been used in games that have similar structure to our own (see survey in Andreoni & Croson, 2008). How people behave differently towards partners or strangers? Many previous literature find that people treat their counterparts better in partner condition. For instance, in public goods game, subjects contribute more in partner groups (Croson, 1996; Keser and Winden, 2000; Sonnemans et al., 1999). In trust game, Warnick and Slonim (2004) found that trustors send more to trustee in indefinite repeated game than in definite repeated game. Cochard, Van, and Willinger (2004) found that not only trustors send more in repeated game, but trustee also return more than in one-shot game, and interpreted their findings with "reciprocity hypothesis". Similarly, we hypothesize that people treat counterparts better in terms of giving less negative feedback in repeated interactions.

Hypothesis 1 (PTFB vs STFB). For a given message (combination of reported number and payoff), receivers will be more likely to provide negative feedback in one-time interactions (STFB) than in repeated interactions (PTFB).

What drives the behavioral difference between repeated interactions and one-shot interaction? Andreoni(1988) argued that a comparison between partner and stranger conditions indicates the role of strategies and learning played in the game. After ruling out learning effect by “restart” mechanism, the excessive contribution in partner conditions is attributed to rational strategic play. In following studies, Andreoni (1990) and Croson(1998) suggest some alternative explanations, including “warm glow”, altruism, and reciprocity, to explain higher level of cooperation among partners.

Our study builds upon previous work that acknowledges the difference between partner and strangers stems from learning, strategic and non-strategic play, and attempts to parse out the effect of strategy. First, we elaborate each of them in the current context:

1) Learning: partners can learn more about their counterpart’s honesty level due to accumulative knowledge during repeated interactions. Feedback provision not only reflects an immediate response to lies/truth-telling, but also encompasses influence from the past.

2) Strategy: The strategic objective is to induce the sender to behave honestly in the future. In experiments with costly punishment, many participants punish norm violators

despite the cost to themselves, and this tends to generate fewer subsequent violations of norms (e.g., Fehr and Fischbacher 2002). Similarly, due to repeated interactions, if receivers believe that feedback would affect senders' behavior in the next round, partners may intentionally give certain types of feedback for the best self-interests. For instance, a person who suspects that she was lied to may strategically give positive feedback in the anticipation of positive reciprocity – truth telling –in the future. People may abstain from giving negative feedback all the time (even if they suspect that their counterpart has lied) if they fear that giving erroneous criticism to an honest report may cause retaliatory lies thereafter.

One of the difficulties to identify strategic behavior is that it pertains large variations across individuals – people hold different beliefs about what kind of strategy works may behave differently. For instance, previous research captures many types of strategies in prisoner dilemma and public goods game, such as tit-for-tat (i.e. play the action that was previously played by the other), pavlov (win-stay, lose-shift), or grim trigger (punishment thereafter). However, the most important distinction between strategic and non-strategic response is whether one is given out of best self-interests.

3) Non-strategic motives: aside from strategy for self-interest, non-strategic feedback may be used to express emotions or in-group favoritism. The emotional reaction is a response to a perceived good, bad or ambiguous act by the sender. The emotional response of being treated fairly or unfairly is a visceral reaction that generally reflects a receiver's norms and moral judgment. Moral approval arouses positive emotions and

disapproval negative ones (Haidt 2001). In addition, though we never induce participants to form the sense of in-group, we cannot rule out the possibility that partners may regard each other as a group for mere repeated interactions, and thus behave friendly towards a long-term counterpart.

In STFB, receivers balance the emotional-moral benefits to themselves of telling senders what they think, and if they care at all about senders, the psychological cost to them from being wrongly accused or wrongly praised. In PTFB, the same considerations apply, and also the consideration how senders will respond to feedback in future interactions. In repeated interactions, senders can react to evaluative feedback with financial rewards or punishment, the reward by being honest and punishment by being dishonest. Hence receivers may worry about giving undeserved negative feedback, which may lead some receivers to give more positive feedback than they would in one-time interactions, especially when the risk of being wrong is higher – in response to ambiguous messages.

***Hypothesis 2.** Behavioral differences between PTFB and STFB hypothesized in **H1** are driven by strategic responses: receivers are strategically rather than genuinely nicer towards partners than strangers, to maximize self-interests.*

To offer more evidence on whether strategy is the dominant factor drives the behavioral difference between STFB and PTFB, we incorporate a novel way to isolate the effect of strategy by introducing hypothetical conditions, in which strategy is eliminated: we let receivers interact with strangers and partners in the same way as before, but only allow

them to give *hypothetical feedback*. As subjects respond with the understanding that hypothetical feedback will not be sent out and thus, any strategy to affect future interactions will not take effects, we expect a strong decrease for strategic type of feedback. However, partners can still learn about their counterparts' honesty level and can also transfer their emotions to the type of feedback they choose.

