Doshisha University

The Effect of Escalating Lies on Business Ethics: An Experimental Study of the Repeated Deception Game

Satoshi Taguchi, Kazunori Miwa, Tatsushi Yamamoto

ITEC Working Paper
22-02
March 2022

ITEC
Institute for Technology,
Enterprise and Competitiveness

The Effect of Escalating Lies on Business Ethics:
An Experimental Study of the Repeated Deception Game

Satoshi TAGUCHI* (Doshisha University)

Kazunori MIWA (Osaka University)

Tatsushi YAMAMOTO (Doshisha University)

Abstract.

Many reports of dishonest acts such as financial fraud describe how minor dishonest decisions gradually snowballed into significant ones over time. Despite the impact of these acts, we do not have a clear understanding of how and why small transgressions may gradually lead to larger ones. We developed hypotheses founded on the lying aversion hypothesis—which states that people tend to avoid lying, particularly when others are likely to experience serious damage from the deception—and the lying escalation hypothesis based on the moral disengagement theory, in which people tell a small lie initially and escalate lying behavior subsequently, even if such escalation would eventually cause serious damage to others.

We tested the hypotheses by laboratory experiments and the main results supported both. Additionally, the results of the follow-up experiment revealed the robustness of the lying escalation hypothesis with repetitive situations. This paper provides new evidence for the vast literature on lying behavior and its psychological mechanism.

Keywords:

lying aversion; lying escalation; harmful lie; moral disengagement; moral cleansing

I. Introduction

Lying behavior is a major concern in business ethics. Many accounts of dishonest acts, such as financial fraud, describe how minor dishonest decisions gradually snowballed into significant ones over time (e.g., Fleming and Zyglidopoulos 2008; McLean and Elkind 2013). Despite the impact of these acts, we do not have a clear understanding of how and why small transgressions can gradually lead to larger ones. Therefore, verifying the mechanism of such snowballing lying behavior is a pressing issue.

There are two streams of research on lying behavior. One stream argues that people usually avoid telling a lie, as the proverb says, "Honesty is the best policy." Gneezy (2005), a seminal paper on lying research, shows that people tend to avoid lying, particularly when others are likely to experience serious damage from the deception. Erat and Gneezy (2012) and Gino et al. (2013) support the lying aversion proposal of Gneezy (2005)[1] and argue that people might feel guilty about their dishonest behavior when others cannot benefit from such behavior.

Another stream of research argues that people continue to lie for their own benefit. A famous English poet in the 17th century, George Herbert, said "Show me a liar, and I will show you a thief."[2] This means that a small lie grows bigger like a snowball and eventually turns into a downright lie. Fleming and Zyglidopoulos (2008) develop a model that explains the escalation of deception in corrupt firms. This is a typical consequence of lying escalation.[3] A theory that might be related to people's lying escalation behavior is moral disengagement. Moral disengagement was originally introduced by Bandura (1986), and is defined as "a set of eight cognitive mechanisms that decouple one's internal moral standards from one's actions, facilitating engaging in unethical behavior (Moore 2015, p. 199)."[4] Shu et al. (2011) conduct four

[1] The lying aversion in this paper is based on consequences, that is, changes in wealth resulting from a lie (Gneezy 2005). In contrast, López-Pérez and Spiegelman (2013) investigate the relevance of pure lying aversion, that is, a dislike for lies independent of their consequences. This paper focuses only on the consequence-dependent lying aversion of Gneezy (2005).

[2] See Pickering (1846), p. 325.

[3] The escalation behavior of commitment to bad, failing, or unethical courses of action other than lying has been extensively examined in the literature (e.g., Brockner 1992; Kramer and Maas 2020; Staw 1976, 1981; Staw and Ross 1989).

[4] The eight mechanisms include distortions of consequences, diffusion of responsibility, advantageous comparison, displacement of responsibility, moral justification, euphemistic labeling, dehumanization, and attribution of blame. See Moore (2015) for a short review on moral disengagement.

experiments and find that one’s moral disengagement level increases after behaving dishonestly. This suggests that people repeat and even escalate the dishonest behavior because the increased moral disengagement makes their internal moral standards ineffective and can justify their dishonest behavior.

Which condition holds true in the real world, lying aversion or lying escalation? In this study, we assume that both lying aversion and lying escalation hold true. On one hand, people are assumed to avoid telling lies, particularly when others are likely to experience serious damage from the deception without the recurrent circumstance. On the other hand, if people have sufficient opportunities to tell lies under multiple circumstances, they are more likely to tell a small lie initially and to escalate lying behavior subsequently, even if such escalation would eventually cause serious damage to others.

We test the two hypotheses, lying aversion and lying escalation, by conducting two experiments in which a total of 176 participants participated. An experimental approach enables us to create a hypothetical environment and to control all other factors that might affect lying behavior. Specifically, our experiments are based on the deception game of Gneezy (2005) and Erat and Gneezy (2012), and we extend the game to repeated settings.

The main results are as follows. First, we observe the tendency of lying aversion. That is, participants tend to avoid lying, particularly when others are likely to experience serious damage from the deception. Second, however, if the payoff structures gradually change with time from a harmless lie to a harmful lie, in the sense that the harmless lie does not harm the partner but the harmful lie does serious damage to the partner, then participants tend to continue lying to the end. This supports the lying escalation hypothesis. Additionally, the results of the follow-up experiment reveal the robustness of the lying escalation hypothesis with repetitive situations in which the self-payoff increases but the others-payoff decreases.

The contributions of our findings are two-fold. First, we provide evidence that corroborates not only the lying aversion hypothesis of Gneezy (2005) but also our lying escalation hypothesis. Gneezy (2005) claims that people tend to avoid lying, particularly when it would cause serious damage to others. This might be explained by the fact that people feel remorse over what they have done. In addition, we clarify that people tend to report a harmless lie initially, escalate lying, and eventually report a serious harmful lie, even if it would cause serious damage to others. We further explain the phenomenon by the psychological cost of lying and the decision-making time of the sender. Therefore, our findings add a piece of new evidence to the vast literature on lying behavior.

Second, we contribute to the two streams on the psychological mechanism of lying behavior: the *moral disengagement theory* (Bandura 1986; Moore 2015; Rosenbaum et al. 2014; Shu et al. 2011; Vincent et al. 2013) and the *moral cleansing theory* (Blanken et al. 2015; Chowdhury et al. 2021; Cojoc and Stoian 2014; Gneezy et al. 2014; Lasarov and Hoffmann 2020; Ploner and Regner 2013; Schurr and Ritov 2016; West and Zhong 2015). The results of our experiments partially demonstrate the moral cleansing behavior of the sender but generally support the moral disengagement theory and the lying escalation behavior. Therefore, our findings reveal robust evidence regarding the moral disengagement theory.

II. Model and Hypotheses

II-1. Deception Game

Our experiments are based on the deception game of Gneezy (2005) and Erat and Gneezy (2012). In particular, we extend the deception game of Erat and Gneezy (2012) to a repetitive setting. In the deception game, two players act sequentially in the role of sender and receiver, respectively. Figure 1 shows the timeline of the game.

[Insert Figure 1 about here.]

A computer rolls a six-sided die before the start of the game, and only the sender knows the outcome. The sender chooses a message to the receiver from a pool of six possible messages which are “the outcome of the roll of the die was *i*,” where i $\in \{1, 2, 3, 4, 5, 6\}$. After observing a message from the sender, the receiver chooses a number from the set $\{1, 2, 3, 4, 5, 6\}$. If the receiver chooses the real outcome of the die roll, payoff option A is implemented, and for any other choice, payoff option B is implemented.

To test our hypotheses shown in the next subsection, we set two types of payoff for options A and B (Table 1).

[Insert Table 1 about here.]

Under the first type, the payoff for option A was (300, 300), and that for option B was (500, 300). We called this a "harmless lie," in which the receiver would suffer no damage on being deceived. Under the second type, the payoff for option A was (300, 300), and that for option B was (500, *100*). We called this a "harmful lie," in which the receiver would suffer serious damage on being deceived. Importantly, the receiver was not informed what the actual payoffs associated with option A and option B were, and the sender knew that the receiver would not know the payoffs.

II-2. Hypotheses Development

There are two streams of research on lying behavior. One stream argues that people generally avoid telling a lie. Using an experimental approach, Gneezy (2005) shows that people tend to avoid lying, particularly when others are likely to experience serious damage from the deception. Erat and Gneezy (2012) classify lies into four categories—selfish black lies, spiteful black lies, altruistic white lies, and Pareto white lies—depending on how a sender's lie affects both the sender's and receiver's payoffs. While selfish black lies increase the sender's payoff and decrease the receiver's payoff, spiteful black lies decrease both the payoffs. Altruistic white lies decrease the sender's payoff and increase the receiver's payoff, and Pareto white lies increase both the payoffs. Erat and Gneezy (2012) suggest that people are more likely to tell "Pareto white lies" than "selfish black lies." Gino et al. (2013) argue that people might feel less guilty about their dishonest behavior when others can benefit from it. These papers reinforce the conclusions of Gneezy (2005). Hence, we first verify the lying aversion hypothesis of Gneezy (2005).

Hypothesis 1 (H1): Lying aversion hypothesis. People tend to avoid lying, particularly when others are likely to experience serious damage from the deception.

Based on hypothesis 1 (lying aversion hypothesis), the difference between the levels of the fraction of lies for the harmless lie and those for the harmful lie would be statistically significant.

Another stream of research argues that people continue lying for their own benefit. Many reports of dishonest acts such as financial fraud describe how minor dishonest decisions gradually snowballed into significant ones over time (e.g. McLean and Elkind 2013). Fleming and Zyglidopoulos (2008) develop a model that explains the escalation of deception in corrupt firms. They claim that if undetected, an initial lie can begin a process whereby the ease, severity, and pervasiveness of deception increases over time so that it eventually becomes an organization-level phenomenon. Welsh et al. (2015) experimentally show that individuals engage in a slippery slope of increasingly unethical behavior in response to escalating rewards. Garrett et al. (2016) also strived to empirically demonstrate escalation of dishonesty in a controlled laboratory setting and examine the underlying mechanism. Their results provide empirical evidence that dishonesty gradually increases with repetition when all else is held constant, and also offer a mechanistic account of how dishonesty escalates, showing that it is supported by reduced activity in brain regions previously associated with emotion, predominantly the amygdala. Lee et al. (2019) experimentally unpack that dishonest conduct reduces one's generalized empathic accuracy—the ability to accurately read other people's emotional states.

