AN APOLOGY FOR LYING

Keiko Aoki
Kenju Akai
Kenta Onoshiro

August 2010
Revised April 2013

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

An apology for lying

Keiko Aoki[a], Kenju Akai[b], Kenta Onoshiro[a]

[a] *Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 5670047, Japan.*

[b] *Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 1138656, Japan.*

Abstract: We investigate what types of social factors affect apology behavior for a previous lie and credibility levels for that apology. We abruptly provide subjects an opportunity to send an apology message after completion of the deception game (Gneezy, 2005) and investigate the effects of three main variables: burden of guilt based on the difference of stakes to be earned from lying and those from telling the truth (large vs. small), socio-economic background (students vs. non-students), and social distance (anonymity vs. face-to-face). The results show that none of these variables affect lying behavior. Students trust their counterparts less than non-students. After the deception game, students are less likely to send the message of having told a lie than non-students, but neither the burden of guilt nor social distance affects the motivation for sending such a message. Students give lower credibility levels to the additional messages sent after the deception game than non-students. Lifting anonymity raises credibility levels. The most powerful variables to affect apology behavior and credibility levels are subjects own previous decisions: whether to lie or not and whether to trust or not. That is, liars are more likely to send the message of having told a lie or keep silent than honest subjects, and trustors grant higher credibility than non-trustors.

Keywords: Apology, Lying, Disapproval, Deception game, Experiment, Face

JEL classification: C91, C72, D81

[*] Corresponding author. E-mail: k_aoki@iser.osaka-u.ac.jp. Address: Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaragi, Osaka 567-0047, Japan.

Tel.: +81-6-6879-8552;

Fax: +81-6-6879-8584.

E-mail: akai@css.t.u-tokyo.ac.jp (K. Akai), fge006ok@gmail.com (K. Onoshiro)

George was staggered by it, but quickly recovering himself he cried: "I cannot tell a lie, father, you know I cannot tell a lie! I did cut it with my little hatchet." – M. L. Weems, "The Cherry Tree."

1. Introduction

Apologies are useful instruments of communication for rebuilding damaged interactions and repairing trust in economics. Abeler et al. (2010) found that an apology results in better customer evaluations than monetary compensation in an eBay field experiment. Fischbacher and Utikal (2010) found that an apology prevents punishment in the sequential prisoner's dilemma game. Andreoni and Rao (2011) found that selfish allocators in the dictator game frequently reveal their remorse accompanied with an apology. Van Dijke and De Cremer (2011) found that an apology results in more frequent acceptance of an unfair ultimatum offer. De Cremer (2012) founds that apologies are more effective in promoting trust than financial compensation in the case of dividing gains. Leunissen et al. (2012) found that perpetrators are willing to apologize when the victim has a high likelihood of forgiving them. Schniter et al. (2012) found that apology is a useful tool for rebuilding trust in a variation of the repeated trust game. Ho (2012) founds that an apology induces high trust in a variation of the trust game.

While these studies consider the effects of apologies in strategic interactions with monetary incentives or strategic altruistic situations, our motivation in this paper is to investigate pure apologies that are independent of any monetary incentive or strategic altruism. We explore what kinds of social variables encourage these types of purely motivated apologies.

We replicate the deception game developed by Gneezy (2005) to employ lying as a means of having subjects feel guilty. In the original study, a sender chooses to send either a true or a deceptive message containing receiver's payoffs in the two options that detect actual payoffs between two subjects. Gneezy discovered that higher stakes achieved by lying induce a higher fraction of lying. The most important observation was that the deception game reduces selfishness relative to the dictator game using the same stakes, which implies an existence of lying aversion. That is, lying behavior in the deception game imposes a heavier burden of guilt than selfish behavior in the dictator game.

To investigate whether a pure incentive for making a confession with an apology exists, we abruptly provide a chance after the deception game for senders to apologize for their previous lies. Andreoni and Rao (2011) found that revealing remorse and/or apologizing along with allocating money enhances selfish behavior. To prevent such an effect and compare lying behavior with the original study of Gneezy (2005), we do not tell subjects that they have a chance to apologize later. The sender can send one of four messages: (i) "I told a lie. I am sorry." (ii) "I told a lie because I have some reasons." (iii) "I told the truth." (iv) "I told the truth. Please trust me." The sender could also choose to send no-message (i.e., keep silent). Afterwards, we ask receivers to self-report the credibility of each message to investigate evaluation of receivers' approval for each message.

We compare the distribution of these messages and the credibility of each message under three main treatment variables: burden of guilt (large vs. small), socio-economic

background (students vs. non-students), and social distance (anonymity vs. face-to-face). First, the burden of guilt is defined as the difference between the stakes that can be obtained by lying and those obtained by telling the truth. A larger difference in stakes induces a higher fraction of lying (Gneezy, 2005), which implies that the larger difference in stakes imposes a heavier burden of guilt on the subjects.

On the basis of the results of Gneezy, we suggest a hypothesis: *a larger burden of guilt is associated with a greater number of apologies*. We suppose that if the senders lie, they will incur a burden of guilt the weight of which will be proportional to the difference in stakes obtained. That is, the larger the difference in stakes, the heavier the burden of guilt incurs. Under this assumption, the larger difference in stakes will provide a stronger incentive to release the burden of guilt, so that a confession of guilt with an apology will be encouraged. On the other hand, since receivers do not know the stakes in the deception game, we propose the hypothesis: *the burden of guilt does not affect receivers' evaluations of credibility*.

Second, socio-economic background has various demographic components, such as age, gender, or income. To investigate the effects of these various components, Gächter et al. (2004) found students to be less trusting than non-students in the trust game. In the deception game, only the difference of gender has been considered as an effect of subject type. Dreber and Johannesson (2010) found that men are more likely to lie than women in situations with small stakes, while Childs (2012) found no difference between genders in situations with large stakes.

On the basis of the results of Gächter et al. (2004), we suggest a hypothesis: *students are less likely to apologize than non-students*. If students are less trusting than non-students, as in Gächter et al., students will be more likely to reduce their beliefs for earning approvals from the receivers by the apology than non-students. Therefore, they will apologize less than non-students. Additionally, the fact that students are less trusting in the trust game also implies the hypothesis: *students give lower evaluations of credibility than non-students*.

Third, social distance is represented by the degree of anonymity between senders and receivers. In the trust game, Scharlemann et al. (2001) showed that a facial expression elicits corporation under non-verbal communication environment. Holm and Kawagoe (2010) employed a bluffing game in which subjects played cards to deceive their counterparts. They discovered that lifting anonymity between players reduces lying. No study has evaluated the effects of social distance by using the deception game.[1] However, we can refer to Sutter (2009), who found that groups are less likely to lie than individuals. This result implies that lifting anonymity of the senders reduces lying.

On the basis of the arguments above, we suggest a hypothesis: *lifting anonymity reduces apology behavior*. Since reducing social distance encourages an emphasis on

---

[1] According to Roth (1995), face-to-face communication, one of the central issues in social interaction, encourages cooperative behavior in public goods games (Isaac and Walker, 1988; Brosig et al., 2003; Bochet et al., 2006) and prisoner's dilemma games (Frohlich and Oppenheimer, 1998; Bohnet and Frey, 1999b). It also increases offers in dictator games (Bohnet and Frey, 1999a, 1999b; Burnham, 2003; Charness and Gneezy, 2008).

others, lifting anonymity reduces senders' beliefs for earning approvals from the receivers, or it makes them less willing to let receivers down. On the other hand, the effect of increased cooperative behavior suggests the hypothesis: *lifting anonymity enhances evaluations of credibility.*

We obtain the following results. Approximately 40% of senders told a lies, and more than 60% of receivers trusted the messages from the senders. None of the main treatment variables affect the fraction of lying, but socio-economic background affects trust. That is, students are less trusting of the senders than non-students. Approximately 30% of liars confessed their lying, and half of these liars apologized. Students are less likely to prefer sending the message of having told a lie than non-students. Neither the burden of guilt nor lifting anonymity affects the motivation of sending the message of having told a lie. Lifting anonymity raises the credibility levels for every message. Students give lower credibility levels than non-students. Additionally, liars less frequently send messages of having told the truth than honest subjects. The receivers who have trusted senders once also grant higher credibility for the second message. The former result implies that subjects do not like to lie under a no-monetary-incentive environment. The latter result suggests that trustors avoid suspecting they are wrong.

The remainder of the paper is organized as follows. Section 2 presents the experimental design and procedures. Section 3 analyzes the results at the deception stage, and Section 4, the results at the apology stage. Section 5 discusses the pure motivation of guilty confession with apology. Finally, Section 6 summarizes the conclusions.

2. Experimental design and procedures

2.1 Design

Our game design includes two stages: a deception stage and an apology stage. In the deception stage, we replicate the deception game (Gneezy, 2005). The subjects were allocated to separate rooms, designated as Rooms A and B. Subjects in Room A are randomly matched with subjects in Room B.[2] Subjects in Room A plays the role of senders, and subjects in Room B plays the role of receivers.

Only senders are informed about the monetary payoffs for the two different options, A and B, as follows[3]:

> Option A: X yen to you and Y yen to the other student
> Option B: Y yen to you and X yen to the other student

Receivers know only that there were two options, A and B, but do not know the context. Then, the senders send receivers one of the two messages below:

> Message 1: Option A will earn you more money than Option B.

---

[2] All instructions and questionnaires are provided in the supplementary materials.
[3] We used Japanese yen (JPY). One US dollar was equivalent to 90 JPY at the time of the experiment.

Message 2: Option B will earn you more money than Option A.

Receivers receive the messages and choose one of the two options. Receivers know nothing but the amount of money they will earn at the end of the experiment. Therefore, receivers can not judge whether the sender has told a lie.

After the deception stage, the apology stage begins. This stage is only informed to senders. We provide a chance for senders to send a second message to reveal whether the first message is true or not:

(i) I told a lie. I am sorry.
(ii) I told a lie because I have some reasons.
(iii) I told the truth.
(iv) I told the truth. Please trust me.

Senders choose one of the four messages or send nothing (i.e., keep silent). Receivers can only know the message sent from senders.