Therefore, by comparing actual and hypothetical feedback among partners (PTFB vs. PT), the conditional difference, if any, represents the existence of strategic feedback. However, among strangers (STFB vs. ST), we should expect no difference as strategic feedback play is never possible among one-shot interaction. In addition, any behavioral difference in hypothetical feedback between strangers and partners (ST vs PT) should stem from learning or non-strategic differences, such as exhibiting stronger emotions or being more altruistic towards partners.

In sum:

- **PTFB-STFB=strategy+learning+non-strategic**
- **PTFB-PT=strategy**
- **STFB-ST=no difference**
- **PT-ST=learning+non-strategic**

An important assumption for our analysis is that strategic motivation is substantially lower in repeated interactions when feedback is hypothetical, whereas non-strategic motivation and learning remains. Previous studies offer supports for this assertion. For

instance, an fMRI study by Kang et al. (2011) provides neuroscience evidence of real versus hypothetical choices. They find that while there exists large neuro-level overlap between those two choices, stronger neural activity shows up during real choices, particularly, activity in valuation and cognitive control areas (e.g., mOFC, ACC, caudate, inferior frontal gyrus), appears more responsive. Moreover, FeldmanHall et al. (2012) find that people are more likely to make self-serving moral decisions in real moral situation than in hypothetical one. And by enhancing the salience of personal gains, subjects' hypothetical responses get closer to real choices. This result aligns with our argument that strategic plays to gain maximum self-interest is an important aspect in real situation, but not so much in hypothetical ones.

***Hypothesis 2a (PT vs PTFB).** Hypothetical feedback in PT is more negative than feedback in PTFB.*

***Hypothesis 2b (PT vs ST).** Hypothetical feedback in PT will be indifferent to hypothetical feedback in ST.*

***Hypothesis 2c (ST vs STFB).** Hypothetical feedback in ST will be indifferent to actual feedback in STFB.*

To offer more insights on the strategic play in PTFB, we investigate the conditional difference in more details by comparing across conditions the dynamics of feedback provision over current and past experience, and the evolution of feedback over time.

Repeated engagement in similar interactions provides an opportunity to learn about the senders. In STFB, learning concerns the participant pool rather than any individual sender. In PTFB, learning concerns the same sender. What receivers learn from the two signals they receive in each interaction varies with the mode of thinking of receivers. Some may take into account little of past experience and react predominantly to the current message (akin to zero-level thinking), others may keep track of multiple past messages and interpret them carefully, with other participants recalling the immediate past message. Some participants may consider not only the effect of their feedback on senders discussed earlier but may also be aware of how senders responded to past feedback. Receivers learn about the honesty of senders, but the effects of past information are weak for most receivers because of their complexity.

Hypothesis 3. *Experience affects the choice of current feedback in repeated interactions (PTFB and PT), but not in one-shot interactions (STFB and ST).*

Hypothesis 3a. *The effects of the type of message on feedback provision weaken over time – the strongest effect comes from current messages, followed by the effect of message in the previous round.*

The effects of past experience and other considerations in repeated interactions with the same partner (PTFB) do not matter much in a series of one-time interactions with

strangers (STFB). Experience accumulated with different senders may affect the decision-making of some receivers in STFB but less than in PTFB.

Information about a sender's honesty is most valuable when the message received is ambiguous. A sender who appeared mostly honest in the past will be more likely to be treated as honest in case of an ambiguous message as compared to a sender whose history of messages indicated mostly dishonesty.

***Hypothesis 3b.** A current ambiguous message preceded by mostly honest-appearing messages will more likely to earn a positive feedback as compared to one that was preceded by mostly dishonest-appearing messages.*

5.3 Results

In the following analyses, for tractability, and more importantly, to reflect plausible interpretations by receivers of their experience, we do not use the 6x2 combinations of reported numbers and payoffs. Instead, we use the heuristics described in Table 2 to identify *report types*, based on combinations of report number and receiver payoff, to represent the signals (and experience) receivers get. We use report types as our principal explanatory variable for the response of receivers.