A psychological theory behind lying escalation behavior is *moral disengagement* (Bandura 1986; Moore 2015).[5] In general, people have their internal moral standards and, hence, they tend not to behave in a way that violates their moral standards. *Moral disengagement* is a set of cognitive mechanisms that makes people's self-regulation for ethical behavior inactive. Shu et al. (2011) show that dishonest behavior itself increases the level of moral disengagement. Because people have internal moral standards and care about

[5] Cojoc and Stoian (2014) and Jacobsen et al. (2018) also claim a similar psychological mechanism called *conscience numbing* in which an initial transgression numbs one's conscience, decreases a sense of guilt, and leads to another transgression.

behaving ethically, they feel distress arising from cognitive dissonance after behaving dishonestly. The increased moral disengagement can serve to reduce such cognitive dissonance. Furthermore, several studies show that moral disengagement predicts future unethical behavior (e.g., Rosenbaum et al. 2014; Vincent et al. 2013). This suggests that dishonest behavior increases the level of moral disengagement, and the increased moral disengagement makes it easy to behave dishonestly in the future. Taken together, previous research on moral disengagement predicts that people tend to repeat and even escalate dishonest behavior such as lying escalation.

Hypothesis 2 (H2): Lying escalation hypothesis. The sender is likely to escalate lying behavior, even if lying messages would be gradually harmful and eventually cause serious damage to others.

Specifically, we test the above hypotheses using three indicators: *the levels of the fraction of lies*, *the psychological cost of lying of the sender*, and *the decision-making time of the sender.*

H2a: The levels of the fraction of lies. Based on H2 (lying escalation hypothesis), when the lies gradually escalate from harmless to harmful with repetition, the sender would continue to report the lying messages and there would be no significant difference between the levels of the fraction of lies under the harmless setting [(1) in Table 1] and those under the harmful setting [(2) in Table 1]. However, based on H1 (lying aversion hypothesis), the difference between the levels of the fraction of lies under the harmless setting and those under the harmful setting would significantly differ.

Next, we focus on *the psychological cost of lying of the sender*. We assume that the greater the lie distance, the difference between the true value and the message value of the lie as described below, the greater the psychological cost of lying to the sender (e.g., Abeler et al. 2014, 2019; Dufwenberg and Dufwenberg 2018; Gibson et al. 2013; Gneezy et al. 2018; Kartik 2009; Lundquist et al. 2009; Rosenbaum et al. 2014). Gneezy et al. (2018) show, for example, that the intrinsic cost of lying depends on the size of the lie, which is measured by the outcome dimension (the distance between what the agent observes and what she/he says).[6] We define *the lie distance ratio* as the rate obtained by dividing the lie distance by *the maximum lie distance*.

$$The\ lie\ distance\ ratio = \frac{The\ lie\ distance}{The\ maximum\ lie\ distance}.$$

Figure 2 shows the outline of the definition of the lie distance ratio.

[Insert Figure 2 about here]

First, we define *the lie distance* of the sender as the absolute value of the difference between the true value and the message value that the sender sent. For example, when the true value is equal to “1” and the sender sent the message “4,” the level of the lie distance is equal to “3.”

The maximum lie distance depends on the true value: when the true value is equal to “1,” for example, the maximum lie distance is equal to “5.” When the true value is equal to “2,” “3,” “4,” “5,” and “6,” the maximum lie distance is equal to “4,” “3,” “3,” “4,” and “5,” respectively. In the case of Figure 2, we calculate the ratio to be 0.6. The greater the lie distance ratio, the greater the psychological cost of lying of the sender.

H2b. The psychological cost of lying of the sender. Based on H2 (lying escalation hypothesis), when the lies gradually escalate from harmless to harmful with repetition, the sender would gradually choose a message with a lower lie distance ratio to reduce the psychological cost of lying.

Finally, we focus on *the decision-making time* of the sender, which we define as the time between the sender’s observation of the true value and the sender’s report of the message. Related works of literature

[6] Lundquist et al. (2009) also define the size of the lie as the deviation from the true number.

discuss whether lying behavior is associated with system 1 or 2 of the brain: intuitive or deliberation behavior, respectively (Capraro 2017; Foerster et al. 2013;Lohse et al. 2018; Shalvi et al. 2012). Garrett et al. (2016) show, for example, that lying behavior is supported by reduced activity in brain regions previously associated with emotion, predominantly the amygdala. We propose the following hypothesis:

H2c. Decision-making time of the sender. Based on H2 (lying escalation hypothesis), when the lies gradually escalate from harmless to harmful with repetition, as the lies gradually escalate, the sender intuitively lies and the decision time gradually becomes shortened.

III. Experimental Design

The first step in the empirical exploration of the escalation of lying is examining the effect of the number of rounds. To test our hypotheses, we conduct the deception game experiment described in the previous section and extend the game to repeated settings. In this initial study, we adopt a 2 × 1 experimental design, where *the number of rounds* is manipulated between participants at two levels: five rounds (*escalation condition*) and two rounds (*no escalation condition*). For the “harmless lie” and “harmful lie,” we define the two conditions: the escalation and no escalation conditions, respectively (Figure 3).

[Insert Figure 3 about here.]

Under the escalation condition, a sender has numerous opportunities to tell lies because there are five rounds—the payoff for the first round is a “harmless lie” and that for the final round is a “harmful lie.” The payoff structures gradually change with time from a harmless lie to a harmful lie, in the sense that the harmless lie does not harm the receiver but the harmful lie causes the receiver serious damage. As the number of rounds increases, the degree to which the receiver is hurt by the lie will gradually increase. However, under the no escalation condition, a sender has few opportunities to tell lies because there are only two rounds, although the payoff structures of the first and final rounds are the same as that under the escalation condition.

III-1. Participants

We recruited participants using a standard research participant pool of undergraduate and graduate students at a large private university by advertisements and e-mail.[7] In total, 88 undergraduate and graduate students participated in our experiment.[8] Participants were 20.61 years old on average (SD = 1.31). The maximum and minimum ages of the participants were 24 and 18 years, respectively, and 48.86% of the participants were female. Monetary rewards were emphasized as an incentive for participation. The allocation of participants to the conditions was completely random; the number of participants was 44 each under the escalation and no escalation conditions. Because we adopted a between-participant design, no one participated in more than one experimental session.

III-2. Procedures

The experiment was programmed with z-Tree software and administered in an experimental laboratory of a large private university (Fischbacher 2007).[9] We conducted nine sessions (four and five sessions for the escalation and no escalation conditions, respectively) of computerized experiments in October 2017 and June and November 2018.

Each participant participated in only one session, comprising five or two rounds of decision-making. Participants were assigned the role of a sender or a receiver, which was predetermined randomly by

[7] The use of students as surrogates for employed adults and professionals has long been an issue in business research (Dickhaut et al. 1972). However, several studies have suggested that business students are appropriate proxies for professionals when assessing basic traits or perceptions (Ward 1993). Remus (1986) and Greenberg (1987) address the student-as-surrogates issue by studying business students and employed adults simultaneously, and both researchers conclude that the results show no differences. Geiger and Smith (2010) also argue that the use of business students as surrogates for employed professionals is appropriate.

[8] A total of 98 unique participants took part in eight sessions. A computer failure occurred in one session, requiring us to run a replacement session. Thus, our final sample comprised 88 individuals who participated in seven sessions.

[9] All experiments in this study were approved by the IRB of the university in which the experiment was conducted.

the computer at the beginning of the experiment. Roles remained unchanged throughout the rounds. Under each condition, half of the participants were assigned the role of a sender, and the other half that of a receiver.[10] To minimize the influences of role-playing (Haynes and Kachelmeier 1998), the two roles were labeled "role A" for the sender and "role B" for the receiver in the experimental instructions (see Supplement 1). Participants interacted anonymously through a computer network using the z-Tree software.

We informed all participants that the partner would be determined randomly by the computer at the beginning of each round and that the partner would not change throughout an experiment.

The participants were separated by dividers in each experimental session. At the beginning of each session, participants read an initial set of instructions (see Supplement 1). We used the instructions based on Erat and Gneezy (2012). Therefore, the structure of the game was explained to the participants using neutral terminology in order to increase experimental control and reduce the risk of undesired contextually induced incentives (e.g., Arnold 2015; Friedman and Sunder 1994; Moser 1998). After the instructions were read, participants were asked to answer questions about the experiment. Participants had to answer all questions correctly before they began the experimental task. Hence, we ensured that all participants accurately understood the details of the experiment.

The feedback information at the end of each round for a sender and a receiver differed. For a sender, it was as follows: the number the computer chose randomly, his/her own message, receiver's decision, and his/her own payoff. For a receiver, it was as follows: partner's message and his/her own action. There was less information provided to the receiver, and both the sender and the receiver were aware of the asymmetry of the feedback information. This setting was based on previous research (Erat and Gneezy 2012). In all of the treatments, participants were not provided with any information, either individually or in aggregate, about the results of the other pairs of participants.

At the end of the experiment, participants filled out an exit questionnaire that gathered demographic information and personal perceptions (see Supplements 2 and 4). Each session lasted about 60

[10] In total, under both escalation and no escalation conditions, 22 participants each were assigned the role of sender and receiver.

minutes (escalation condition) or 40 minutes (no escalation condition), including the reading of the instructions and answering the exit questionnaire. Participants received a $9 show-up fee plus their rewards from the game in cash.[11] The average earnings were $17.06 ($20.50 for the escalation condition and $13.62 for the no escalation condition). Supplement 6 summarizes the experimental design.

IV. Results

Panels A, B, and C of Table 2 present the descriptive statistics for the fraction of lies, the lie distance ratio, and the decision time by each condition, respectively, for each experimental condition.[12]

[Insert Table 2 about here]

We first focus on the senders' lying behavior. The fraction of lies under the no escalation condition was higher in the first (86.36%) than in the last round (59.09%) (Table 2 Panel A). There were also significant differences in the fraction levels between the first and last rounds under the no escalation condition—Fisher's Exact Test (one-tailed) indicated that $p = 0.044$ and *effect size* (ϕ) = 0.3610.[13] This result provides evidence consistent with H1, suggesting that our findings are in agreement with those of previous studies, such as Gneezy (2005).