2.2 Treatments and variables

Our design features three main treatment variables, as shown in Table 1. The first variable is the burden of guilt as defined by the difference in stakes between X and Y: large vs. small. In situations with a large difference in stakes, (X, Y) = (1,500, 500), and in situations with a small difference in stakes, (X, Y) = (1,000, 900). The second variable is the socio-economic background of the subjects. The subjects are divided into two groups: students vs. non-students. Student subjects were recruited from the campus of Osaka University. Non-student subjects were neighborhood residents invited by leaflets in the four most popular Japanese newspapers to voluntarily sign up for this experiment.[4] The third variable is social distance: anonymity vs. face-to-face. In the anonymity treatments, both senders and receivers stay in their separate rooms and the experimenters deliver the messages. In the face-to-face treatment, each pair move from their own rooms to a common space and the sender personally hands his or her message to the receiver without any verbal communication.

In this study, we focused on the effect of social distance under the treatment combination that achieved the highest fraction of lying. To this end, we first conducted a full matrix design of two types of burden of guilt and two subject groups, and then we explored what treatment combination induced the highest fraction of lying. The combination of large burden of guilt and student subjects (treatment A1000S) achieved the highest fraction of lying. Therefore, we conducted the face-to-face treatment under the same combination of variables, that is, F1000S.

Along with these main treatment variables, we investigate the effects of four types of variables self-reported in the questionnaire. The first type includes demographic variable, such as gender, age, and income. This is also used for evaluating differences in

---

[4] The leaflets were provided to a total of 30,000 households around the university.

socio-economic backgrounds between students and non-students. The second type is a psychometric scale such as General Social Survey (GSS) scales and trust scales used in many previous studies. The third type, employed only in the face-to-face treatment, includes information about acquaintance levels, subjects' impressions of their partners, and paired gender effects. These variables can be used to check social distance in each pair. The fourth variable is the willing to re-pair with the same senders if they have told a lie, which is defined by *Re-pair* and *Re-pair2*. We use these variables for evaluating the degree of receivers' aversion to the sender's lying. The definitions of the variables we employed in this paper are summarized in Table 2.

Tables 1 and 2 are about here

2.3 Procedures

We conducted the laboratory experiments at Osaka University. No one participated in more than one session. The anonymity treatments proceeded in the following steps:

Step 1: Subjects gathered in one room before the experiment. The experimenter read a consent form aloud in front of the subjects. The subjects were then divided by a lottery into two rooms, A (senders) and B (receivers).

Step 2: Senders and receivers were given separate instructions at the deception stage and read them silently and independently. Senders were given two cards with Messages 1 and 2. Receivers were given two cards describing Options A and B.

Step 3: Senders inserted one card (Message 1 or 2) into an envelope. The experimenter delivered the envelope to the paired receiver in the other room. The receiver read the message and inserted a card (Option A or B) into the envelope. Then, everyone answered a short questionnaire about their expectations.

Step 4: The experimenter delivered the envelope from the sender to the receiver. At this point, senders were given instructions about the apology stage along with four message cards containing Messages (i) to (iv). Senders inserted a card, if desired, into the envelope. The experimenter delivered the envelope to the receiver. If the sender decided not to send a second message, he or she kept the envelope on the desk and the receiver received nothing.

Step 5: Senders answered questions about their reasons for choosing a certain message or no-message with the questionnaire. Only receivers who had received a second message were given questionnaire and answered questions about the credibility of that message, their willingness to re-pair with the same person, and so forth. Then, everyone answered the general demographic and psychometric questionnaires.

Step 6: Everyone privately received their earnings in cash, which was the amount of money contained in the option receivers had chosen, plus a show-up fee of 500 JPY.

5

In the face-to-face treatment, Steps 3 and 4 were customized. In Step 3, each pair moved to a common space and stood face to face across a table separated by a partition to prevent receivers from seeing senders' decisions as they were being made. Senders inserted one of the two cards (Message 1 or 2) into an envelope and then personally handed it to their paired receiver. Verbal communication was prohibited during this procedure. After returning to their original rooms, receivers read the message and inserted one of two cards (Option A or B) into the envelope, which the experimenter then delivered. In Step 4, in addition to the same questions as in the anonymity treatments, both senders and receivers indicated whether they knew each other and gave their impressions of their partners.

3. Results in the deception stage

3.1 Overview

We conducted a total of 53 sessions with 640 persons, including 524 students and 116 non-students. The sessions lasted for 40–60 minutes. Each session consisted of 4–26 subjects, and they earned an average of approximately 1,500 JPY. Table 3 summarizes the socio-economic backgrounds of students and non-students. Females were 41% and 77% of students and non-students, respectively. The average age was 20.4 and 47.5 years for students and non-students, respectively. Most of the non-students were married. Income and size of household were also higher for non-students than for students.

Table 4 summarizes the results. In the deception stage, the fraction of lying was less than 0.5, and students lied more than non-students. The fraction of trust was higher than 0.6. In the apology stage, the distribution of second messages was identical across all treatments. The message, "I told the truth," which included Messages (i) and (ii), was sent most often; no-message was the second most popular choice; and the message, "I told a lie," which included Messages (iii) and (iv), was third. Among all treatments, F1000S achieved the highest fraction of messages of having told a lie. Messages of having told the truth obtained higher credibility levels than messages of having told a lie, except in treatment A100NS.

On the basis of these results, in this section, we analyze which variables affect lying and trust behavior at the deception stage. In Section 4, we move on to the apology stage to analyze what kinds of variables induce apology and the credibility of that message.

Tables 3 and 4 are about here

3.2 Lying behavior of the senders

Figure 1 compares the fraction of lies in all treatments. Treatment A1000S had the highest fraction of lies. The fraction in A1000S was more than 0.4, while other treatments had fractions below 0.4. However, according to Fisher's exact test, there is no significant difference among the treatments (p = 0.27). Although not statistically significant, our

observation is consistent with Gneezy (2005) and Sutter (2009), who found that larger stakes for students in the anonymity condition induces a higher fraction of lying. Table 5 summarizes the comparisons of the main treatment variables by Fisher's exact tests. None of the treatments affects lying behavior at the 5% significance level. At the 10% significance level, comparing A1000S and A1000NS, students lied more than non-students when stakes were large.

Next, we include answers to the questionnaire to evaluate the robustness of the above main treatment effects and explore what other factors affect lying behavior. We employ logit regressions of the lying behavior, which is a dummy variable taking the value 1 if a lie was told and 0 otherwise. Table 6 summarizes the results of this regression. We separate regressions in pooled all treatments and those in only the face-to-face treatment. In all treatments, based on regression (1), which considers the three main treatment effects, we cumulatively add demographics[5] to (2) and the GSS and trust scales to (3). In the face-to-face treatment, based on regression (1), which considers demographics, we cumulatively add the GSS and trust scales to (2) and acquaintance levels to (3).

In all treatments, the three main treatment variables still have no significant effects at the 5% level. Only *Non-student* has significance at the 10% level but this effect is not robust, as significance disappears in regressions (2) and (3). *Face* has negative coefficients but is not significant, but this observation is consistent with Holm and Kawagoe (2010), who found that the face-to-face treatment reduces the fraction of lying. Our result implies that face-to-face communication closes social distance and enhances altruistic behavior. The results in this regression are consistent with the results in Table 5. Therefore, we conclude that none of the main treatment variables affect lying behavior.

Neither demographics nor psychometrics affect lying behavior. No gender difference is found, consistent with Childs (2011), but not with Dreber and Johannesson (2008).

Figure 1 and Tables 5 and 6 are about here

3.3 Trust behavior of the receivers

We define receiver trust as the receiver choosing Option A (B) in response to the sender's Message 1 (2). At the beginning of this analysis, we compare trust behavior between Messages 1 and 2 to check whether the receivers are biased toward either of the messages. In any treatment, the senders' messages do not affect the receivers' trust at the 5% significance level according to Fisher's exact test. This is the natural result, because receivers know only the text of the message, as well as the face of the sender in the face-to-face treatment. Therefore, we pool trust across the two messages in the analysis below.

---

[5] We eliminated variables of income and marriage because of high correlation of subject variables.

7

Figure 2 compares the trust behavior among all treatments. Fractions of trust were more than 0.6 for all treatments, which is consistent with Gneezy (2005) and Sutter (2009), who found more than 60% of receivers trust messages. Especially, treatment A1000NS achieved the highest fraction of trust, whose value of more than 0.9 is significantly different from fractions for the other treatments according to Fisher's exact test (p = 0.016).

However, since receivers do not know the amounts of the stakes, we focus on the effects of subject groups and anonymity conditions. Comparing students with non-students in the anonymity treatments, the fraction of trust for students, 0.67, was less than that for non-students, 0.81, and this difference is marginally significant according to Fisher's exact test (p = 0.056). Even if we include treatment F1000S, the fraction of trust for students, 0.69, still remained less than that for non-students, and the difference also remains marginally significant according to Fisher's exact test (p = 0.078).

Comparing pooled anonymity treatments with F1000S, the fraction of trust for the pooled anonymity treatments, 0.71, was the same as that for F1000S, 0.71, and the difference is not significant according to Fisher's exact test (p = 0.899). Thus, lifting anonymity does not affect trust behavior.

Next, we include answers to the questionnaire to evaluate the robustness of the above main treatment effects and explore what other factors affect trust behavior. We employ logit regressions of trust behavior, which is a dummy variable taking the value 1 if the receiver trusted the message and 0 otherwise. Table 7 summarizes the results of this regression. In all treatments, based on regression (1), which considers the three main treatment effects, we cumulatively add demographics to (2), the GSS and trust scales to (3), and the willingness to re-pair to (4). In the face-to-face treatment, based on regression (1), which considers demographics, we cumulatively add the GSS and trust scales to (2), the willingness to re-pair to (3), and acquaintance levels to (4).

In all treatments, *Face* has no significance, and *Non-student* remains significant at the 10% level in all regressions. Therefore, we conclude that lifting anonymity has no effect, and students have marginally less trust than non-students. The latter result is consistent with Gächter et al. (2004), who found that students show less trust than non-students in the trust game.

Among demographic factors, *Female* has significant positive coefficients at the 5% level in all treatments, but significance disappeares in the face-to-face treatment. This result is consistent with Dreber and Johannesson (2008) but inconsistent with Childs (2011). *Age* has marginally significant negative coefficients at the 10% level in regressions (3) and (4). The significance levels increase to 5% in the face-to-face treatment. Therefore, we conclude that females show more trust, but older subjects show less trust.