Dishonest reports are combinations of reported number 5 or 6 and payoff 0 ECUs, thus very lies; *honest reports* are combinations of reported number 1, 2, 3 or 4 and payoff 40

ECUs, thus likely to be truthful; *ambiguous reports* are combination of reported number 5 or 6 and payoff 40 ECUs or reported number 1, 2, 3 or 4 and payoff 0 ECUs, thus more difficult to conjecture whether they are lies or truth as compared to the “honest” and “dishonest” reports. We chose the cutoff point by examining the proportion of negative and positive feedback at each reported number and payoff (see analysis in Appendix A). We experimented with different cutoff points as well as with distinguishing two types of “ambiguous” report reports (high reported number and high payoff and low reported number and low payoff), with similar findings to those reported below.

We proceed in two steps, first exploring the comparisons among STFB, PTFB, ST, and PT, then evaluating the effect of past and current experience on feedback provision across conditions.

First, we examine differences in provision of feedback among the four conditions - feedback transmitted to senders in STFB and PTFB and hypothetical feedback in ST and PT, in Table 5-4, which presents an OLS linear probability estimation, for ease of interpretation (logistic models yield similar results). The dependent variable is the receiver provision of feedback (FB) in round t ($t=1, 2, \dots, 18$), where $FB = 1$ if feedback is negative and $FB = 0$ if positive. All analyses control for individual characteristics, round and adjusted standard errors clustered by individual-level.

Column (1) of Table 5-4 examines the treatment difference between STFB and PTFB, after controlling for report types, round, and gender. We find support for **H1**: receivers

are significantly less likely to give negative feedback to partners than to strangers, given observation of the same report type. (The estimate on “condition” is similar if we control for reported number and payoff instead of report type). Receivers are more lenient towards partners than strangers.

However, it remains to be seen what is the explanation behind such leniency. Are partners just try to be nice (non-strategic), or to seek best self-interest (strategic), or simply because they know more (learning)? We explored this questions with a discussion of remaining columns of Table 4, and find support for **H2** – strategic leniency. First of all, we do not find differences in strangers’ feedback provision between actual and hypothetical conditions(**H2b**), nor any difference between partners and strangers in hypothetical conditions (**H2c**), which implies that partners’ mere learning of sender’s history does not explain why there exist a strong forgiving tendency.

Moreover, just as partners are less likely to give negative feedback than strangers, partners in actual feedback conditions are less likely to give negative feedback than partners in hypothetical conditions, based on columns (3) of Table 5-4, supporting **H2a**. The similar magnitude implies that partners are likely to be strategically lenient.

Next, we explored the dynamics of feedback provision across the four conditions. The effect of report type history on feedback provision is presented in Table 5-5, where the omitted report type is Honest report (at t, t-1, and t-1). The table includes report types for the current round, previous round (t-1) and one round before that (t-2). The association

between past experience and current feedback choice only shows up in PTFB but not in STFB and ST, as predicted in **H3**. Besides, such relationship in PTFB weakens over time: report types in t always matter most, followed by report types in $t-1$, whereas the effects of $t-2$ is small and marginal. These findings support **H3a**.

Surprisingly, we history does not associate with current feedback in PT as in PTFB, which indicates that though receivers in PT and PTFB both can learn about senders' type via repeated interactions, only receivers in PTFB use what they learn of the past to determine current feedback choices.

Next, we investigate further how do receivers use the past in Table 5-6. The independent variables are combinations of current round report type and report type history. The variables are titled using H for Honest report, D for Dishonest report, and A for Ambiguous report, so that, for example, AtHt-1 means that the receiver faced an ambiguous report in the current round and an honest report in the previous round.

The results from Table 5-6 are summarized in the four panels of Figure 5-1, predicting the probability of the receiver providing negative feedback in the current period in response to each type of report and conditioned on type of report in the previous round. In STFB (Figure 5-1a), current report type is the most important determinant of feedback regardless of previous round experience. In contrast, receivers in PTFB (Figure 5-1b) weighed previous round experience heavily when they face ambiguous reports in the current round; compared to receiving an ambiguousness report in $t-1$, receivers are more

likely to give sender-partners negative feedback if they had a dishonest(t-1) experience and less likely to give negative feedback if they had an honest(t-1) experience.

In face of a current dishonest report, receivers are somewhat forgiving if the t-1 report was honest, and being most likely to give negative feedback to those whose reports were dishonest in both t and t-1, and even if the report was ambiguous in t-1. However, when faced with a dishonest report the likelihood of providing negative feedback in STFB is always higher than in PTFB (regardless of history).