There were no significant differences in the fraction levels between the first (86.36%) and the last rounds (68.18%) under the escalation condition (Table 2 Panel A)—Fisher's Exact Test (one-tailed) $p = 0.140$ and *effect size* (ϕ) = 0.2310.[14] This result implies that under the escalation condition, in which the

[11] The method of calculating the variable rewards is follows: the sender obtained 0.6 times gain of earned points, while the receiver obtained 1.0 times gain of earned points. Incentives of this kind, which include ex-ante adjustments of various kinds to equalize mean (not individual) payoffs across roles, are often used in the experimental literature in finance and accounting (e.g., Bloomfield et al. 2005; Hales 2009; Nelson et al. 2001).

[12] Participants responded to a number of statements in an exit questionnaire, which was designed to test the effectiveness of our experimental manipulation and controls using a seven-point Likert-type scale (1 = strongly disagree and 7 = strongly agree; see Supplement 2). The tests measured the mean difference from the neutral response (all $p < 0.01$). Therefore, the manipulation and controls were effective for our experiment.

[13] Pearson's Chi-squared test also indicated that $X^2_{(1)} = 2.864$, $p = 0.090$.

[14] Pearson's Chi-squared test also indicated that $X^2_{(1)} = 1.164$, $p = 0.280$.

payoff structures gradually changed with repetition from a harmless to a harmful lie, participants tended to continue lying to the end. In other words, the lying escalation effect was stronger than the lying aversion effect. This result supports H2a. The above results are robust because the same result was derived even when a probit regression analysis was used to control for individual factors (e.g., gender, morality, and lie acceptability) (see Supplements 3 and 4).

IV-1 Psychological Cost of the Sender Lying: Analysis of the Lie Distance Ratio

Next, we focus on analysis of the lying cost of the sender. Figure 4 shows the box plot of the lie distance ratio by round and each condition and supplement 7 shows the statistical test of differences of the lie distance ratio for each condition.

[Insert Figure 4 about here]

Only for the escalation condition did the lie distance ratio decreased from the first to the last round (Figure 4); the median levels significantly ($p < 0.05$) decreased from the first (0.50) to the last round (0.25)—Mann–Whitney U test (one-tailed) $p = 0.032$, *effect size* $r = 0.491$ (supplement 7). Additionally, under the no escalation condition, the difference between the median levels of the lie distance ratio in the first (0.60) and the last round (0.67) did not significantly differ—Mann–Whitney U test (one-tailed) $p = 0.622$, *effect size* $r = 0.113$) (supplement 7).

In sum, only in the escalation condition did the lie distance ratio significantly differ between the first and last rounds. The result supports H2b. Under the escalation condition, as the lies gradually escalated from harmless to harmful lies, the sender gradually chose a message with a lower psychological cost of lying.

IV-2 Analysis of Decision Time

We now focus on analysis of senders' decision time. Panel C of Table 2 shows the descriptive statistics of the decision time of the senders for each condition and supplement 8 shows the corresponding statistical test of the differences in decision time.

Under the escalation condition, the median levels of decision time significantly ($p < 0.05$) decreased from the first (20.50 seconds) to the last round (18.00 seconds), with Mann–Whitney U test (one-tailed) showing $p = 0.026$, *effect size* $r = 0.472$ (supplement 8). However, under the no escalation condition, although the median levels of the decision time decreased from the first (23.50 seconds) to the last round (21.50 seconds), the difference was not significant—Mann–Whitney U test (one-tailed) $p = 0.306$, *effect size* $r = 0.218$ (supplement 8). In sum, only in the escalation condition was the difference between the decision time in the first and the last round significant. The result supports H2c. Under the escalation condition, as the lies gradually escalated, the senders intuitively lied and the decision time was gradually shortened.

V. Follow-up Experiment

The results of the experiment described in the previous section supported both H1 (lying aversion) and H2 (lying escalation). In particular, under the escalation condition, the lying escalation effect was stronger than the lying aversion effect with the repeated setting.

We conducted a follow-up experiment to examine whether the lying escalation effect could be reproduced in another payoff structure of the deception game with repetition, especially when self-servicing dishonesty was amplified. Welsh et al. (2015) experimentally show, for example, that individuals engage in a slippery slope of increasingly unethical behavior in response to escalating self-rewards. In our follow-up experiment, we combined the setting of Welsh et al. (2015) with the setting of the others-payoff reduction employed in our main experiment.

[Insert Figure 5 and Table 3 about here]

In the follow-up experiment, we set different payoffs associated with the three types of lies and options (Table 3). Under the first type [see (1) in Table 3], the payoff for option A was (4,000, 4,000) and that for option B was (5,000, *3,000*). We call this a "*Little selfish and harmless lie*," in which the sender would obtain little additional payoff and the receiver would suffer little damage on being deceived. Under the second type [see (2) in Table 3], the payoff for option A was (4,000, 4,000) and that of option B was (*7,000*, *3,000*). We call this a "*Selfish and harmless lie*," in which the sender would obtain a high payoff but the receiver would suffer little damage on being deceived. Under the third type [see (3) in Table 3], the payoff for option A was (4,000, 4,000) and that of option B was (*7,000*, *1,000*). We call this a "*Selfish and harmful lie*," in which the sender would obtain a high payoff and the receiver would suffer great damage on being deceived.

Based on these types, we conducted the deception game experiment described in Figure 5. We adopted a 2 × 1 experimental design, where the self-payoff and the others-payoff were manipulated with repetition between participants at two levels. First, the payoff structure of the deception game gradually changed from (1) *Little selfish and harmless lie* to (2) *Selfish and harmless lie* with repetition, where the self-payoff increased but the others-payoff was constant. We call this treatment the **Inc-Con condition** (control condition). Second, the other payoff structure of the deception game gradually changed from (1) *Little selfish and harmless lie* to (3) *Selfish and harmful lie* with repetition, where the self-payoff increased and the others-payoff decreased. We call this treatment the **Inc-Dec condition**. The payoff of the Inc-Dec condition was a combination of the Inc-Con and the others-payoff reduction settings employed in our main experiment. Although the payoffs of both conditions gradually changed selfishly, the two trends differed in whether the others-payoff decreased.

Next, we predict the results of the follow-up experiment based on the lying escalation hypothesis. From the lying escalation hypothesis, even if the payoffs of the others gradually decrease, the fraction of lies would not decrease with repetition. However, from the lying aversion hypothesis, the Inc-Dec condition is expected to result in a significant reduction in lies due to the others-payoff reduction with repetition. Therefore, we generated the following research question:

***Research Question (the lying escalation hypothesis in the follow-up experiment)*:** Would the transition of the fraction of lies under the Inc-Dec condition, where the self-payoff increases and the others-payoff decreases, differ from that under the Inc-Con condition, where the self-payoff increases but the others-payoff is constant?

V-1. Experimental Design

We recruited participants using a standard research participant pool of undergraduate and graduate students at a large private university by advertisements and e-mail. In total, 88 undergraduate and graduate students participated in our experiment. Participants were 19.55 years old on average (SD = 1.19). The maximum and minimum ages of the participants were 26 and 18 years, respectively, and 54.55% of the participants were female. Monetary rewards were emphasized as an incentive for participation.

The experiment was programmed with z-Tree software and administered in an experimental laboratory of a large private university (Fischbacher 2007). We conducted three sessions of computerized experiments in April 2019. Each participant participated in only one session, comprising five rounds of decision-making. The allocation of participants to the conditions was completely random. Because we adopted a between-participant design, no one participated in more than one experimental session. Participants were assigned the role of a sender or a receiver, predetermined randomly by the computer at the beginning of the experiment. Roles remained unchanged throughout the rounds. Under each condition, half of the participants were assigned the role of a sender and the other half that of a receiver. The experimental procedure was the same as the main experiment described in the previous section. Each session lasted about 60 minutes, including the reading of the instructions and answering the exit questionnaire. Participants received a $9 show-up fee plus their earnings from the game in cash.[15] The average earnings were $22.27

[15] The way of calculating the variable earnings follows: the sender obtained 0.06 times gain of earned points, while the receiver obtained 0.09 times gain of earned points. Incentives of this kind, which include ex-ante adjustments of various kinds to equalize mean (not individual) payoffs across roles, are often used in the experimental literature in finance and accounting (e.g., Bloomfield et al. 2005; Hales 2009; Nelson et al. 2001).

($23.04 under the Inc-Con condition and $21.49 under the Inc-Dec condition). Supplement 9 summarizes the experimental design.

V-2. Result

Panels A–D of Table 4 present the descriptive statistics for the fraction of lies, the lie distance ratio, the decision time (full sample), and the decision time (sub-sample) for each experimental condition, respectively.

[Insert Table 4 about here]

We now focus on the senders' lying behavior. Under the Inc-Con condition, the fraction of lies slightly decreased from the first (68.18%) to the last round (63.64%) (Table 4 Panel A). There were no significant differences in the fraction levels between the first and the last round under the Inc-Con condition—Fisher's Exact Test (one-tailed) $p = 0.500$. Under the Inc-Dec condition also, there were no significant differences in the fraction levels between the first (59.09%) and the last round (59.09%)—Fisher's Exact Test (one-tailed) $p = 0.62$. This result implies that under the Inc-Dec condition, in which the payoff structures gradually changed from little selfish and harmless lie to selfish and harmful lie with repetition, the sender tended to continue lying to the end at the same level as for the Inc-Con condition. In other words, the lying escalation effect was stronger than the lying aversion effect with repetition.

In sum, the follow-up experiment revealed the robustness of the lying escalation hypothesis with repetitive situations in which the self-payoff increased but the others-payoff decreased.

V-2-1. Psychological Cost of the Sender Lying: Analysis of the Lie Distance Ratio

Next, we focus on the analysis of the lie distance ratio. Panel B of Table 4 shows the descriptive statistics of the lie distance ratio of the sub-sample restricted to the sender reporting a lie under each condition. Figure 6 shows the box plot of the lie distance ratio by round and each condition, and supplement 10 shows the corresponding statistical test of differences.