Among psychometric variables, *TrustWorthiness* has a marginal effect only in regression (3), so we deny this result. *Re-pair* has strongly significant negative coefficients at the 1% level in all regressions, both in all treatments and in the face-to-face treatment. It seems odd that the receivers who want to re-pair with a liar has less trust for the senders' messages. According to the reasons why they chose "yes, they want" in the freeform questionnaire, many of them did not mind lying at all. A possible explanation for these answers is that these receivers have a tendency to trust others less, so that they do not care a great deal about others' lying. In the face-to-face treatment, *R_known* has a marginally

significant positive effect. Therefore, we conclude that the subjects who trust others less in the first place do not mind their counterpart's lying. Additionally, the acquaintance level enhances trust levels marginally.

Figure 2 and Table 7 are about here
----

4. Results in the apology stage

Here we focus on the second messages, including the guilty confessions with an apology at the apology stage. First, we analyze the senders' second messages and what factors affect their choices. Next, we consider the level of credibility receiver's attributed to those messages.

4.1 The apology

First, we consider the main treatment effect on the second messages including the apology. As shown in Table 8, we divide the second messages into three categories: no-message, the message of having told a lie, and the message of having told the truth. The distribution of these three messages was identical across all treatments. There is no significant difference of the distribution across all treatments according to Fisher's exact test ($p = 0.205$). Table 8 compares the three main treatment effects. None of the main treatment variables affect the distributions. Therefore, we pool the categorized message across all treatments.

Next, we compare the categorized messages between those who lied and those who were honest. Figure 3A compares the fraction of categorized messages pooled across all treatments between honest subjects and liars. Liars kept silent or sent messages of having told a lie more often than honest subjects, and liars sent a smaller proportion of messages of having told the truth. There is significant difference in the distribution of second messages between liars and honest subjects according to Fisher's exact test ($p < 0.001$).

Figure 3B shows the distribution of four original second messages and no-message. We refer to "I told a lie. I am sorry" as *apology,* "I told a lie because I have some reasons" as *self-justification,* "I told the truth" as *just telling the truth,* and "I told the truth. Please trust me" as *asking for trust.* The distributions of these four messages are different between liars and honest subjects according to Fisher's exact test ($p < 0.001$). The liars who sent a message of having told a lie were equally divided into those sending an apology and those sending a message of self-justification. The liars who sent a message of having told the truth preferred just telling the truth to asking for trust. This tendency was also observed in the honest subjects who sent the message of having told the truth; they also preferred just telling the truth. A possible reason for this same distribution is that excess asking will reduce credibility for the receivers. This is the same as excessive asking for money from the receivers in the dictator game reducing allocation, as Andreoni and Rao

9

(2011) discovered. We compare credibility levels of those two messages in Section 4.2 and consider the explanation above.

Next, we include answers to the questionnaire to evaluate the robustness of the above main treatment effects and explore what factors affect the second messages. We employ multinomial logit regressions of the categorized second messages with a baseline of no-message. Table 9 summarizes the results of this regression. In all treatments, based on regression (1), which considers the three main treatment effects along with an effect of lying behavior in the deception stage, we cumulatively add demographics to (2), and the GSS and trust scales to (3). In the face-to-face treatment, based on regression (1), which considers the effect of lying behavior in the deception stage along with demographics, we cumulatively add the GSS and trust scales to (2), and acquaintance levels to (3).

First, we consider subjects' motivation for changing their decisions from no-message to the message of having told a lie. For the three main treatment variables, only *Non-student* has a significantly positive value in regression (1) in all treatments at the 10% level. These results are consistent with Table 8. Therefore, we conclude that students have a weaker motivation for sending the message of having told a lie including an apology than non-students. This significance is marginal, but it supports the hypothesis that *students apologize less than non-students*. However, neither hypotheses for burden of guilt nor social distance is statistically supported.

Lie has significantly positive effects in all regressions in all treatments, with stronger significance in regressions (2) and (3) in the face-to-face treatment. Thus, liars prefer choosing the message of having told a lie rather than no-message. This result implies that even liars do not like to lie under the no-monetary-incentive environment.

For the other variables, *GSS* has significantly negative effects at the 10% level only in the face-to-face treatment. *S_Known* and *S_Impression* have significantly positive and negative effects, respectively, at the 10% level. Subjects who know their partners send messages of having told a lie, but subjects who receive a better impression from their counterpart keep silent more frequently.

Next, we investigate subjects' motivation for changing their decisions from no-message to the message of having told the truth. There are no significant effects of any of the three main treatment variables. This is consistent with the results shown in Table 8. Therefore, we conclude that none of the treatment variables affect subjects' motivation for sending the message of having told the truth.

Lie maintains a significantly negative effect in all regressions in all treatments at the 1% level. The effect becomes weaker but remains significant in the face-to-face treatment. Thus, we conclude that liars do not prefer sending the message of having told the truth rather than no-message even though they have met the receiver face-to-face. This implies that liars are aversion to lying under no-monetary-incentive environment.

In all treatments, no other variables have significant effects. In the face-to-face treatment, *Female* has a significantly positive effect at the 5% level only in regression (3). *TrustWorthiness* has a significantly negative effect at the 10% level in regressions (2) and (3). Thus, subjects with higher trustworthiness do not prefer sending the message of having told the truth. *S_Impression* also has a significantly positive effect at the 10% level, and this is the opposite result for the motivation for sending the message of having told a lie.

10

Subjects who receive better impressions of their counterparts more frequently send the message of having told the truth. These two results seem to indicate a kind of need for approval or an aversion to let down their counterpart (Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006). We discuss these possibilities in Section 5.

Figures 3A and 3B and Tables 8 and 9 are about here

4.2 Credibility of the apology

Here we compare the credibility of each message. The receivers' credibility levels were self-reported on a five-point scale: 1, "I absolutely believe it"; 2, "I somewhat believe it"; 3, "I don't know"; 4, "I somewhat do not believe it"; and 5, "I never believe it."

First, we investigate the treatment effects. As shown in Table 4, the message of having told the truth obtained higher credibility levels than the message of having told a lie in any treatment condition. According to Fisher's exact test, there is no significant difference in the distribution of credibility levels for each categorized message ($p = 0.147$). Comparing anonymity and face-to-face treatments, lifting anonymity increased credibility for both types of messages, but there are no significant differences for either type of message between the two treatments at the 5% level according to Wilcoxon rank-sum tests ($p = 0.102$ for the message of having told a lie and $p = 0.132$ for the message of having told the truth). Comparing students with non-students, non-students gave higher credibility levels for both messages, but there are no significant differences between these two groups at the 5% level according to Wilcoxon rank-sum tests ($p = 0.949$ for the message of having told a lie and $p = 0.395$ for the message of having told the truth). Therefore, we pool credibility across treatments. When we use the pooled samples, the message of having told the truth obtained higher credibility levels than the message of having told a lie, but the difference between these two messages is marginally significant according to the Wilcoxon rank-sum test ($p = 0.098$).

Second, we divide the categorized messages into the original four types of second messages. Figure 4A shows the credibility levels of each original message. The message of just telling the truth obtained the highest credibility. Asking for trust reduced credibility, and the credibility level for this type of message was the same as for apology and self-justification. This result supports our expectation in Section 4.1 such that the lower credibility of asking for trust makes senders prefer just telling the truth rather than asking for trust. Apology and self-justification achieved almost the same values. According to Fisher's exact test, there is no significant difference in the credibility levels among all messages ($p = 0.342$).

Figure 4B compares the pooled credibility levels for all messages between anonymity and face-to-face treatments. The face-to-face treatment achieved higher credibility levels for all second messages than anonymity treatments. Thus, lifting anonymity led to higher credibility, although the differences between these treatments are not significant at the 10% level in the Wilcoxon rank-sum test.

11

Next, we include answers to the questionnaire to evaluate the robustness of the above main treatment effects and explore what factors affect the credibility levels. We employ ordered logit regressions of credibility levels. Table 10 summarizes the results of this regression. In all treatments, based on regression (1), which considers the main treatment effects along with effect of trust behavior in the deception stage, we cumulatively add the type of second message to (2), demographics to (3), the GSS and trust scales to (4), and the willingness to re-pair to (5). In the face-to-face treatment, based on regression (1), which considers the effect of trust behavior in the deception stage along with the type of second message, we cumulatively add demographics to (2), the GSS and trust scales to (3), the willingness to re-pair to (4), and acquaintance levels to (5).

For the main treatments, *Face* has a significant positive effect in all regressions in all treatments at the 1% to 5% levels. Lifting anonymity induces higher credibility levels. This is consistent with the suggestion of Roth (1995). *Non-student* also has positive significance in regressions (3), (4), and (5) in all treatments. This result is consistent with results in the trust game in Gächter et al. (2004), in which students were found to be less trusting than non-students. Comparing the trust effect on credibility between students and non-students, students seemed to have a downward bias for trusting their counterpart. The results we obtained here support the hypotheses: *lifting anonymity enhances levels of credibility,* and *students give lower credibility levels than non-students.* However, the hypothesis for burden of guilt is not supported.

Trust has significantly positive effects in all regressions in all treatments at the 1% significance level, but this effect becomes much weaker in the face-to-face treatment and finally disappears. Thus, subjects who have already trusted their partner give higher credibility levels for the second messages. *SM_truth* has significantly positive effects in all treatments, except for regression (2). Receivers prefer the message of having told the truth to the message of having told a lie. This result may have two possible explanations: credulity or self-deception. The receivers who trusted the sender once did not have any incentive to deviate from their previous belief of trusting their counterpart, leading to credulity in the second encounter. However, if they suspected they were wrong to trust their counterpart regardless of the messages sent from the senders, they would be likely to become self-deception in the second encounter to avoid incurring a mental burden from being wrapped themselves up in suspicion. To distinguish credulity from self-deception, we cumulatively add the multiple variable *Trust*SM_truth*, which indicate credulity, in Table 10. However, we find no significant effect in all regressions both in all treatments and face-to-face treatment.

Among the other variables, *GSS* and *TrustBehave* have significantly positive effects in all treatments and in the face-to-face treatment. *TrustWorthiness* has a significantly positive effect only in regression (4) in all treatments. Thus, higher trust induces higher credibility levels. Some of these psychometric variables exhibit higher significance than those in Table 7 for the analysis of trust at the deception stage. *Re-pair2* has significantly positive effects in all treatments. The receivers who want to re-pair with the same person after receiving the second message exhibit higher credibility evaluations than those who do not. This result implies that the subjects who receive the good message want to re-pair, and subjects who receive the bad message do not mind to re-pair at all.