By looking at Figure 5-1, we find that the leniency predominantly happens in front of current dishonest reports. While it is unintuitive that partners become forgiving towards current dishonest reports, instead of ambiguous or honest ones, it is reasonable if such leniency is a strategic way to earn future reciprocity. By comparing PTFB (Figure 5-1b) and PT (Figure 5-1d), the main difference lie at how partners treat dishonest or ambiguous report with an honest history. We observe “conditional reciprocity”: if the sender has shown honesty in the past, receivers are more encouraged to induce their future honesty by giving positive feedback to dishonest reports. This is in contrast to a “pure reciprocity” in which people will send positive feedback as a “reciprocity” to honesty, without expectation of future pay-back.

Table 5-4. Difference in giving negative feedback, compared across conditions (OLS)

	PTFB vs. STFB (base = STFB)	PT vs. ST (base = ST)	PTFB vs. PT (base = PT)	STFB vs. ST (base = ST)
	(1)	(2)	(3)	(4)
Condition	-0.0801**	0.00516	-0.0787**	0.0227
	(0.0348)	(0.0450)	(0.0396)	(0.0402)
Ambiguous Reports (L0)	0.270***	0.285***	0.284***	0.264***
	(0.0445)	(0.0530)	(0.0479)	(0.0487)
Ambiguous Reports (H40)	0.177***	0.270***	0.201***	0.250***
	(0.0298)	(0.0383)	(0.0330)	(0.0323)
Dishonest Reports	0.704***	0.641***	0.588***	0.752***
	(0.0355)	(0.0428)	(0.0404)	(0.0347)
N	1926	1782	1926	1782
r ² _a	0.373	0.272	0.272	0.378
F	30.25	22.04	19.56	39.43
Controlling for report types (base = Honest reports), individual characteristics, round.				
* p<0.1 ** p<0.05 ***p<0.001				

Table 5-5. The effect of report type history on giving negative feedback, by condition (OLS)

	(1)	(2)	(3)	(4)
	STFB	PTFB	ST	PT
Honest Reports	0	0	0	0
	(.)	(.)	(.)	(.)
Ambiguous Reports (L0)	0.216***	0.172***	0.310***	0.266***
	(0.0296)	(0.0345)	(0.0490)	(0.0394)
Ambiguous Reports (H40)	0.794***	0.600***	0.744***	0.562***
	(0.0440)	(0.0537)	(0.0522)	(0.0500)
Dishonest Reports	0	0	0	0
	(.)	(.)	(.)	(.)
Honest Reports (t-1)	0.000713	0.0659*	-0.0213	0.0194
	(0.0361)	(0.0388)	(0.0579)	(0.0470)
Ambiguous Reports (t-1) (L0)	0.0512	0.147***	0.0166	0.0474
	(0.0329)	(0.0427)	(0.0592)	(0.0548)
Ambiguous Reports (t-1) (H40)	0	0	0	0

	(.)	(.)	(.)	(.)
Dishonest Reports (t-1)	0.0325	0.0388	-0.0299	0.00616
	(0.0324)	(0.0407)	(0.0440)	(0.0395)
Honest Reports (t-2)	0.0471	0.0736*	0.0148	0.0364
	(0.0310)	(0.0430)	(0.0455)	(0.0444)
Ambiguous Reports (t-2) (L0)	0.0528	0.0651	0.115	0.303**
	(0.0746)	(0.108)	(0.112)	(0.125)
Ambiguous Reports (t-2) (H40)	896	816	688	896
	0.444	0.310	0.319	0.277
Dishonest Reports (t-2)	55.34	12.20	57.85	29.06
Constant				

Figure 5-1. Predicted probability of giving negative feedback, by report type history