[Insert Figure 6 about here]

Under both conditions, the lie distance ratio increased from the first to the last round (Figure 6). Under the Inc-Con condition, the median levels of the lie distance ratio increased from the first (0.40) to the last round (0.67), and this difference was significant at $p < 0.01$—Mann–Whitney U test (one-tailed) $p = 0.000$, *effect size* $r = 1.076$ (supplement 10). Under the Inc-Dec condition, the difference between the median levels of the lie distance ratio in the first (0.67) and the last round (0.75) was also significant at $p < 0.01$—Mann–Whitney U test (one-tailed) $p = 0.000$, *effect size* $r = 1.065$ (supplement 10).

In sum, under both conditions, the lie distance ratio significantly increased with repetition. For the psychological cost of lying of the sender (the lie distance ratio), the result of the follow-up experiment was opposite to the result of the main experiment. This was due to the difference in the self-payoff structure: under the follow-up experiment, the self-payoff when the receiver was deceived gradually increased with repetition. Under both the Inc-Con and Inc-Dec conditions, as the lies gradually escalated from little selfish to selfish, the sender gradually chose a message with a higher psychological cost to ensure deceiving the receiver.

V-2-2. Analysis of Decision Time

We now focus on the analysis of senders' decision time. Panels C and D of Table 4 show the descriptive statistics of the decision time of senders for each condition. Panel C presents the result of the full sample; Panel D presents that of the sub-sample restricted to the sender reporting a lie. Supplement 11 shows the statistical test of differences in the decision time of the sender for each condition.

For the sub-sample restricted to the sender reporting a lie, only in the Inc-Dec condition was the difference between decision time in the first and the last round significant at $p < 0.05$—Mann–Whitney U test (one-tailed) $p = 0.025$, *effect size* $r = 0.476$) (Supplement 11 Panel B).[16]

For the decision time of the sender with the sub-sample restricted to the sender reporting a lie, the result of the follow-up experiment was the same as the result of the main experiment. Under the Inc-Dec condition, as the lies gradually escalated, the sender intuitively lied and the decision time was gradually shortened.

VI. Discussion: Possible Alternative Psychological Mechanism

Both experiments with different payoff structures supported the lying escalation hypothesis. Therefore, our results indicate that the psychological mechanism of moral disengagement is robust. This section describes the relationship between our experimental results and an alternative psychological mechanism: *moral cleansing theory*. Cojoc and Stoian (2014) point out that not only the moral disengagement theory but also the moral cleansing theory may be psychological mechanisms that can explain dishonest behavior. The moral cleansing theory is described as follows: ethical decisions are "substitutes" and past transgressions enhance one's conscience, leading to a desire to atone through compliance with social norms, and vice versa (e.g. Blanken et al. 2015; Chowdhury et al. 2021; Cojoc and Stoian 2014; Gneezy et al. 2014; Lasarov and Hoffmann 2020; Ploner and Regner 2013; Schurr and Ritov 2016; West and Zhong 2015). The moral cleansing hypothesis, therefore, predicts in our experiments that the sender deliberately chooses to be inconsistent with repetition: honest behavior in the first round may lead to a dishonest behavior later, and vice versa. Therefore, such an inconsistent or compensatory pattern of behavior may be observed with repetition.

Figure 7 shows the transition of the sender's behavior by round and each condition.

[Insert Figure 7 about here]

[16] However, in the Inc-Con condition, the difference between the decision times in the first and last rounds was not significant [Mann–Whitney U test (one-tailed) indicated that $p = 0.162$, *effect size* $r = 0.297$].

Panel A of Figure 7 shows the fraction of lies, and Panel B shows the mean levels of the lie distance ratio, by round and each condition. They show, as predicted by the moral cleansing theory, an inconsistent or compensatory pattern of behavior of the sender with repetition.

To further analyze the possibility of the moral cleansing theory in our experiments, we focus on the individual behavior of the sender. Table 5 posits an analysis of *the decision switching ratio of the sender* by each condition, defined as follows:

$$The\ decision\ switching\ ratio = \frac{The\ number\ of\ switching}{The\ number\ of\ all\ rounds - 1}.$$

[Insert Table 5 about here]

This ratio indicates the rate of the decision switching for all opportunities and implies the degree of moral cleansing: the higher the ratio, the greater the degree of moral cleansing of the sender. Panel A of Table 5 shows that the mean and median levels of the ratio, and Panels B and C show the frequency of decision switching (the number and even–odd classification, respectively).

An even number of switches indicates that the same decision is made at the end; however, an odd number of switches indicates a different decision. For example, in an odd-numbered case, if a sender first reports a lie and then changes the decision once, the decision will eventually be changed and an honest report will be made in the end (lie→honest). In contrast, in an even-numbered example, if a sender first reports a lie and then changes the decision twice, the decision will be maintained and a lying report will be made in the end (lie→honest→lie).

Notably, the switching frequency was often an even number under the escalation, Inc-Con, and Inc-Dec conditions (Table 5 Panel C), indicating that the same decision was frequently made at the end.

Therefore, the experimental results partially demonstrate moral cleansing behavior of the sender[17] but generally support the moral disengagement theory and lying escalation behavior.

VII. Conclusion

A multitude of reports concerning dishonest acts, such as financial fraud, describe how minor dishonest decisions gradually snowballed into significant ones over time. Despite the impact of these acts, we do not have a clear understanding of how and why small transgressions may gradually lead to larger ones. We developed hypotheses founded on the lying aversion hypothesis (Gneezy 2005) and the lying escalation hypothesis based on the moral disengagement theory (Bandura 1986; Shu et al. 2011) and tested them in laboratory experiments under multiple circumstances. The main results supported both hypotheses. Additionally, the results of the follow-up experiment revealed the robustness of the lying escalation hypothesis in multiple circumstances.

Despite offering new evidence to the vast literature of lying behavior and its psychological mechanism, this study is limited in three aspects. First, we did not consider the context in our experiments. For example, the effect of lying escalation and lying aversion may change depending on the nature of the context. For example, Bauer et al. (2020) examine the moral cleansing theory in the context of internal control over financial reporting. This issue may be analyzed in detail in future studies.

The second limitation of the study is the influence of personal characteristics. Prior research has discussed the relationship between lying behavior and personal characteristics (e.g. Cohn et al. 2014; Gibson et al. 2013). Although a growing body of evidence supports heterogeneity in both social preferences and lying behavior, our knowledge concerning the relationship between these two propensities at the individual level is scarce. Kerschbamer et al. (2019) find, for example, that altruists lie less when lying hurts another party but they find no evidence supporting the hypothesis that altruists are more (or less) averse to lying than

[17] In our experiments, because the payoff structure changed through rounds, it is considered that the change affected the decision making of the sender. For this reason, it may be difficult to say that switching decisions was due to the effect of moral cleansing.

others in environments where lying has no effects on the payoffs of others. We discussed the relationship between personal characteristics and lying behavior in Supplement 3, but a more in-depth examination may be necessary.

The third limitation of the study is how to prevent people from lying. For example, Pennycook et al. (2020) indicate the importance of reconsideration through experiments that examine whether a nudge of "room for reconsideration" prevents the spread of COVID-19 fake news. Saijo (2019) and Masuda et al. (2014) also show that, under a social dilemma game, "second thoughts" is human wisdom that plays an important role in resolving social dilemmas. The results of Saijo (2019) and Masuda et al. (2014) indicate that second thoughts change the payoff structure of the game in favor of cooperation and the second-thought-mechanism is robust even when players deviate from a payoff maximizing behavior. However, it should also be noted that nudges against lying may not work. For example, Dimant et al. (2020) examine framing effects in nudging honesty, and find compelling null effects with tight confidence intervals, showing that none of the norm-nudge interventions work. This issue may be analyzed in detail in future studies.

References

Abeler, J., Becker, A., Falk, A. (2014). Representative evidence on lying costs. Journal of Public Economics. 113, 96–104. https://doi.org/10.1016/j.jpubeco.2014.01.005

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. Econometrica. 87(4): 1115–1153. https://doi.org/10.3982/ECTA14673

Arnold, M.C. (2015). The effect of superiors' exogenous constraints on budget negotiations. Accounting Review. 90(1): 31–57. https://doi.org/10.2308/accr-50864

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall: Englewood Cliffs, NJ.

Bauer, T. D., Bucaro, A. C., & Estep, C. (2020). The unintended consequences of material weakness reporting on auditors' acceptance of aggressive client reporting. Accounting Review. 95 (4): 51–72. https://doi.org/10.2308/accr-52610

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. Personality and Social Psychology Bulletin. 41(4):540–558. https://doi.org/10.1177/0146167215572134

Bloomfield, R., O'hara, M., & Saar, G.(2005). The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. Journal of Financial Economics.75(1): 165–199. https://doi.org/10.1016/j.jfineco.2004.07.001

Brockner, J. (1992). The escalation of commitment to a failing course of action: Toward theoretical progress. Academy of Management Review. 17(1): 39–61. https://doi.org/10.5465/amr.1992.4279568

Capraro, V. (2017). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. Economic Letters.158: 54–57. https://doi.org/10.1016/j.econlet.2017.06.015

Chowdhury, S. M., Kim, C., & Kim, S. H.(2021). Pre-planning and its effects on repeated dishonest behavior: An experiment. Bulletin of Economic Research.73(2): 143-153. https://doi.org/10.1111/boer.12238

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. Nature. 516: 86–89. https://doi.org/10.1038/nature13977

Cojoc, D., & Stoian, A. (2014) Dishonesty and charitable behavior. Experimental Economics.17(4): 717–732. https://doi.org/10.1007/s10683-014-9391-2

Dickhaut, J. W., Livingstone, J., & Watson, D. (1972). On the use of surrogates in behavioral experimentation. Accounting Review. 47:455-470.