12

Figures 4A and 4B and Table 10 are about here

----

5. Discussion

Our results show that subjects have a weak incentive to apologize, and that this incentive is not affected by any of the main treatment variables. The variables that affected incentive most are the subjects' own previous behaviors, that is, whether to lie or not at the deception stage. Here we discuss the liars' incentive to send the second message or keep silent based on their previous lying behavior. We focus on lying aversion and disapproval aversion to compare our results with those of Andreoni and Rao (2011), the seminal paper in the study of apology.

5.1 Lying aversion

The results we obtained show that the combination of lying at the deception stage and choosing the message of having told the truth obtained higher monetary profits and credibility levels. If the liars naturally like to lie, they must choose this combination. Only 36% of liars, however, chose such behaviors, and the remaining liars chose no-message or the message of having told a lie. That is, 64% of liars did not lie a second time. This result implies that subjects do not like to lie under the no-monetary-incentive environment. Liars may have succumbed to venal temptations at the deception stage. If so, our result is consistent with Hurkens and Kartik (2009), who showed that a decision to lie is independent of monetary incentives as long as subjects prefer the outcomes that come from lying.

Comparing liars with honest subjects, liars are less likely than honest subjects to prefer sending the message of having told the truth to sending no-message, and this effect is significant at the 1% level. On the other hand, the liars are more likely than honest subjects to prefer sending the message of having told a lie to sending no-message, and this effect is significant at the 10% level. The former result suggests that liars do not prefer deception, and the latter result implies that they prefer the sincere behavior. That is, these results also support the hypothesis that liars do not like to lie in the no-monetary-incentive environment. This evidence allows us to investigate what motivation makes liars who do not like to lie choose either no-message or the message of having told a lie. Next, we discuss the differences in motivation between choosing no-message and the message of having told a lie.

5.2 Disapproval aversion

To consider the motivation for choosing either no-message or the message of having told a lie, we can refer to Andreoni and Rao (2011). When only allocators can express the message to the receivers along with their decision makings, approximately 70% of allocators provided no money. 39% of allocators expressed remorse and/or an apology and

13

18% kept silent. These messages were the top two among all messages. On the other hand, our results show that 28% of liars chose the message of having told a lie, half of them apologized, and 36% of liars kept silent. The fractions of those who expressed remorse and/or an apology was not so different from those in Andreoni and Rao's study, but the distribution of messages was reversed.

Our hypothesis of the explanation for this difference in the distribution of messages is a difference in the degree of aversion to disapproval or letting receivers down (Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006). We believe that this difference in the degree of guilt aversion is caused by the burden of guilt for lying. As the original study of Gneezy (2005) showed, the fraction of selfishness in the deception game is less than that in the dictator game. In other words, people do not like to lie, as lying imposes a moral burden. If this interpretation is correct, liars can easily infer that sending the message of having told a lie lets receivers down or causes them to reduce their approval levels.

This hypothesis is supported by the credibility levels self-reported by the receivers. The message of having told a lie obtained lower credibility levels than the message of having told the truth. This tendency would be easily inferred by the liars, which makes them more averse to disapproval. Since the liars have strong incentives to be averse to disapproval, they prefer keeping silent to expressing their remorse, more so than selfish allocators in the study of Andreoni and Rao (2011).

This idea is also consistent with Ekman and Friesen (1969) and Ekman (1992) in psychology, who categorized liars' emotions into three types: fear of being caught, deception guilt, and duping delight. The fear of being caught is similar to disapproval aversion. Lying enhances the fear of being caught compared to selfish behavior without lying, because lying is a morally worse behavior. This idea is also supported by Fischbacher and Utikal (2010), who found that an apology was not offered as an expression of remorse but in hopes of preventing punishment, and Utikal (2012), who found that a confession was more likely to be expressed in a no-punishment environment.

Our results also showed evidence of duping delight. Some honest subjects chose the message of having told a lie. The higher monetary incentive will impose a heavier burden of guilt, which makes them dislike lying. At the apology stage, however, there is no monetary incentive, so that those honest subjects will be free from such a burden of guilt. Therefore, some of them can enjoy telling a lie, which is considered duping delight.

5.3 Apology, self-justification, and asking for trust

The message of having told a lie has two additional messages: apology and self-justification. Our results showed that the credibility of each message exhibited different values in different situations. Self-justification obtained higher credibility levels than apology in the face-to-face treatment, but apology obtained higher credibility levels than self-justification in the anonymity treatments. This result implies that lifting anonymity grants more credibility to self-justification than to apology. We can interpret this as facial expressions enhancing the probability of others accepting morally bad behavior that is excused by self-justification.

Generally, people are curious about why they do bad things. In a court of justice, expression of a valid reason provides grounds for extenuating circumstances. Andreoni and Rao (2011) discovered that the empathy induces the remorse and/or apology. It is also supported in psychology, for instance, by Ohbuchi et al. (1989), who found that an apology led victims to have more favorable impressions of the perpetrator and to be more likely to refrain from severe aggression toward the perpetrator. Thus, in the face-to-face treatment, the self-justification may have satisfied the curiosity of the receivers and make the lie more acceptable, so that the credibility level for self-justification was higher than for an apology which has no justifications. Recently, a brain imaging study showed that an apology dismisses the motivation for anger but does not resolve the negative emotion of the anger (Kubo et al., 2012). Conceivably, providing the justification for having done a bad thing could be a useful way of dismissing the negative emotion of anger.

Finally, we also discuss the power of asking for trust. Asking for trust had a credibility level similar to apology, and it had a lower credibility level than just telling the truth. Asking for trust adversely affected receivers' credibility levels. This finding is similar to Andreoni and Rao (2011), who found that excess asking from the receivers in the deception game reduces allocation from the proposer. Excess asking often reduces credibility or increases disapproval.

6. Conclusions

In this study, we investigate what social factors encourage an apology for previous lying and raise the credibility for such a message of apology. Our design features three main treatment variables: the burden of guilt based on the difference of stakes to be earned from lying and those from telling the truth (large vs. small), socio-economic backgrounds of the subjects (students vs. non-students), and the social distance between senders and receivers (anonymity vs. face-to-face). We also evaluate the effects of other variables, including demographics and psychometric and other scales.

Deception stage: Approximately 40% of subjects told a lie, and none of the main treatment variables has a significant effect on lying behavior. No gender differences are evident in the lying behavior. More than 60% of receivers trusted the messages from the senders. Students are less trusting of the messages from senders than non-students. In addition, females are more trusting of senders than males.

Apology stage: Approximately 30% of liars and 70% of honest subjects confessed the truth. Half of these liars expressed an apology. For both liars and honest subjects, students are less likely to prefer sending the message of having told a lie including an apology than non-students. Neither the burden of guilt nor lifting anonymity affects the guilty confession with apology. Liars less frequently send messages of having told the truth than honest subjects. Receivers give higher credibility levels to the message of having told the truth than the message of having told a lie. Facial expression grants higher credibility to messages of self-justification than to apologies, and lifting anonymity increases the credibility levels for all messages. On the other hand, students give lower credibility levels than non-students. Additionally, receivers who have trusted the senders once are more likely to grant credibility a second time.

15

In the apology stage, there is no monetary incentive or strategic altruistic effects. Therefore, the apology was motivated by a tradeoff in the sender's mind between lying aversion and disapproval aversion. Our results showed that most subjects placed a higher value on lying aversion at a first and then disapproval aversion. This is a kind of lexicographical preferences. This preference is weakly affected by socio-economic background, as Gächter et al. (2004) showed in the trust game, but not by the burden of guilt or social distance. On the other hand, social distance and socio-economic background affect the credibility levels as one of the measures for disapproval the receivers attach to the senders' messages.

Our results showed that behavior at the deception stage has a stronger impact than the main treatment variables. Liars are less likely to prefer to lie in the apology stage than honest subjects, and trustors grant higher credibility than non-trustors. These results imply that behavior at the deception stage anchor behavior at the apology stage, even though these two stages are independent from the viewpoint of monetary incentives and strategic altruism. Especially, liars could not ignore their previous lying behavior as the sunk cost. This is the aversion to lying we found in this study.

The results we obtained shed new light on the fact that morally worse behavior raises the degree of disapproval aversion so that a voluntary apology is not as likely. Our results imply that expression of an apology requires approval or an acceptance of the apology from the counterpart, or provision of some monetary incentives instead of low approvals. This idea can be a basis for the discussion of monetary compensation schemes and whistle-blowing systems for inducing voluntary guilty confessions with apologies from economic criminals, as in situations of fraud in the marketplace and corner-cutting in procurement auctions. Furthermore, in the view point of law and economics, this idea suggests the importance of apologies in victim-offender reconciliation programs and victim-offender conferences in the U.S. and other countries to resolve negative emotions and encourage victims' approval of aggressors, as suggested by Mungan (2012).

Acknowledgement

References

Abeler, J., Calaki, J., Andree, K., & Basek, C. (2010). The power of apology. *Economics Letters, 107*(2), 233-235.

Andreoni, J., & Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics, 95*(7–8), 513-520.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization, 60*(1), 11-26.

Bohnet, I., & Frey, B. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior & Organization, 38*(1), 43-57.

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review, 89*(1), 335-340.

Brosig, J., Weimann, J., & Ockenfels, A. (2003). The effect of communication media on cooperation. *German Economic Review, 4*(2), 217-241.

Burnham, T. (2003). Engineering altruism: a theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior & Organization, 50*(1), 133-144.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74*(6), 1579-1601.

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization, 68*(1), 29-35.

Childs, J. (2011). Gender differences in lying. *Economics Letters, 114*(2), 147-149.

De Cremer, D. (2010). To pay or to apologize? On the psychology of dealing with unfair offers in a dictator game. *Journal of Economic Psychology, 31*(6), 843-848.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30*(2), 163-182.

Ekman, P. Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. 1992. In: WW Norton and Company, New York.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 288–298.

Fischbacher, U., & Utikal, V. (2010). On the acceptance of apologies. Bibliothek der Universität Konstanz. Research paper Series No.53.

Frohlich, N., & Oppenheimer, J. (1998). Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior & Organization, 35*(3), 389-403.

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

Gächter, S., Herrmann, B., & Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior & Organization, 55*(4), 505-531.

Ho, B. (2012). Apologies as Signals: With Evidence from a Trust Game. *Management Science, 58*(1), 1-18.

Holm, H. J., & Kawagoe, T. (2010). Face-to-Face Lying–an experimental study in Sweden and Japan. *Journal of Economic Psychology, 31*(3), 310-321.