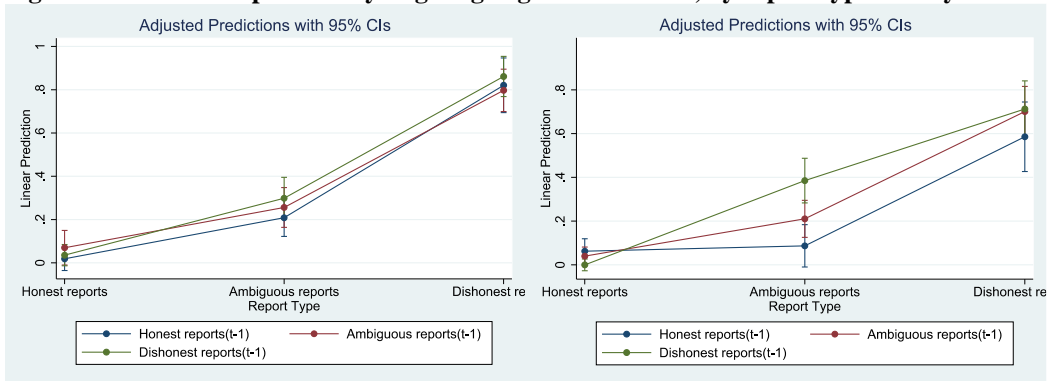


Figure 1a. STFB

Figure 1b. PTFB

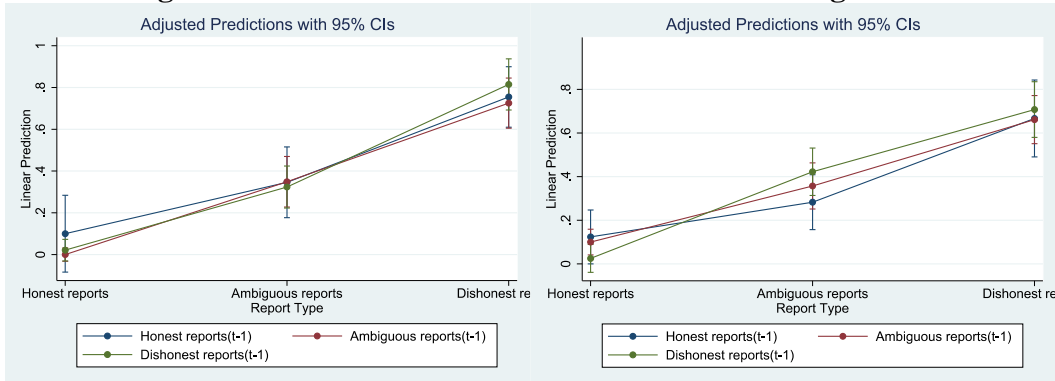


Figure 1c. ST

Figure 1d. PT

	Actual feedback		Hypothetical feedback	
	<i>STFB</i>	<i>PTFB</i>	<i>ST</i>	<i>PT</i>
<i>HtHt-1</i>
<i>HtAt-1</i>	0.0681	-0.0223	-0.0845	-0.0263
<i>HtDt-1</i>	0.0380	-0.0853**	-0.0689	-0.121
<i>AtHt-1</i>	0.208***	0.0276	0.259**	0.162*
<i>AtAt-1</i>	0.244***	0.138***	0.253**	0.235***
<i>AtDt-1</i>	0.298***	0.322***	0.214**	0.284***
<i>DtHt-1</i>	0.822***	0.510***	0.668***	0.517***
<i>DtAt-1</i>	0.798***	0.639***	0.630***	0.530***
<i>DtDt-1</i>	0.859***	0.638***	0.730***	0.547***

Note: *HtHt-1* = Honest report at t and Honest report at t-1, this is the baseline scenario.

At = Ambiguous report at t; *At-1* = Ambiguous report at t-1.

Dt = Dishonest report at t; *Dt-1* = Dishonest report at t-1.

5.4 Conclusions & Limitations

In this study, we attempt to explore how people react to lies. We investigate the motivations behind people's decision of giving feedback to potential lies. Our results imply that feedback provision is different between repeated long-term partner and one-shot stranger. The main difference comes from the manipulation of feedback to maximize benefits in future interactions.

We find evidence of “conditional reciprocity” among partners in which positive feedback is given to ambiguous or dishonest reports only if there is a possibility for senders to pay back (i.e., be honest) in future interactions. Suppose senders cannot acknowledge receivers' leniency, as in hypothetical conditions, receivers stop to be reciprocal.

As a consequence, we observe that repeated interactions lead to a higher proportion of false rewards on lying: 45.7% of lies are falsely rewarded in PTFB, compared to 32.1% in STFB.

One limitation of our study is that we cannot differentiate non-strategic motives and learning. Particularly, we cannot distinguish if the pure learning about someone's previous actions or the emotions associated with someone's previous action make a difference. For instance, suppose one gives positive feedback towards senders with an honest history track, is it because he/she learns that the sender may be an honest person, or because the receiver feels happy about not being lied to (e.g., emotions). Though this

is not the focus of our study, it leaves great potential for future research to examine the independent effect of learning, strategy, and non-strategic motivations.