Dimant, E., Van Kleef, G. A., & Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. Journal of Economic Behavior and Organization.172: 247-266. https://doi.org/10.1016/j.jebo.2020.02.015

Dufwenberg, M., & Dufwenberg, M. A. (2018). Lies in disguise – A theoretical analysis of cheating. Journal of Economic Theory.175: 248-264. https://doi.org/10.1016/j.jet.2018.01.013

Erat, S., & Gneezy, U.(2012). White lies. Management Science. 58(4): 723–733. https://doi.org/10.1287/mnsc.1110.1449

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics. 10(2):171–178. https://doi.org/10.1007/s10683-006-9159-4

Fleming, P., & Zyglidopoulos, S. C. (2008). The escalation of deception in organizations. Journal of Business Ethics. 81(4): 837–850. https://doi.org/10.1007/s10551-007-9551-9

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W.(2013). Honesty saves time (and justifications). Frontiers in Psychology. 4 (473): 1–2. https://doi.org/10.3389/fpsyg.2013.00473

Friedman. D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* Cambridge University Press: Cambridge, MA.

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. Nature Neuroscience. 19(12): 1727–1732. https://doi.org/10.1038/nn.4426

Geiger, M., & Van Der Laan Smith, J. (2010). The effect of institutional and cultural factors on the perceptions of earnings management. Journal of International Accounting Research. 9(2): 21–43. https://doi.org/10.2308/jiar.2010.9.2.21

Gibson, R., Tanner, C., & Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. American Economic Review. 103(1): 532–548. https://doi.org/10.1257/aer.103.1.532

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. Journal of Economic Behavior and Organization. 93: 285–292. https://doi.org/10.1016/j.jebo.2013.04.005

Gneezy, U. (2005). Deception: The role of consequences. American Economic Review. 95(1): 384–394. https://doi.org/10.1257/0002828053828662

Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: emotion dynamics and social behavior. Management Science. 60(11): 2645–2658. https://doi.org/10.1287/mnsc.2014.1942

Gneezy, U., Kajackaite, A., & Sobel, J.(2018). Lying aversion and the size of the lie. American Economic Review.108(2): 419–453.https://doi.org/10.1257/aer.20161553

Greenberg, J. (1987). The college sophomore as guinea pig: Setting the record straight. Academy of Management Review.12(1): 157-159. https://doi.org/10.5465/amr.1987.4306516

Hales, J. (2009). Are investors really willing to disagree? An experimental investigation of how disagreement and attention to disagreement affect trading behavior. Organizational Behavior and Human Decision Processes.108(2): 230–241. https://doi.org/10.1016/j.obhdp.2008.08.003

Haynes, C. M., & Kachelmeier, S. J.(1998). The effects of accounting contexts on accounting decisions: A synthesis of cognitive and economic perspectives in accounting experimentation. Journal of Accounting Literature.17:97–136.

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. https://doi.org/10.1111/joes.12204

Kartik, N. (2009). Strategic communication with lying costs. Review of Economic Studies.76(4): 1359–1395. https://doi.org/10.1111/j.1467-937X.2009.00559.x

Kerschbamer, R., Neururer, D., & Gruber, A. (2019). Do altruists lie less? Journal of Economic Behavior and Organization.157(C): 560–579. https://doi.org/10.1016/j.jebo.2018.10.021

Kramer, S., & Maas, V. S.(2020). Selective attention as a determinant of escalation bias in subjective performance evaluation judgments. Behavioral Research in Accounting.32(1): 87–100. https://doi.org/10.2308/bria-18-021

Lasarov, W., & Hoffmann, S. (2020). Social moral licensing. Journal of Business Ethics.165: 45–66. https://doi.org/10.1007/s10551-018-4083-z

Lee, J. J., Hardin, A. E., Parmar, B., & Gino, F. (2019). The interpersonal costs of dishonesty: How dishonest behavior reduces individuals' ability to read others' emotions. Journal of Experimental Psychology: General.148(9): 1557–1574. https://doi.org/10.1037/xge0000639

Lohse, T., Simon, S. A., & Konrad, K. A. (2018). Deception under time pressure: Conscious decision or a problem of awareness? Journal of Economic Behavior and Organization.146: 31–42. https://doi.org/10.1016/j.jebo.2017.11.026

López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. Experimental Economics.16(3): 233–247. https://doi.org/10.1007/s10683-012-9324-x

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. Journal of Economic Behavior and Organization. 70(1–2): 81–92. https://doi.org/10.1016/j.jebo.2009.02.010

Masuda, T., Okano, Y., & Saijo, T. (2014). The minimum approval mechanism implements the efficient public good allocation theoretically and experimentally. Games and Economic Behavior. 83: 73–85. https://doi.org/10.1016/j.geb.2013.10.003

McLean, B., Elkind, P. (2013). *The smartest guys in the room: The amazing rise and scandalous fall of Enron.* Penguin: Calgary, Canada.

Moore, C. (2015). Moral disengagement. Current Opinion in Psychology. 6: 199–204. https://doi.org/10.1016/j.copsyc.2015.07.018

Moser. D.V. (1998). Using an experimental economics approach in behavioral accounting research. Behavioral Research in Accounting.10: 94–110.

Nelson, M. W., Bloomfield, R., Hales, J. W., & Libby, R. (2001). The effect of information strength and weight on behavior in financial markets. Organizational Behavior and Human Decision Processes. 86(2): 168–196. https://doi.org/10.1006/obhd.2000.2950

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological Science. 31(7): 770–780. https://doi.org/10.1177/0956797620939054

Pickering, W. (1846). *The works of George Herbert in prose and verse, Vol. I.* George Routledge, London.

Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. Journal of Economic Behavior and Organization. 93: 374–383. https://doi.org/10.1016/j.jebo.2013.03.030

Remus, W. (1986). Graduate students as surrogates for managers in experiments on business decision making. Journal of Business Research. 14(1): 19–25. https://doi.org/10.1016/0148-2963(86)90053-6

Rosenbaum, S. M., Billinger, S., & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. Journal of Economic Psychology.45: 181–196. https://doi.org/10.1016/j.joep.2014.10.002

Saijo, T. (2019). Second thoughts of social dilemma in mechanism design. In: W. Trockel W. (ed.). Social design: Essays in memory of Leonid Hurwicz, (pp.157–171). Springer Nature, London.

Schurr, A., & Ritov, I. (2016). Winning a competition predicts dishonest behavior. Proceedings of the National Academy of Sciences.113(7):1754–1759. https://doi.org/10.1073/pnas.1515102113

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). Psychological Science. 23(10): 1264–1270. https://doi.org/10.1177/0956797612443835

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. Personality and Social Psychology Bulletin. 37(3): 330–349. https://doi.org/10.1177/0146167211398138

Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. Organizational Behavior and Human Performance. 16(1): 27–44. https://doi.org/10.1016/0030-5073(76)90005-2

Staw, B.M. (1981). The escalation of commitment to a course of action. Academy of Management Review.6(4): 577–587. https://doi.org/10.5465/amr.1981.4285694

Staw, B. M., & Ross, J. (1989). Understanding behavior in escalation situations. Science. 246(4927): 216–220. https://doi.org/10.1126/science.246.4927.216

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. https://doi.org/10.1177/0956797612458806

Ward, E. A. (1993). Generalizability of psychological research from undergraduates to employed adults. Journal of Social Psychology.133(4): 513–519. https://doi.org/10.1080/00224545.1993.9712176

Welsh, D. T., Ordóñez, L. D., Snyder, D. G., & Christian, M. S. (2015). The slippery slope: how small ethical transgressions pave the way for larger future transgressions. Journal of Applied Psychology.100: 114–127. https://doi.org/10.1037/a0036950

West, C., & Zhong, C. B. (2015). Moral cleansing. Current Opinion in Psychology. 6: 221–225. https://doi.org/10.1016/j.copsyc.2015.09.022

Wu, Y., Blue, P. R., & Clark, L. (2017). Commentary: Winning a competition predicts dishonest behavior. Frontiers in Neuroscience. 11: 417. https://doi.org/10.3389/fnins.2017.00417

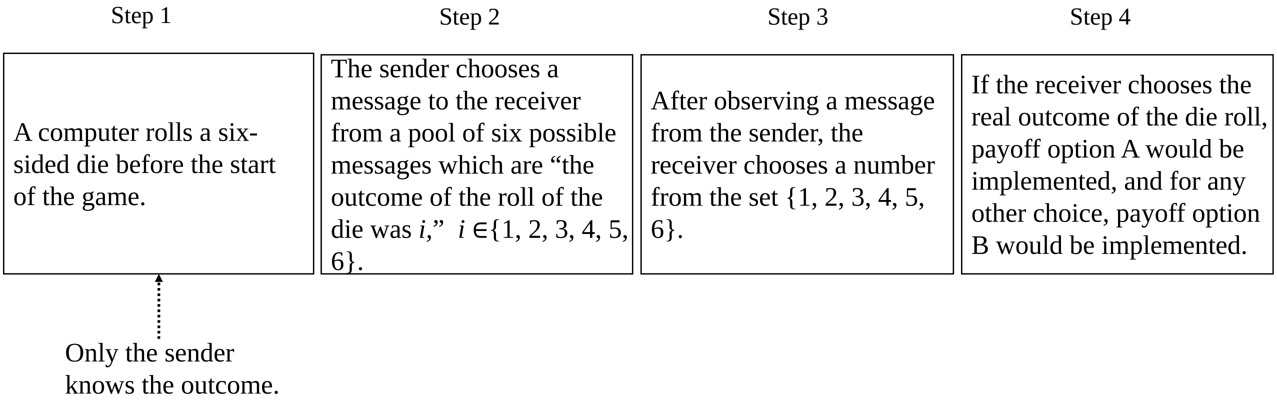

Step 1
Step 2
Step 3
Step 4
A computer rolls a six-sided die before the start of the game.
The sender chooses a message to the receiver from a pool of six possible messages which are “the outcome of the roll of the die was *i*,” *i* ∈{1, 2, 3, 4, 5, 6}.
After observing a message from the sender, the receiver chooses a number from the set {1, 2, 3, 4, 5, 6}.
If the receiver chooses the real outcome of the die roll, payoff option A would be implemented, and for any other choice, payoff option B would be implemented.
Only the sender knows the outcome.

Figure 1. The Timeline of the Deception Game

Note: In this study, we used the simple deception game that Erat and Gneezy (2012) developed in their study. In the deception game, two players act sequentially in the role of sender and receiver, respectively.