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics, 12*(2), 180-192.

Isaac, R. M., & Walker, J. M. (1988). Communication and free-riding behavior - The voluntary contribution mechanism. *Economic Inquiry, 26*(4), 585-608.

Kubo, K., Okanoya, K., & Kawai, N. (2012). Apology Isn't Good Enough: An Apology Suppresses an Approach Motivation but Not the Physiological and Psychological Anger. *PloS one, 7*(3), e33006.

Leunissen, J. M., Cremer, D. D., & Reinders Folmer, C. P. (2012). An instrumental perspective on apologizing in bargaining: The importance of forgiveness to apologize. *Journal of Economic Psychology, 33*(1), 215-222.

Mungan, M. C. (2012). Don't Say You're Sorry Unless You Mean It: Pricing apologies to achieve credibility. *International Review of Law and Economics, 32*(1), 178-187.

Ohbuchi, K., Kameda, M., & Agarie, N. (1989). Apology as aggression control - Its role in mediating appraisals of and response to harm. *Journal of Personality and Social Psychology, 56*(2).

Roth, A. (1995). Bargaining Experiments. *Handbook of Experimental Economics*, ed. by J. Kagel and A. Roth. Princeton, NJ: Princeton University Press, pp.253-348.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*(5), 617–640.

Schniter, E., Sheremeta, R. M., & Sznycer, D. Building and rebuilding trust with promises and apologies. *Journal of Economic Behavior & Organization*. Forthcoming.

Sutter, M. (2009). Deception through telling the truth?! experimental evidence from individuals and teams. *The Economic Journal, 119*(534), 47-60.

Utikal, V. (2012). A fault confessed is half redressed—Confessions and punishment. *Journal of Economic Behavior & Organization, 81*(1), 314-327.

Van Dijke, M., & Cremer, D. D. (2011). When social accounts promote acceptance of unfair ultimatum offers: The role of the victim's stress responses to uncertainty and power position. *Journal of Economic Psychology, 32*(3), 468-479.

Supplementary materials

1. Instructions in anonymity treatment (original text in Japanese) [*with changes for face-to-face treatment*]

1.1 Instructions for Role A
There are instructions and two kinds of messages, A and B, in the "Message 1" envelope. First, please check the envelope for these. Next, we will explain the procedure in depth.

Overview
In the experiment, one person in this room and one person in another room will interact with each other. Each of them is randomly assigned to one of two possible roles: either Role A or Role B. First, Role A will send a message to Role B. Next, Role B will choose a monetary payment option after Role B has received the message from Role A. The monetary payments for both Roles A and B in the experiment depend on the choice of Role B.
Please read the instructions carefully. Do not talk or in any way try to communicate with other participants during the experiment. The earnings are calculated by summing the monetary payment in the experiment and the show-up fee (500 JPY). You will be paid privately and in cash at the end of the experiment.

Your role
Your role in this experiment is <u>Role A</u>. Role B, with whom you will be paired in the experiment, is in another room. You will send either Message 1 or 2 to Role B. No information about your partner will be provided to you during or after the experiment. At the end of the experiment, you will not meet Role B.

Alternatives
Two possible monetary payments are available to Role A and Role B in the experiment. The two payment options are as follows:

In Treatments A1000S, A1000NS, and F1000S, the following options are available:
Option A: 500 JPY to Role A and 1500 JPY to Role B
Option B: 1500 JPY to Role A and 500 JPY to Role B

In Treatments A100S and A100NS, the following options are available:
Option A: 900 JPY to Role A and 1000 JPY to Role B
Option B: 1000 JPY to Role A and 900 JPY to Role B

Role B will have to choose either Option A or Option B. The only information Role B will have is information sent by you in a message. That is, Role B will not know the monetary payments associated with each choice.

Message
We now ask you to choose one of the following two possible messages that you will send to Role B:

Message 1: "Option A will earn you more money than option B."
Message 2: "Option B will earn you more money than option A."

How to send a message to Role B
In the "Message" envelope, you are provided with two cards. Message 1 is written on one card, and Message 2 is written on the other. You will choose one of the two messages and then enclose that card in the "Message" envelope. You will put the card that you do not choose on the desk. After all Role A participants finish doing this, the experimenters will collect the "Message" envelopes and provide the selected message to Role B in each pair.
[*In the "Message" envelope you are provided with two cards. Message A is written on one and Message B is written on the other. You provide this "Message" envelope to Role B personally. First, you leave carrying two messages in the "Message" envelope at the experimenter's signal. All Role A players leave simultaneously. Next, you will choose one of the two messages and place the card on which it is written in the "Message" envelope at the table with partitions. You lay the card which you do not choose face down. If you want to check the instructions, please read the instructions affixed to the partitions. Last, you provide the "Message" envelope to Role B with whom you stand face to face. After you do this, please return to your seat. Please do not speak to Role B during this time.*]

Earnings
Role B chooses one of two alternatives, Option A or B, after Role B receives the message from Role A. The monetary payments for Roles A and B depend on this choice. The earnings are calculated by summing the monetary payment in the experiment and the show-up fee (500 JPY). You will be paid privately and in cash at the end of the experiment.

To repeat, you can know the monetary payments for both yourself and Role B. However, Role B will never know what sums were actually offered in the option that was not chosen (that is, he or she will never know whether your message was true or not). Furthermore, Role B will also never know the sums to be paid to you according to the different options.

If you have any questions, please raise your hand silently. The experimenters will answer your questions privately.

1.2 Instructions for Role B
There are instructions and two kinds of messages, A and B, in the "Message" envelope. First, please check the envelope for these. Next, we will explain the procedure in depth.

Overview
In the experiment, one person in this room and one person in another room will interact with each other. Both of them are randomly assigned to one of two possible roles: either

Role A or Role B. First, Role A will send a message to Role B. Next, Role B will choose a monetary payment option after Role B has received the message from Role A. The monetary payments for both Roles A and B in the experiment depend on the choice of Role B.

Please read the instructions carefully. Do not talk or in any way try to communicate with other participants during the experiment. The earnings are calculated by summing the monetary payment in the experiment and the show-up fee (500 JPY). You will be paid privately and in cash at the end of the experiment.

Your role

Your role in this experiment is <u>Role B</u>. Role A, with whom you will be paired in the experiment, is in another room. You will choose one of two alternatives, Option A or B. No information about your partner will be provided to you during or after the experiment. At the end of the experiment, you will not meet Role A.

Alternatives

Two possible monetary payments are available to Role A and Role B in the experiment. Role B will have to choose either Option A or Option B. The only information that Role B will have is information sent by Role A in a message. That is, Role B will not know the monetary payments associated with each choice.

Message

We ask Role A to choose one of the following two possible messages, which you will send to Role B:

Message 1: "Option A will earn you more money than option B."
Message 2: "Option B will earn you more money than option A."

One of the two messages, either Message 1 or Message 2, in the "Message" envelope from Role A is brought to you, Role B, by the experimenter.
[*One of the two messages, either Message 1 or Message 2 in the "Message" envelope from Role A is provided by the experimenter. First, you leave when the experimenters signal you to do so. Role A leaves simultaneously. Next, you will receive the "Message" envelope from Role A after Role A chooses one of the two messages, 1 or 2, and then places the card on which it is written in the "Message" envelope at the table with partitions. After you receive the message, please return to your seat. Please do not speak to Role A during this time.*]

How to send an alternative to Role A

You will choose one of the two alternatives, Option A or B, after you receive the message. You are provided cards on which are written two alternatives, Options A and B, in the "Decision" envelope. You will choose one of the two alternatives and then place that card in the "Decision" envelope. You will put the card that you do not choose on the desk. After all Role B participants finish doing this, the experimenters will collect the "Decision" envelopes and provide the selected alternative to Role A in each pair.

Earnings

You choose one of two alternatives, Option A or B, after you receive the message from Role A. The monetary payments for Roles A and B depend on this choice. The earnings are calculated by summing the monetary payment in the experiment and the show-up fee (500 JPY). You will be paid privately and in cash at the end of the experiment.

To repeat, Role A will know the monetary payments for both you and Role A when Role A chooses a message. However, you will never know what sums were actually offered in the option not chosen (that is, you will never know whether the message from Role A was true or not during and after the experiment). You will also never know the sums to be paid to you according to the different options.

If you have any questions, please raise your hand silently. The experimenters will answer your questions privately.

2. Instructions for second messages

There are instructions, the alternative from Role B, and four kinds of other messages, (i), (ii), (iii), and (iv), in the envelope.
First, please check the envelope for these.
Next, we will explain the messages in depth. You may choose to send or not send a message to Role B. When one sends a message, it is chosen from among the following four messages written on the cards:

I told a lie. I am sorry.
I told a lie because I have some reasons.
I told the truth.
I told the truth. Please trust me.

Finally, one who wants to send a message places it in the "Message 2" envelope, and all other messages are put in the first envelope that was used. One who does not send a message places all the messages in the first envelope. Then, the experimenters will collect them and send the "Message 2" envelope to Role B in each pair.

3. Questionnaires
3.1 Questionnaire for Role A
3.1.1 First questionnaire in anonymity treatment [*in face-to-face treatment*]

Please respond to the following.

[1. About Role B in the pair:
A: I have never seen him/her before.
B: I have seen but never talked to him/her.
C: I just exchange greetings with him/her.
D: I talk with him/her sometimes.
E: I often talk with him/her; he/she is a good friend of mine.

2. How do you feel about Role B?
A: Very good.
B: Good.
C: Neither good nor bad.
D: Not very good.
E: Not good at all.]

3. Do you expect Role B to believe your message is true?
A: Role B will always believe it is true.
B: Role B will sometimes believe it is true.
C: Don't know
D: Role B will sometimes not believe it is true.
E: Role B will never believe it is true.

4. Which option do you expect Role B to choose?
A: Option A.
B: Option B.

3.1.2 Second questionnaire

Please answer the following questions.

1. Did you send a message to Role B?
A: Yes→ To question 2.
B: No→ To question 3.

2. (For individuals who chose answer A in question 1) What kind of message did you send?
I told a lie. I am sorry.
I told a lie because there was a reason.
I told the truth.
I told the truth. Please believe me.

Reason:

3. (For individuals who chose answer B in question 1) Why did you *not* send a message?
Reason:

3.2 Questionnaire for Role B
3.2.1 First questionnaire in anonymity treatment [*in face-to-face treatment*]

Please respond to the following.