A practical implication of our results in human resource management is that the accuracy of feedback could be in danger (i.e., not accurately reflect the true performance) even if it is only used for developmental purposes without monetary punishment or rewards. This is in contrary to the common wisdom that feedback used for incentive purposes is likely to suffer from mean convergence biases, but not those used for developmental purposes.

REFERENCES

- Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of public Economics*, 37(3), 291-304.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal*, 100(401), 464-477.
- Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, 891-904.
- Andreoni, J., & Croson, R. (2008). Partners versus strangers: Random rematching in public goods experiments. *Handbook of experimental economics results*, 1, 776-783.
- Brandts, J., & Schram, A. (2001). Cooperation and noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics*, 79(2), 399-427.
- Burlando, R., & Hey, J. D. (1997). Do Anglo-Saxons free-ride more? *Journal of Public Economics*, 64(1), 41-60.
- Croson, R. T. (1996). Partners and strangers revisited. *Economics Letters*, 53(1), 25-32.
- Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99(1), 197-199.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, 71(3), 397-404.
- Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review*, 108(2), 419-453.
- Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93, 293-300.
- Grosch, K., & Rau, H. A. (2017). Gender differences in honesty: The role of social value orientation. *Journal of Economic Psychology*, 62, 258-267.
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic theory*, 27(2), 245-252.
- Maschler, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1), 366-380.
- Palfrey, T. R., & Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *The American Economic Review*, 829-846.
- Sonnemans, J., Schram, A., & Offerman, T. (1999). Strategic behavior in public good games: when partners drift apart. *Economics Letters*, 62(1), 35-41.
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119(534), pp.47-60.
- Thielmann, I., & Hilbig, B. E. (2019). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology*, 71, 126-137.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle back: distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757

- Ben-Ner, A., & Putterman, L. (2009). Trust, Communication and Contracts,: An experiment. *Journal of Economic Behavior & Organization*, 70(1-2), 106-121.
- Ben-Ner, A., Putterman, L., & Wang, Y. (2018). Effort and Peer Pressure in Teams: An Experiment with Real Effort and Costly Feedback. Unpublished manuscript.
- Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*, 199(3), 457-480.
- Bicchieri, C., Dimant, E., & Sonderegger, S. (2019). It's not a lie if you believe it: Lying and belief distortion under norm-uncertainty. Available at SSRN.
- Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General*, 142(4), 1001.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88-97.
- Clay-Warner, J., Robinson, D. T., Smith-Lovin, L., Rogers, K. B., & James, K. R. (2016). Justice standard determines emotional responses to over-reward. *Social Psychology Quarterly*, 79(1), 44-67.
- Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34, 1-7.
- Dugar, S. (2013). Non-Monetary Incentives and Opportunistic Behavior: Evidence from a Laboratory Public Good Game. *Economic Inquiry*, 51(2), 1374-1388.
- Ellingsen, T., & Johannesson, M. (2008). Anticipated verbal feedback induces altruistic behavior. *Evolution and Human Behavior*, 29(2), 100-105.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, 58(4), 723-733.
- Faillo, M., Grieco, D., & Zarri, L. (2013). Legitimate punishment, feedback, and the enforcement of cooperation. *Games and economic behavior*, 77(1), 271-283
- Faravelli, M., Friesen, L., & Gangadharan, L. (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization*, 110, 160-175.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.
- Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes*, 119(1), 15-26.
- Gino, F., Gu, J., & Zhong, C.-B. (2009). Contagion or restitution? When bad apples can motivate ethical behavior. *Journal of Experimental Social Psychology*, 45(6), 1299-1302.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384-394.
- Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review*, 108(2), 419-453.
- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4), 191-210.
- Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93, 293-300.
- Gross, J., Leib, M., Offerman, T., & Shalvi, S. (2018). Ethical Free Riding: When Honest People Find Dishonest Partners. *Psychological science*, 29(12), 1956-1968.