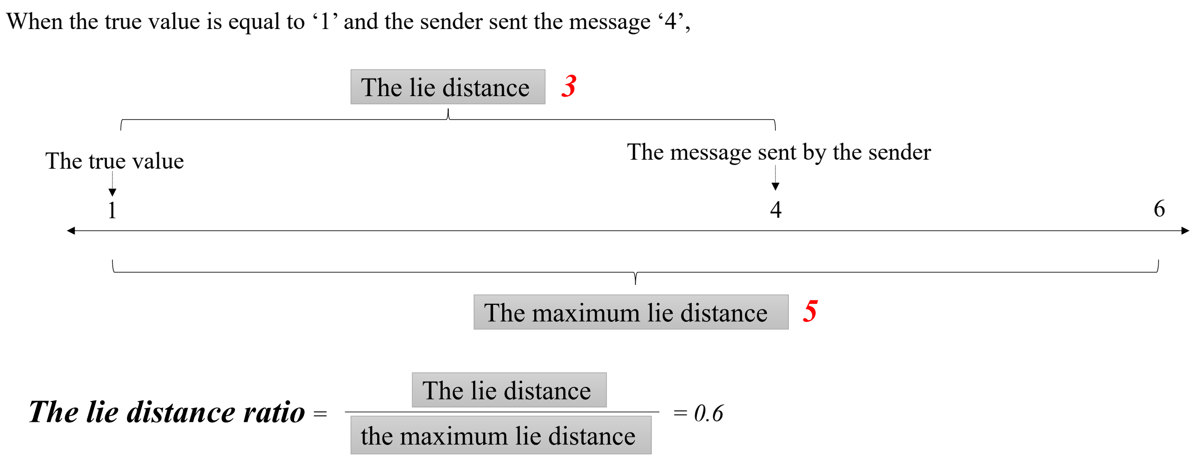
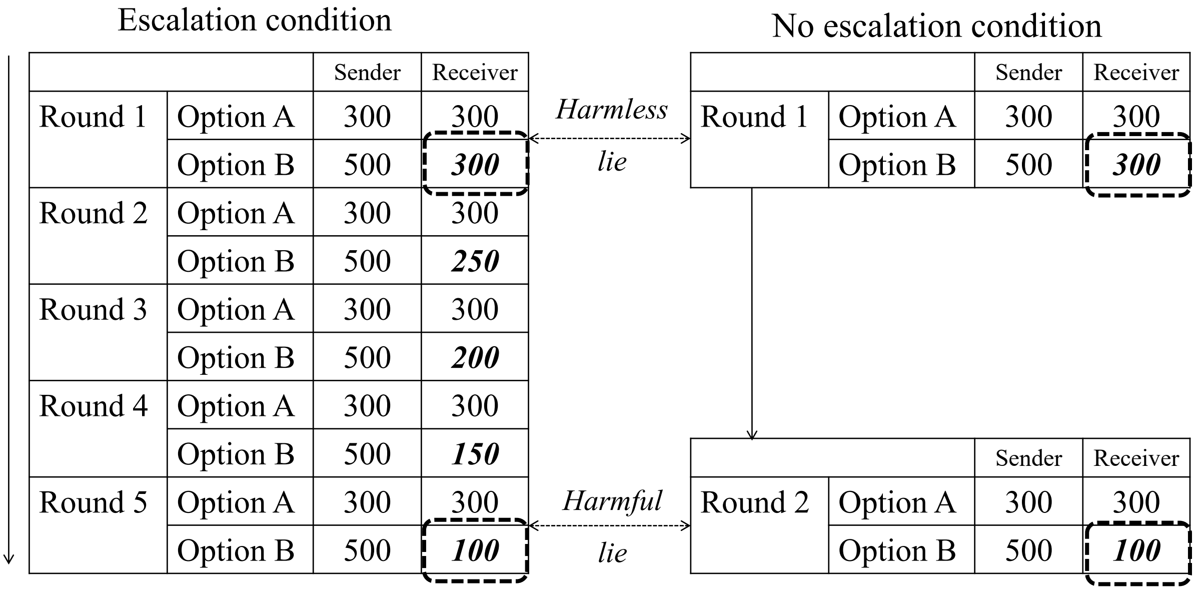


Figure 2. Definition of the Lie Distance Ratio of the Sender

Note: This figure shows the lie distance ratio of the sub-sample restricted to the sender reporting a lie under each condition. We define *the lie distance ratio* as the ratio obtained by dividing *the lie distance* by *the maximum lie distance*. Firstly, we define *the lie distance* of the sender as the absolute value of the difference between the true value and the message value which the sender sent. When the true value is equal to "1" and the sender sends the message "4," for example, the level of the lie distance is equal to "3." *The maximum lie distance* depends on the true value: when the true value is equal to "1," for example, the maximum lie distance is equal to "5." When the true value is equal to "2," "3," "4," "5," and "6," the maximum lie distance is equal to "4," "3," "3," "4," and "5," respectively. In the case of this figure, we calculate the lie distance ratio equal to 0.6.



Escalation condition

		Sender	Receiver
Round 1	Option A	300	300
	Option B	500	***300***
Round 2	Option A	300	300
	Option B	500	***250***
Round 3	Option A	300	300
	Option B	500	***200***
Round 4	Option A	300	300
	Option B	500	***150***
Round 5	Option A	300	300
	Option B	500	***100***

No escalation condition

		Sender	Receiver
Round 1	Option A	300	300
	Option B	500	***300***

		Sender	Receiver
Round 2	Option A	300	300
	Option B	500	***100***

Figure 3. The Two Conditions of the Experiment

Note: Under the escalation condition, a sender has numerous opportunities to report lies because there are five rounds. The payoff gradually changes from “harmless lie” in the first round to “harmful lie” in the last round with repetition under the escalation condition. However, under the no escalation condition, a sender has few opportunities to report lies because there are only two rounds, although the payoff structure of the first and the last rounds is the same as that under the escalation condition.

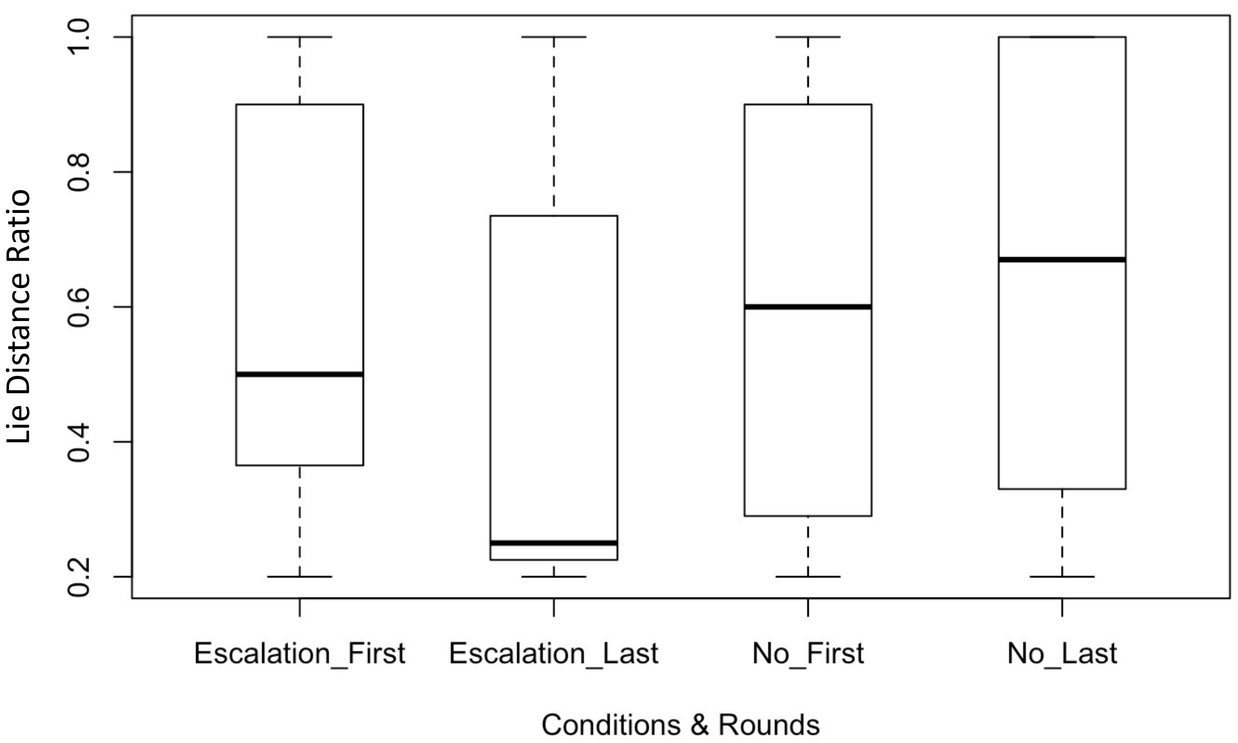
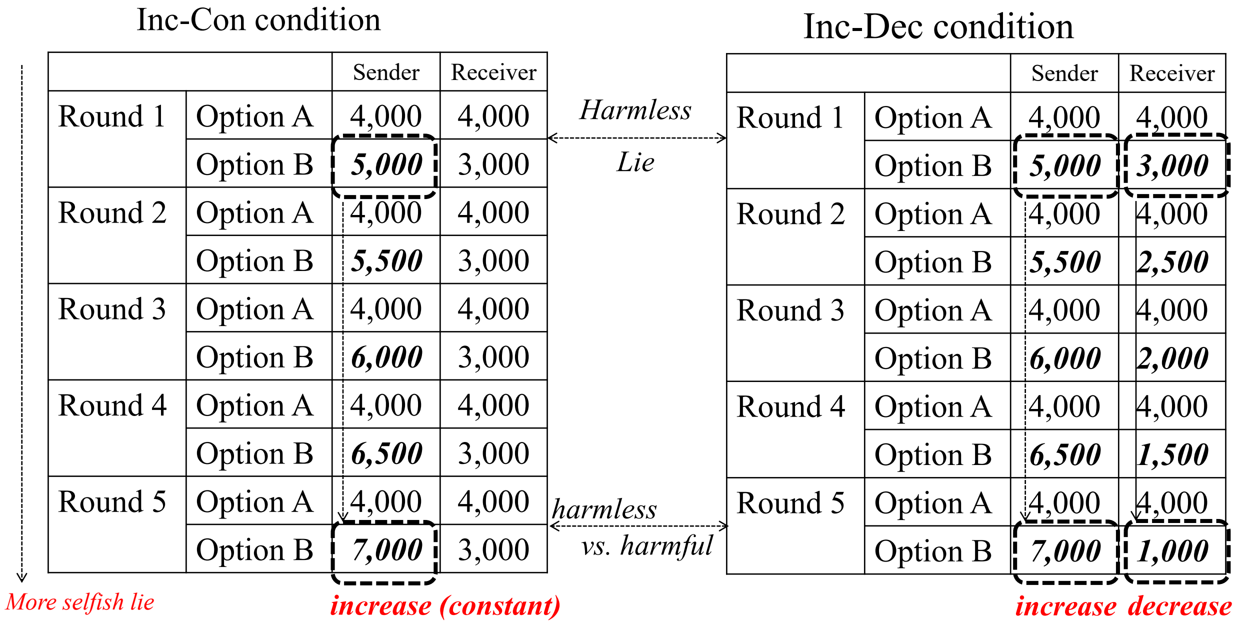