[1. About Role A in the pair:
A: I have never seen him/her before.
B: I have seen but never talked to him/her.
C: I just exchange greetings with him/her.
D: I talk with him/her sometimes.
E: I often talk with him/her; he/she is a good friend of mine.

2. How do you feel about Role A?
A: Very good.
B: Good.
C: Neither good nor bad.
D: Not very good.
E: Not good at all.]

1 *[3]*. Do you believe the message from Role A to be true?
A: I absolutely believe it to be true.
B: I sometimes believe it to be true.
C: Don't know.
D: I sometimes do not believe it to be true.
E: I never believe it to be true.

2 *[4]*. Would you pair again with this partner if the message is a lie?
A: Yes, I would.
B: No, I would not.

3.2.2 Second questionnaire

Please check the message in the "Message 2" envelope and answer the following questions:

1. What kind of message did you receive?
The message:

2. Do you believe the message from Role A to be true?
A: I absolutely believe it to be true.

24

B: I sometimes believe it to be true.
C: Don't know.
D: I sometimes do not believe it to be true.
E: I never believe it to be true.

3. Do you tell someone the contents of the message which you receive?
A: Yes, I do.
B: No, I do not.
C: Don't know.

4. Would you pair again with this partner if the message is a lie?
A: Yes, I would.
B: No, I would not.

Table 1. Experimental design

Treatment	Stakes of (X, Y)	Subjects	Social distance	# of pairs
A1000S	(1500, 500)	Students	Anonymity	70
A1000NS	(1500, 500)	Non-students	Anonymity	27
A100S	(1000, 900)	Students	Anonymity	63
A100NS	(1000, 900)	Non-students	Anonymity	31
F1000S	(1500, 500)	Students	Face-to-Face	129

Table 2. Definitions of variables

Variables	Description	Answer range
Face	Face-to-face treatment	1: Face-to-face treatment; 0: Anonymity treatment
Non-student	Non-student treatment	1: Non-student; 0: Student
Stake1000	The difference of stakes is 1000 or not	1: Difference of stake is 1000; 0: Others
Lie	Telling a lie or not	1: Sending Message 2; 0: Others
Trust	Role B follows the message from Role A or not	1: Follow the message; 0:Not follow
SM	Second message Role A sent after the deception game	0: No-message; 1: I told the truth; 1: I told the truth. Please trust me; 2: I told a lie. I am sorry; 2: I told a lie. I have some reasons
Credibility	Do you believe that the message sent from Role A is true?	1: I absolutely believe it; 2: I somewhat believe it; 3: I don't know; 4: I somewhat do not believe it; and 5: I never believe it.
SM_truth	Which second messages did receiver receive?	1: SM=1 ; 0: SM=2
Female	Female or Male	1: Female ; 0: Male
Age	Age	
Income	Last year, what was the total income in	1: less than 2.5 million yen;

	your family?	2: 2.5 – 4.0 million yen;
		3: 4.0 – 5.5 million yen;
		4: 5.5 – 7.0 million yen;
		5: more than 7.0 million yen.
Marriage	Are you married?	1: Yes; 0:No
Household	How many people do you live with?	1: 1 person ;2: 2 persons ;3 : 3persons; 4: 4 persons; 5: above 5 persons
GSS	Normalized sum of de-meaned normalized and resigned GSS fair, GSS help, and GSS trust	
	GSS fair: "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?"	1: Would take advantage of you; 2: would try to be fair; 1.5: depends; –: no answer/don't know
	GSS help: "Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?"	1: Try to be helpful; 2: just look out for themselves; 1.5: depends; –: no answer/don't know
	GSS trust: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"	1: Most people can be trusted; 2: can't be too careful; 1.5: depends; –: no answer/don't know
TrustBehave	Normalized and resigned sum of normalized Door unlocked, lend money, and lend possessions	
	Door unlocked: "How often do you leave your door unlocked?"	1: Very often; 2: often; 3: sometimes; 4: rarely; 5: never
	Lend money: "How often do you lend money to friends?"	1: More than once a week; 2: once a week; 3: once a month; 4: once a year or less

	Lend possessions: "How often do you lend personal possessions to friends?"	1: More than once a week; 2: once a week; 3: once a month; 4: once a year or less
TrustStranger	You can't count on strangers anymore	1: More or less disagree ; 0: More or less agree
TrustWorthiness	I am trustworthy	1: Disagree strongly; 2: Disagree somewhat; 3: Disagree slightly; 4: Agree slightly; 5: Agree somewhat; 6: Agree strongly
Re-pair	Do you want to be paired with the same person to share money even if he/she tells a lie?	1: Yes; 0: No
Re-pair2	Do you want to be paired with the same person after receiving the message?	1: Yes; 0: No
S_Known*	Relation to Role B	1: I have never seen him/her before; 2: I have seen but never talked to him/her; 3: I just exchange greetings with him/her; 4: I talk with him/her sometimes; 5: I often talk with him/her; he/she is a good friend of mine
S_Impression*	How do you feel about Role B?	1: Not good at all; 2: Not very good; 3: Neither good nor bad; 4: Good; 5: Very good
R_Known*	Relation to Role A	1: I have never seen him/her before; 2: I have seen but never talked to him/her; 3: I just exchange greetings with him/her; 4: I talk with him/her sometimes; 5: I often talk with him/her; he/she is a good friend of mine
R_Impression*	How do you feel about Role A (B)?	1: Not good at all; 2: Not very good; 3: Neither good nor bad; 4: Good; 5: Very good
Pair_MM*	Sender/receiver pairings	1: Male with male; 0: Others

Notes:* these variables are employed in the face-to-face treatment only.

Table 3. Socio-economic backgrounds

	Students		Non-students	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.41	(0.49)	0.77	(0.43)
Age	20.43	(2.52)	47.47	(13.17)
Income	2.18	(1.67)	3.59	(1.41)
Marriage	0.00	(0.06)	0.90	(0.31)
Household	2.18	(1.51)	3.20	(1.10)
No of Obs.	524		116	

Table 4. Overview of the result

	A1000NS	A1000S	A100NS	A100S	F1000S	Total
Lying	0.26	0.49	0.32	0.40	0.38	0.39
Trust	0.96	0.64	0.68	0.70	0.71	0.71
Second messages						
No-message	0.30	0.30	0.29	0.24	0.28	0.28
I told a lie	0.04	0.16	0.06	0.16	0.23	0.17
I told the truth	0.67	0.54	0.65	0.60	0.49	0.55
Credibility for						
I told a lie	3.00	3.09	4.00	3.70	3.80	3.63
I told the truth	4.06	3.45	3.80	3.78	3.92	3.79
No. of pairs	27	70	31	63	129	320

Table 5. Fisher's exact tests on senders' lying

Issues	Treatments	Difference*	p-value
Stake (S)	A1000S vs. A100S	0.09	0.382
Stake (NS)	A1000NS vs. A100NS	-0.06	0.773
Subject (High)	A1000S vs. A1000NS	0.22	0.066
Subject (Low)	A100S vs. A100NS	0.07	0.507
Face	F1000S vs. A1000S	-0.11	0.176

*The difference = left side treatment – right side treatment

Table 6. Logit regressions of lying

	All treatments			Face-to-face treatment		
	(1)	(2)	(3)	(1)	(2)	(3)
Face	-0.35	-0.34	-0.35			
	(0.29)	(0.29)	(0.30)			
Non-student	-0.64*	-1.06	-1.02			
	(0.34)	(0.68)	(0.68)			
Stake1000	0.18	0.18	0.19			
	(0.30)	(0.30)	(0.31)			
Female		-0.17	-0.16	-0.24	-0.10	-0.13
		(0.25)	(0.25)	(0.38)	(0.39)	(0.53)
Age		0.02	0.02	0.16	0.12	0.09
		(0.02)	(0.02)	(0.11)	(0.12)	(0.12)
Household		0.07	0.06	0.08	0.04	0.02
		(0.08)	(0.08)	(0.12)	(0.13)	(0.13)
GSS			-0.05		0.04	0.04
			(0.12)		(0.22)	(0.22)
TrustBehave			0.12		0.22	0.26
			(0.13)		(0.21)	(0.21)
TrustStranger			-0.36		-0.62	-0.75*
			(0.24)		(0.39)	(0.41)
TrustWorthiness			-0.02		-0.04	-0.09
			(0.09)		(0.15)	(0.15)
S_Known						0.27
						(0.21)
S_Impression						-0.37
						(0.26)
Pair_MM						-0.17
						(0.53)

Constant	-0.32	-0.72	-0.58	-3.81	-2.39	-0.52
	(0.24)	(0.54)	(0.67)	(2.32)	(2.51)	(2.84)
Pseudo R2	0.01	0.01	0.02	0.02	0.03	0.05
No. of Obs.	320	320	320	129	129	129

Notes: Dependent variable is Lying, taking the value 1 for sender's message 2 (that is, lying) and 0 otherwise. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7. Logit regressions of trust

	All treatments				Face-to-face treatment			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Face	0.17	0.10	0.17	0.24				
	(0.27)	(0.27)	(0.28)	(0.30)				
Non-student	0.75*	1.71*	1.80*	1.77*				
	(0.38)	(0.91)	(0.94)	(0.98)				
Female		0.70***	0.74***	0.50*	0.16	0.21	-0.04	-0.42
		(0.27)	(0.27)	(0.29)	(0.40)	(0.41)	(0.44)	(0.65)
Age		-0.04	-0.05*	-0.05*	-0.19	-0.19	-0.29**	-0.31**
		(0.02)	(0.03)	(0.03)	(0.12)	(0.12)	(0.14)	(0.14)
Household		-0.04	-0.05	-0.03	0.11	0.07	0.09	0.13
		(0.09)	(0.09)	(0.09)	(0.14)	(0.14)	(0.15)	(0.15)
GSS			0.05	0.04		-0.08	-0.12	-0.09
			(0.14)	(0.15)		(0.23)	(0.25)	(0.27)
TrustBehave			-0.01	0.03		0.20	0.13	0.10
			(0.15)	(0.16)		(0.25)	(0.27)	(0.28)
TrustStranger			-0.30	-0.27		-0.52	-0.48	-0.53
			(0.27)	(0.28)		(0.41)	(0.45)	(0.47)
TrustWorthiness			0.20*	0.19		0.25	0.28	0.31
			(0.11)	(0.12)		(0.17)	(0.19)	(0.19)
Re-pair				-1.50***			-1.65***	-1.84***
				(0.27)			(0.45)	(0.48)
R_Known								0.68*
								(0.35)
R_Impression								0.02
								(0.30)
Pair_MM								-0.62
								(0.68)