- Hermann, D., & Ostermaier, A. (2018). Be Close to Me and I Will Be Honest. How Social Distance Influences Honesty.
- Hopfensitz, A., & Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *The Economic Journal*, 119(540), 1534-1559.
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.
- Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys*, 32(2), 357-387.
- Khalmetski, K., Rockenbach, B., & Werner, P. (2017). Evasive lying in strategic communication. *Journal of Public Economics*, 156, 59-72.
- Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1-2), 81-92.
- Masclot, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1), 366-380.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6), 633-644.
- Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current directions in psychological science*, 15(5), 260-264.
- Peeters, R., & Vrsatz, M. (2013). Immaterial rewards and sanctions in a voluntary contribution experiment. *Economic Inquiry*, 51(2), 1442-1456.
- Sánchez-Pagés, S., & Vrsatz, M. (2009). Enjoy the silence: an experiment on truth-telling. *Experimental Economics*, 12(2), 220-241.
- Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109(38), 15197-15200.
- Smith, A. 1759. *The Theory of Moral Sentiments*, originally printed for Andrew Millar, in the Strand; and Alexander Kincaid and J. Bell, in Edinburgh. Citations in the present paper are from the online version at <http://www.econlib.org/library/Smith/smMS.html>, Liberty Fund, Inc. 2000
- Vincent, L. C., Emich, K. J., & Goncalo, J. A. (2013). Stretching the moral gray zone: Positive affect, moral disengagement, and dishonesty. *Psychological science*, 24(4), 595-599.
- Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences*, 102(20), 7398-7401.
- Xiao, E., & Houser, D. (2009). Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange. *Journal of Economic Psychology*, 30(3), 393-404.
- Blume, A. and A. Ortmann (2007). "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria." *Journal of Economic theory* 132(1): 274-290.
- Cochard, F., P. N. Van and M. Willinger (2004). "Trusting behavior in a repeated investment game." *Journal of Economic Behavior & Organization* 55(1): 31-44.

Duffy, J. and N. Feltovich (2002). "Do actions speak louder than words? An experimental comparison of observation and cheap talk." Games and Economic Behavior **39**(1): 1-27.

Engle-Warnick, J. and R. L. Slonim (2004). "The evolution of strategies in a repeated trust game." Journal of Economic Behavior & Organization **55**(4): 553-573.

Fehr, E. and S. Gächter (2000). "Cooperation and punishment in public goods experiments." American Economic Review **90**(4): 980-994.

López-Pérez, R. and M. Vorsatz (2010). "On approval and disapproval: Theory and experiments." Journal of Economic Psychology **31**(4): 527-541.

APPENDIX A

EXPERIMENT INSTRUCTIONS

Welcome to the experiment!

Please read these instructions carefully. You may earn a considerable sum of money, depending on the decisions you and other participants make in this experiment.

The experiment will be conducted in Experimental Currency Units (ECUs). 1 ECU = \$0.1. You will be paid in dollars at the end of the experiment.

Participants, Rounds, and Pairs

- At the beginning of the experiment you will be informed of your role, either participant A or participant B. You will remain in that role until the end of the experiment.
- The experiment consists of 18 rounds. The computer randomly pairs one participant A with one participant B for each round (i.e., a new random pair will be generated in each round). There is a small chance that you may be paired with the same person twice during the experiment, but neither of you will know when. You will never know the identity of your counterpart.

Decisions

- Participant A rolls a 6-sided die. The roll will be carried out by the computer.
- Only A knows the number rolled.
- A reports to B a number r : "The number rolled is r ." The reported number r can be any number between 1 and 6, and is not required to be the rolled number.
- B receives the reported number r . B's payment will be determined by a lottery that can be good or bad depending on whether the reported number r was the actual rolled number.
- After B's payment is revealed for the current round, he/she will send a message to A:
 - "Thank you for being an honest person!", **or**
 - "It is not good to be a liar!"

Payment

You will only be paid for ONE round, which will be randomly selected at the end of the experiment.

The experiment will be conducted in Experimental Currency Units (ECUs). 1 ECU = \$0.1. You will be paid in dollars at the end of the experiment.

Participant A

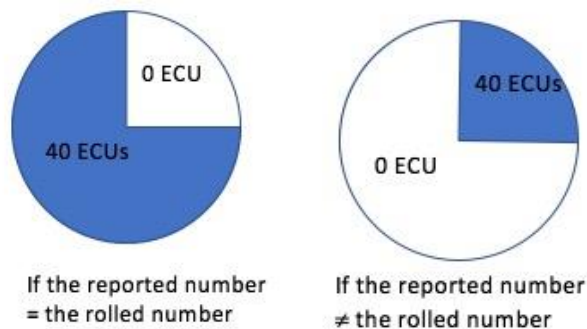
- Participant A receives 30 ECUs plus 10 times the number he/she reported to B.
- For example, if A reports 1, his/her payment will be $30+10 \times 1=40$ ECUs; if A reports 2, the payment will be $30+10 \times 2=50$ ECUs.