Figure 4. Boxplot of the Lie Distance Ratio of the Sender by Round and Each Condition (Sub-sample Restricted to the Sender Reporting a Lie)

Note: This figure shows the boxplot of the lie distance ratio of the sub-sample restricted to the sender reporting a lie under each condition. We define *the lie distance ratio* as the ratio obtained by dividing *the lie distance* by *the maximum lie distance*. Please see the note appended to Figure 2 for a description of the lie distance and maximum lie distance. “Escalation” indicates the escalation condition, “No” the no escalation condition, “First” the data for the first round, and “Last” the data for the last round.



Inc-Con condition

		Sender	Receiver
Round 1	Option A	4,000	4,000
	Option B	***5,000***	3,000
Round 2	Option A	4,000	4,000
	Option B	***5,500***	3,000
Round 3	Option A	4,000	4,000
	Option B	***6,000***	3,000
Round 4	Option A	4,000	4,000
	Option B	***6,500***	3,000
Round 5	Option A	4,000	4,000
	Option B	***7,000***	3,000

Inc-Dec condition

		Sender	Receiver
Round 1	Option A	4,000	4,000
	Option B	***5,000***	***3,000***
Round 2	Option A	4,000	4,000
	Option B	***5,500***	***2,500***
Round 3	Option A	4,000	4,000
	Option B	***6,000***	***2,000***
Round 4	Option A	4,000	4,000
	Option B	***6,500***	***1,500***
Round 5	Option A	4,000	4,000
	Option B	***7,000***	***1,000***

Figure 5. The Two Conditions of the Follow-up Experiment

Note: Under the Inc-Con condition, the payoff structure of the deception game gradually changes from "*Little selfish and harmless lie*" to "*Selfish and little harmless lie*" with repetition, where the self-payoff increases but the others-payoff remains constant (Table 3). However, under the Inc-Dec condition, the payoff structure of the deception game gradually changes from "*Little selfish and harmless lie*" to "*Selfish and harmful lie*" with repetition, where the self-payoff increases and the others-payoff decreases (Table 3). Although the payoffs of both conditions gradually change selfishly, the difference emerges when the others-payoff decreases.

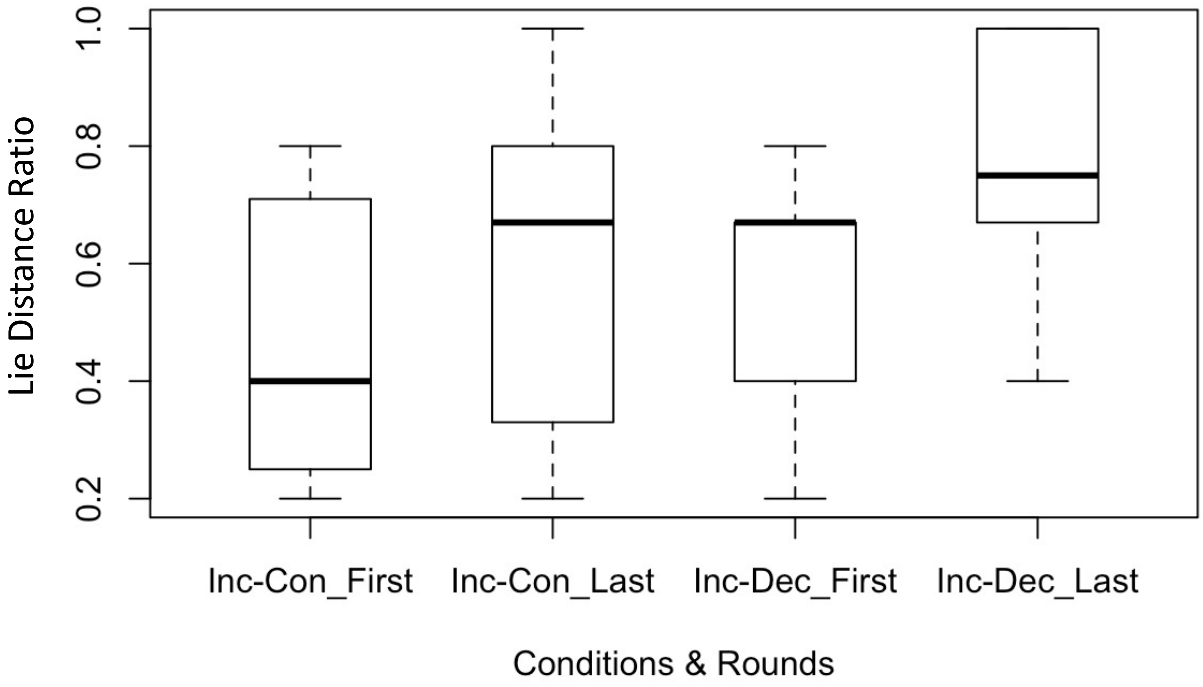
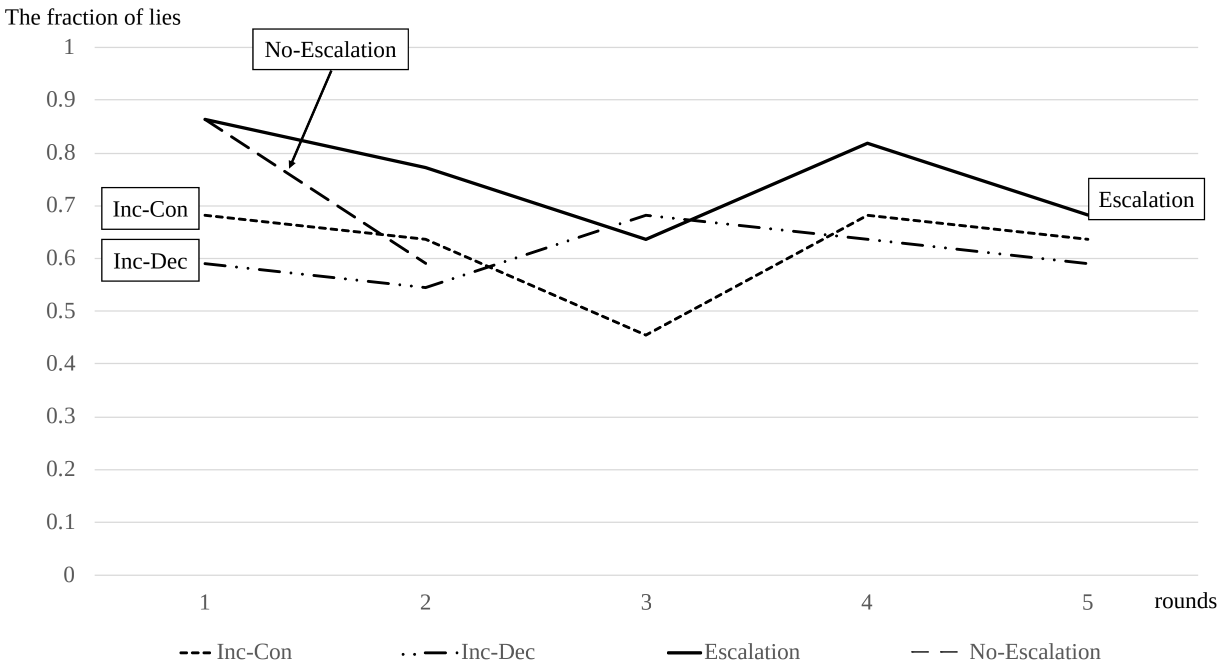


Figure 6. Boxplot of the Lie Distance Ratio of the Sender by Round and Each Condition of the follow-up experiment (Sub-sample Restricted to the Sender Reporting a Lie)

Note: This figure shows the boxplot of the lie distance ratio of the follow-up experiment of the sub-sample restricted to the sender reporting a lie under each condition. We define *the lie distance ratio* as the ratio obtained by dividing *the lie distance* by *the maximum lie distance*. Please see the note appended to Figure 2 for a description of the lie distance and maximum lie distance. "Inc-Con" indicates the Inc-Con condition, "Inc-Dec" the Inc-Dec condition, "First" the data for the first round, and "Last" the data for the last round.