Constant	0.70***	1.35**	0.83	1.51*	4.39*	3.73	6.54**	6.22**
	(0.18)	(0.62)	(0.81)	(0.87)	(2.31)	(2.48)	(2.86)	(3.14)
Pseudo R2	0.01	0.04	0.05	0.13	0.02	0.05	0.14	0.18
No. of Obs.	320	318	317	317	129	129	129	129

Notes: Dependent variable is Trust, taking the value 1 if the receiver follows the sender's message and 0 otherwise. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8. Fisher's exact tests on senders' second messages

Issues	Treatments	Difference*			p-value
		No-message	I told a lie	I told a truth	
Stake (S)	A1000S vs. A100S	6.19	-6.03	-0.16	0.718
Stake (NS)	A1000NS vs. A100NS	0.60	2.15	-2.75	1.000
Subject (High)	A1000S vs. A1000NS	0.37	-12.38	12.01	0.269
Subject (Low)	A100S vs. A100NS	-5.22	-4.20	9.42	0.454
Face	A1000S vs. F1000S	2.09	5.45	-7.55	0.484

*The difference = left side treatment – right side treatment

Table 9. Multiple logit regression of sender's second messages

	All treatments			Face-to-face treatment		
	(1)	(2)	(3)	(1)	(2)	(3)
I told a lie.						
Face	0.56	0.53	0.59			
	(0.44)	(0.44)	(0.45)			
Non-student	-1.15*	-0.39	-0.36			
	(0.69)	(1.23)	(1.29)			
Stake1000	-0.30	-0.27	-0.21			
	(0.50)	(0.51)	(0.52)			
Lie	0.60*	0.61*	0.69*	0.97*	1.15**	1.16**
	(0.36)	(0.37)	(0.38)	(0.52)	(0.56)	(0.59)
Female		-0.34	-0.39	0.03	-0.08	-0.15
		(0.38)	(0.38)	(0.53)	(0.56)	(0.72)
Age		-0.03	-0.02	-0.20	-0.18	-0.26
		(0.05)	(0.05)	(0.16)	(0.16)	(0.18)
Household		-0.01	-0.01	0.02	-0.02	-0.04
		(0.13)	(0.13)	(0.17)	(0.18)	(0.19)
GSS			-0.29		-0.51*	-0.55*
			(0.18)		(0.30)	(0.32)
TrustBehave			0.02		0.07	0.09
			(0.19)		(0.29)	(0.31)
TrustStranger			0.53		0.65	0.36
			(0.39)		(0.58)	(0.60)
TrustWorthiness			-0.04		-0.16	-0.23
			(0.14)		(0.22)	(0.23)
S_Known						0.65*
						(0.34)
S_Impression						-0.68*
						(0.40)
Pair_MM						-0.23

						(0.74)
Constant	-0.76*	-0.09	-0.52	3.14	3.09	6.59
	(0.45)	(1.07)	(1.23)	(3.28)	(3.51)	(4.14)
I told the truth.						
Face	-0.22	-0.19	-0.22			
	(0.34)	(0.34)	(0.35)			
Non-student	-0.10	-0.97	-1.07			
	(0.37)	(0.71)	(0.73)			
Stake1000	-0.18	-0.25	-0.22			
	(0.34)	(0.35)	(0.35)			
Lie	-1.12***	-1.15***	-1.17***	-0.78*	-0.96*	-0.93*
	(0.28)	(0.28)	(0.28)	(0.46)	(0.49)	(0.50)
Female		0.29	0.31	0.47	0.76	1.30**
		(0.29)	(0.29)	(0.44)	(0.47)	(0.63)
Household		0.06	0.06	-0.03	-0.11	-0.10
		(0.10)	(0.10)	(0.14)	(0.15)	(0.15)
Age		0.03	0.03	-0.07	-0.10	-0.06
		(0.02)	(0.02)	(0.13)	(0.14)	(0.15)
GSS			-0.07		-0.14	-0.14
			(0.14)		(0.27)	(0.28)
TrustBehave			-0.05		0.05	-0.03
			(0.15)		(0.25)	(0.25)
TrustStranger			-0.21		-0.70	-0.75
			(0.28)		(0.49)	(0.51)
TrustWorthiness			0.03		-0.35*	-0.33*
			(0.11)		(0.18)	(0.19)
S_Known						-0.02
						(0.35)
S_Impression						0.57*
						(0.33)
Pair_MM						0.98

						(0.65)
Constant	1.34***	0.59	0.58	2.07	4.66	1.19
	(0.31)	(0.60)	(0.76)	(2.64)	(3.00)	(3.62)
Pseudo R2	0.07	0.08	0.09	0.06	0.12	0.18
No. of Obs.	320	320	320	129	129	129

Notes: Dependent variable is categorized second messages; I told a lie, I told a truth, and No-message. The baseline is No-message. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 10. Ordered logit regression of receiver's credibility

	All treatments					Face-to-face treatment				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Face	0.68**	0.76***	0.71**	0.83***	0.77**					
	(0.28)	(0.29)	(0.29)	(0.30)	(0.31)					
Non-student	0.58	0.52	1.61**	1.81**	1.66**					
	(0.37)	(0.37)	(0.76)	(0.79)	(0.83)					
Trust	0.97***	1.28**	1.42***	1.33**	1.49***	1.04	1.25*	1.48*	1.54*	1.19
	(0.29)	(0.53)	(0.54)	(0.56)	(0.57)	(0.71)	(0.74)	(0.77)	(0.79)	(0.84)
SM_truth		0.84	0.98*	1.06*	0.97*	0.65	0.93	1.54*	1.49*	1.63*
		(0.52)	(0.54)	(0.55)	(0.57)	(0.74)	(0.79)	(0.83)	(0.83)	(0.83)
Trust*SM_truth		-0.44	-0.65	-0.50	-0.75	-0.31	-0.72	-1.41	-1.50	-1.20
		(0.62)	(0.64)	(0.65)	(0.67)	(0.89)	(0.97)	(1.03)	(1.04)	(1.08)
Female			0.19	0.25	0.16		0.02	0.17	0.23	0.26
			(0.27)	(0.28)	(0.29)		(0.42)	(0.43)	(0.44)	(0.59)
Age			-0.07	-0.07	-0.07		0.03	0.08	0.09	0.09
			(0.09)	(0.09)	(0.10)		(0.14)	(0.15)	(0.15)	(0.15)
Household			-0.04*	-0.05*	-0.03		-0.18	-0.24*	-0.22	-0.20
			(0.02)	(0.02)	(0.02)		(0.13)	(0.14)	(0.14)	(0.14)
GSS				0.42***	0.42***			0.83***	0.79***	0.78***
				(0.15)	(0.15)			(0.29)	(0.29)	(0.29)
TrustBehave				0.35**	0.42**			0.60**	0.64**	0.62**
				(0.16)	(0.17)			(0.29)	(0.29)	(0.30)
TrustStranger				0.41	0.32			0.33	0.37	0.40
				(0.27)	(0.28)			(0.42)	(0.44)	(0.44)
TrustWorthiness				0.24**	0.16			0.01	-0.01	0.03

				(0.12)	(0.13)			(0.17)	(0.18)	(0.18)
Re-pair2					0.84***				0.23	0.03
					(0.31)				(0.45)	(0.47)
R_Known										0.24
										(0.28)
R_Impression										0.49
										(0.30)
Pair_MM										-0.01
										(0.63)
cut1										
Constant	-2.53***	-1.94***	-2.80***	-1.77*	-2.20**	-2.93***	-6.22**	-7.09***	-6.65**	-4.28
	(0.44)	(0.56)	(0.79)	(0.98)	(1.09)	(0.86)	(2.63)	(2.75)	(2.77)	(3.07)
cut2										
Constant	-1.00***	-0.40	-1.27*	-0.08	0.08	-1.46**	-4.75*	-5.55**	-5.11*	-2.70
	(0.30)	(0.47)	(0.72)	(0.92)	(0.97)	(0.61)	(2.55)	(2.68)	(2.70)	(3.02)
cut3										
Constant	0.02	0.64	-0.22	1.04	1.35	-0.28	-3.56	-4.22	-3.79	-1.35
	(0.28)	(0.47)	(0.72)	(0.92)	(0.97)	(0.57)	(2.53)	(2.65)	(2.68)	(3.01)
cut4										
Constant	2.77***	3.42***	2.60***	4.14***	4.54***	2.37***	-0.87	-1.15	-0.67	1.86
	(0.35)	(0.52)	(0.74)	(0.96)	(1.02)	(0.63)	(2.50)	(2.62)	(2.65)	(3.02)
Pseudo R2	0.04	0.04	0.05	0.09	0.11	0.02	0.03	0.10	0.10	0.12
No. of Obs.	229	229	228	227	217	93	93	93	92	92

Notes: Dependent variable is the credibility levels measured by the five-point scales from 1: Absolutely believe to 5: Never believe. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