Participant B

- Participant B's payment is determined by one of two lotteries (lottery is carried out by the computer).
- If A reports the same number as the rolled number, B's payment will be determined by a good lottery: 75% chance of earning 40 ECUs and 25% chance of earning zero.
- If A reports a different number from the rolled number, B's payment will be determined by a bad lottery: 25% chance of earning 40 ECUs and 75% chance of earning zero.
- For example:
 - If A rolls 4 and reports 6, B's payment will be determined by a bad lottery (25% earning 40 ECUs and 75% earning 0 ECUs).
 - If A rolls 2 and reports 2, B's payment will be determined by a good lottery (75% earning 40 ECUs and 25% earning 0 ECUs).
 - If A rolls 5 and reports 2, B's payment will be determined by a bad lottery (25% earning 40 ECUs and 75% earning 0 ECUs).

Summary

- A knows the rolled number and the number he/she reports to B.
- B knows only the number reported by A.
- A and B will be informed separately only of their own payments.
- A's payment is $30 \text{ ECUs} + 10 * \text{the reported number}$.
- B's payment is based on one of two lotteries:



So the payment to B can be only one of two numbers: 40 or 0 ECUs. If it's the good lottery, 40 will be paid with 75%, whereas if it's the bad lottery, 40 will be paid with only 25%

- B will know only the payment but not whether it is the result of a good or a bad lottery.

APPENDIX B

Exhibit 1. A chooses the number to report to B.

Round 1 of 18

You rolled a 5.

Please choose the message you want to send to B:

- The number rolled is 1.
- The number rolled is 2.
- The number rolled is 3.
- The number rolled is 4.
- The number rolled is 5.
- The number rolled is 6.

[Next](#)

Reminder of Payment Calculation

- A's payment is $30 \text{ ECUs} + 10 * \text{the reported number}$.

If the reported number = the rolled number

If the reported number ≠ the rolled number

So the payment to B can be only one of two numbers: 40 or 0 ECUs. If it's the good lottery, 40 will be paid with 75%, whereas if it's the bad lottery, 40 will be paid with only 25%

Exhibit 2. B receives the reported number from A.

Round 1 of 18

Participant A sent you the following message:

"The number rolled is 5."

Click Next to display your payment.

[Next](#)

Exhibit 3. B chooses which feedback to send to A.

Round 1 of 18

Your payment for this round is 40 ECUs. This payment outcome was generated from the lottery(see reminder below).

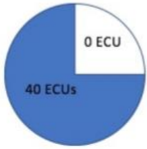
Please choose one of the following messages sent to A:

- Thank you for being an honest person!
- It is not good to be a liar!

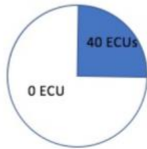
Next

Reminder of Payment Calculation

- A's payment is $30 \text{ ECUs} + 10 * \text{the reported number}$.
- B's payment is based on one of two lotteries:



If the reported number = the rolled number



If the reported number ≠ the rolled number

So the payment to B can be only one of two numbers: 40 or 0 ECUs. If it's the good lottery, 40 will be paid with 75%, whereas if it's the bad lottery, 40 will be paid with only 25%

Exhibit 4. A receives B's feedback

Round 1 of 18

Participant B said:

Thank you for being an honest person!

Next

APPENDIX C

Table c1. Frequency counts of reported numbers at each rolled number – FB treatment

Roll\Report	1	2	3	4	5	6	Total
1	74	0	4	3	12	99	192
2	1	58	4	13	9	82	167
3	0	0	86	5	13	80	184
4	0	0	0	118	2	55	175
5	0	0	0	0	145	53	199
6	0	0	1	1	1	160	163
Total	75	59	95	140	182	529	1080

Table c2. Frequency counts of reported numbers at each rolled number - NO FB treatment

Roll\Report	1	2	3	4	5	6	Total
1	30	0	3	2	5	97	137
2	0	37	8	5	6	98	154
3	2	2	38	1	1	70	114
4	2	2	0	73	2	61	140
5	1	1	1	2	91	27	123
6	1	0	1	0	5	153	160
Total	36	42	51	83	110	506	828

Notes: Tables c1 and c2 show the counts of senders' reported numbers at each rolled number in FB and NO FB, respectively. The diagonal represents, of course, honest reports and below it, smaller reported numbers than the ones rolled. When subjects lie, they lie to the maximum (report 6) most of the time; partial lies occur infrequently, 15.9% in FB and 13.1% in NO FB out of all lies. Those partial lies may not be rational, as suggested by Gneezy et al. (2018), and may be explained as subjects' desire to be perceived as honest by their counterparts, even without the possibility of explicit feedback in NO FB.