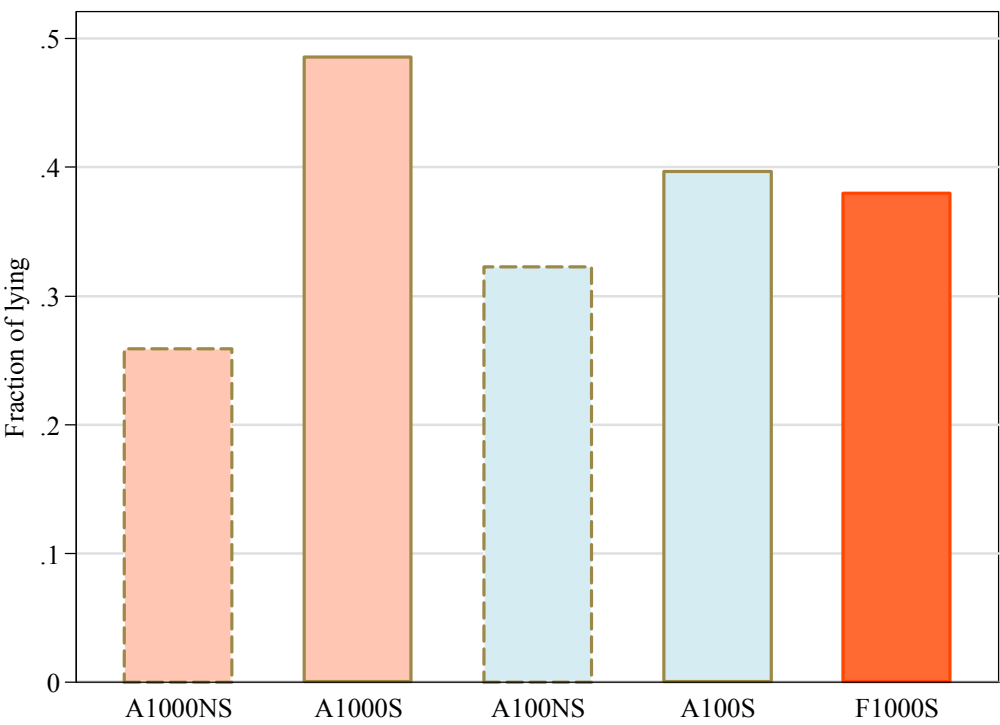Figure 1. Fraction of lies in all treatments
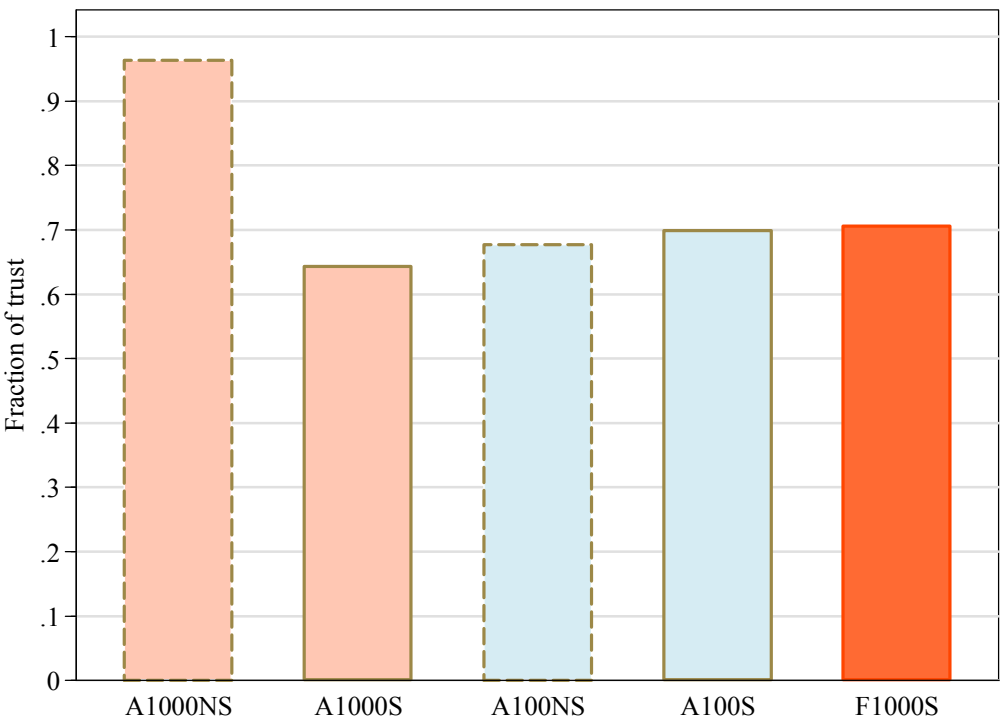

Figure 2. Fraction of trust for sender's message

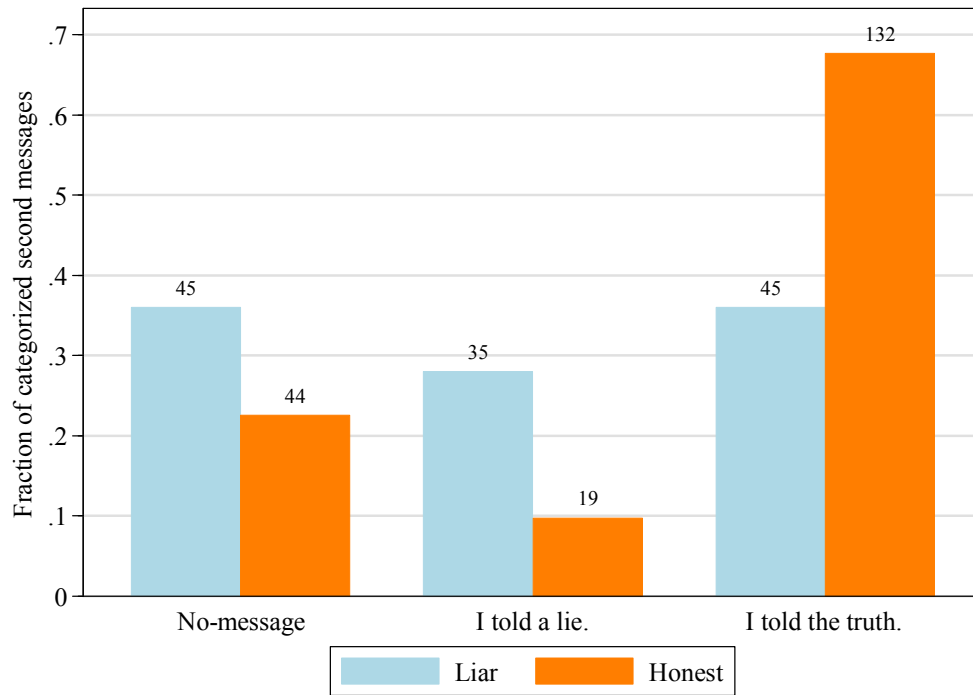Figure 3A. Distribution of categorized second messages

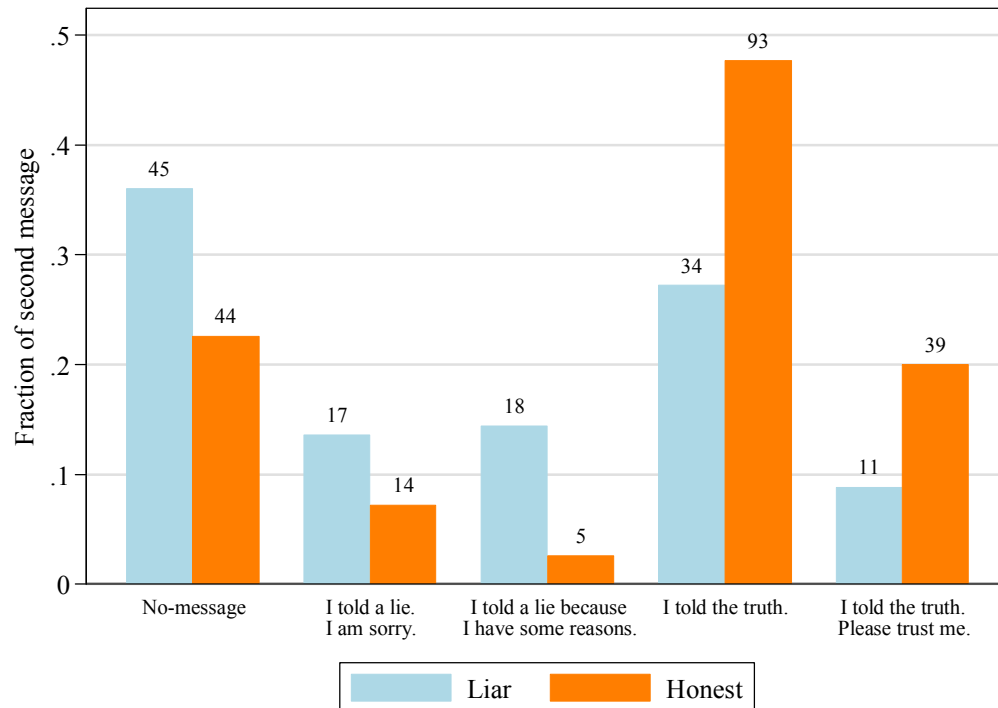Figure 3B. Distribution of all second messages
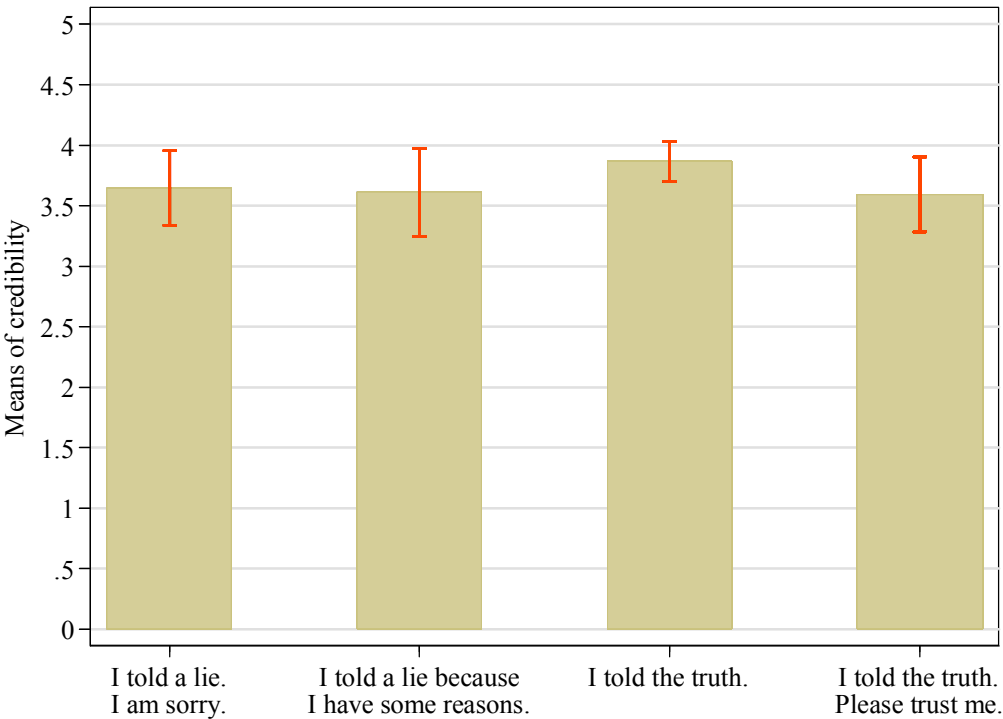
Figure 4A. Credibility levels for all messages.

Figure 4B. Credibility levels in anonymity vs. face-to-face treatments