Panel A. Transition of the fraction of lies by round and each condition



Panel B. Transition of the mean levels of the lie distance ratio by round and each condition

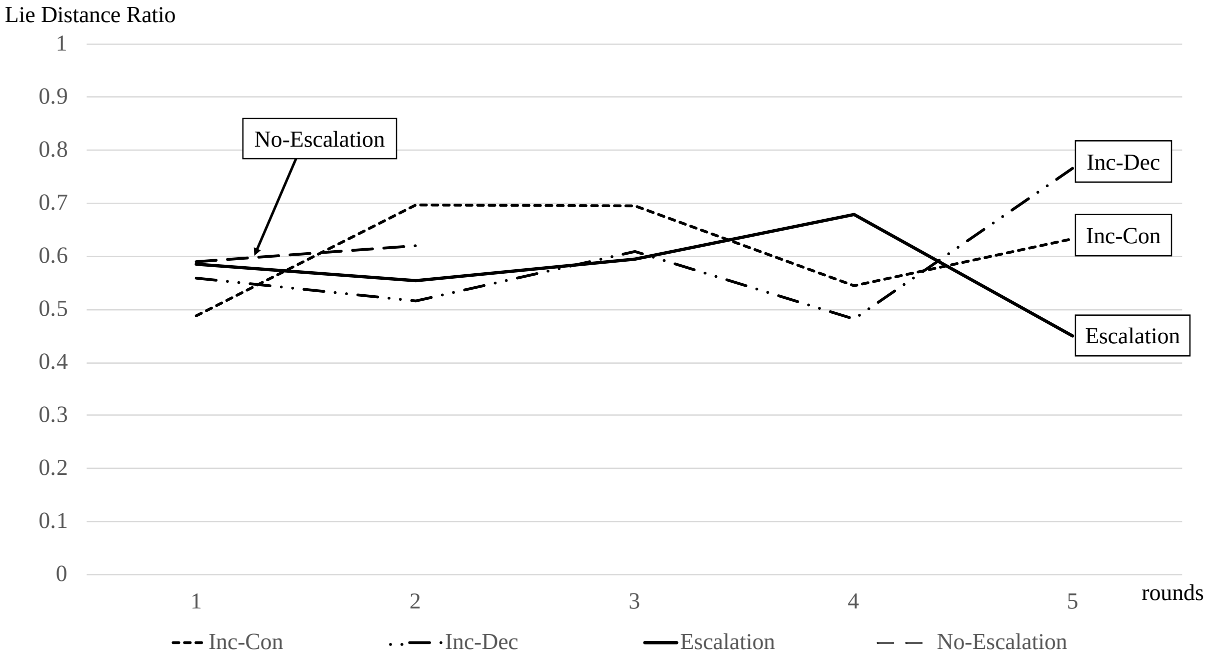


Figure 7. Transition of the Senders' Behavior for Each Round by Each Condition

Note: This figure shows the transition of the senders' behavior for each round by each condition. Panel A shows the transition of the fraction of lies and Panel B shows the transition of the mean levels of the lie distance ratio, by round and each condition. We define *the lie distance ratio* as the ratio obtained by dividing *the lie distance* by *the maximum lie distance*. Please see the note appended to Figure 2 for a description of the lie distance and maximum lie distance.

Table 1. The Different Payoffs Associated with Type of Lies and Options

	(1) Harmless lie		(2) Harmful lie	
	Sender	Receiver	Sender	Receiver
Option A	300	300	300	300
Option B	500	***300***	500	***100***

Note: Under the first type, the payoff for option A is (300, 300) and that of option B is (500, *300*). We call this a "Harmless lie," in which the receiver would suffer no damage on being deceived. Under the second type, the payoff for option A is (300, 300) and that of option B is (500, *100*). We call this a "Harmful lie," in which the receiver would suffer serious damage on being deceived.

Table 2. Descriptive Statistics of Senders' Decisions for Each Experimental Condition

Panel A. The fraction of lies by each condition

	First round	Last round	Fisher's Exact Test (one-tailed)
Escalation condition	86.36%	68.18%	*p = 0.140*, ϕ = 0.2310
No escalation condition	86.36%	59.09%	*p = 0.044*, ϕ = 0.3610

Panel B. The lie distance ratio by each condition (only the lying sub-sample)

Condition		First round	Last round
Escalation	Mean	0.59	0.45
	Median	0.50	0.25
	SD	0.29	0.33
	Obs.	19	15
No escalation	Mean	0.59	0.62
	Median	0.60	0.67
	SD	0.31	0.32
	N	19	13

Panel C. Decision time (seconds) of the sender by each condition (full sample)

Condition		First round	Last round
Escalation	Mean	22.05	19.09
	Median	20.50	18.00
	SD	5.10	7.09
	N	22	22
No escalation	Mean	23.77	22.86
	Median	23.50	21.50
	SD	10.19	8.01
	N	22	22

Note: This table shows the descriptive statistics for each experimental condition.

Panel A shows the descriptive statistics of the fraction of lies under each condition.

Panel B shows the descriptive statistics of the lie distance ratio of the sub-sample restricted to the sender reporting a lie under each condition. The lie distance ratio is the ratio obtained by dividing the lie distance, which means the absolute value of the difference between the true value and the message value which the sender sent, by the maximum lie distance. The maximum lie distance depends on the true value: when the true value is equal to "1," the maximum lie distance is equal to "5" (when the sender sends the message "6"). When the true value is equal to "2," "3," "4," "5," and "6," the maximum lie distance is equal to "4," "3," "3," "4," and "5," respectively. SD, standard deviation; N, number of observations.

Panel C shows the descriptive statistics of decision time (seconds) for each condition. SD, standard deviation; N, the number of observations.

Table 3. Different Payoffs Associated with the Type of Lies and Options of the Follow-up Experiment

	(1) Little selfish and harmless lie		(2) Selfish and harmless lie		(3) Selfish and harmful lie	
	Sender	Receiver	Sender	Receiver	Sender	Receiver
Option A	4,000	4,000	4,000	4,000	4,000	4,000
Option B	5,000	***3,000***	***7,000***	***3,000***	***7,000***	***1,000***

Note: Under the first type, the payoff for option A is (4,000, 4,000) and that of option B is (5,000, *3,000*); we call this a "*Little selfish and harmless lie*," in which the sender would obtain little additional payoff and the receiver would suffer little damage on being deceived. Under the second type, the payoff for option A is (4,000, 4,000) and that of option B is (*7,000*, *3,000*); we call this a "*Selfish and harmless lie*," in which the sender would obtain a high payoff but the receiver would suffer little damage on being deceived. Under the third type, the payoff for option A is (4,000, 4,000) and that of option B is (*7,000*, *1,000*). We call this a "*Selfish and harmful lie*," in which the sender would obtain a high payoff and the receiver would suffer great damage on being deceived. Under the Inc-Con condition, the payoff structure of the deception game gradually changes from (1) to (2) with repetition. However, under the Inc-Dec condition, the payoff structure of the deception game gradually changes from (1) to (3) with repetition. Although the payoffs of both conditions gradually change selfishly, the difference emerges when the others-payoff decreases.

Table 4. Descriptive Statistics for Each Experimental Condition of the Follow-up Experiment

Panel A. Fraction of lies by each condition

	First round	Last round	
Inc-Con condition	68.18%	63.64%	n.s.
Inc-Dec condition	59.09%	59.09%	n.s.

Panel B. Lie distance ratio by each condition (only the lying sub-sample)

		First round	Last round
Inc-Con condition	Mean	0.488	0.633
	Median	0.400	0.670
	SD	0.237	0.273
	N	22	22
Inc-Dec condition	Mean	0.560	0.766
	Median	0.670	0.750
	SD	0.212	0.194
	N	22	22

Panel C. Decision time (seconds) of the sender by each condition (full sample)

Condition		First round	Last round
Inc-Con	Mean	26.04	22.59
	Median	22.00	19.00
	SD	10.05	9.42
	N	22	22
Inc-Dec	Mean	25.00	21.68
	Median	25.00	21.50
	SD	5.53	4.95
	N	22	22

Panel D. Decision time (seconds) of the sender by each condition (only the lying sub-sample)

Condition		First round	Last round
Inc-Con	Mean	27.86	24.42
	Median	26.00	20.00
	SD	11.78	11.12
	N	15	14
Inc-Dec	Mean	26.30	22.23
	Median	27.00	21.00
	SD	6.18	4.72
	N	13	13

Note: This table shows the descriptive statistics for each experimental condition.

Panel A shows the descriptive statistics of the fraction of lies under each condition. n.s., not significant.

Panel B shows the descriptive statistics of the lie distance ratio of the sub-sample restricted to the sender reporting a lie under each condition. The lie distance ratio means the ratio obtained by dividing the lie distance, which means the absolute value of the difference between the true value and the message value which the sender sent, by the maximum lie distance. The maximum lie distance depends on the true value: when the true value is equal to “1,” the maximum lie distance is equal to “5” (when the sender sends the message “6”). When the true value is equal to “2,” “3,” “4,” “5,” and “6,” the maximum lie distance is equal to “4,” “3,” “3,” “4,” and “5,” respectively. SD, standard deviation; N, number of observations.

Panels C and D show the descriptive statistics of decision time (seconds) for each condition. Panel C presents the result of the full sample; Panel D presents the result of the sub-sample restricted to the sender reporting a lie. SD, standard deviation; N, number of observations.

Table 5. Analysis of Individual Behavior of the Sender: Decision Switching Ratio of the Sender for Each Condition

Panel A. Decision switching ratio for each condition

	Main experiment		Follow-up experiment	
	Escalation	No escalation	Inc-Con	Inc-Dec
Mean	0.34	0.55	0.35	0.41
Median	0.25	1.00	0.38	0.50
SD	0.35	0.50	0.33	0.32
N	22	22	22	22

Panel B. Frequency of decision switching by each condition

		Main experiment		Follow-up experiment	
		Escalation	No escalation	Inc-Con	Inc-Dec
Number of decision switching	0	10	10	9	7
	1	2	12	2	1
	2	3	NA	5	8
	3	6	NA	5	5
	4	1	NA	1	1

Panel C Frequency of decision switching by each condition (even–odd classification)

		Main experiment		Follow-up experiment	
		Escalation	No escalation	Inc-Con	Inc-Dec
Number of decision switching	Even	14	10	15	16
	Odd	8	12	7	6

Note: This table shows an analysis of *the decision switching ratio of the sender* by each condition, defined as follows: $The\ decision\ switching\ ratio = \frac{The\ number\ of\ switching}{The\ number\ of\ all\ rounds - 1}$. This ratio indicates the ratio of the decision switching for all opportunities and implies the degree of moral cleansing: the higher the ratio, the greater the degree of moral cleansing of the sender. Panel A shows that the mean and median levels of the ratio, and Panels B and C, the frequency of decision switching (the number and even–odd classification, respectively). In Panel C, an even number of switches indicates that the same decision was made at the end. SD, standard deviation; N, number of observations; NA, not applicable.

Supplementary Information for

The Effect of Escalating Lies on Business Ethics:

An Experimental Study of the Repeated Deception Game

This file includes:

Supplement 1. Instructions used for the experiment

Instructions for role A (Sender)

Welcome to our experiment. Please read these instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment. The remaining instructions describe what the procedure will be should you be chosen.

You will be matched randomly with another participant in this experiment. Neither of you will know the identity of the other.

Before beginning this experiment, a computer has rolled a six-sided die and obtained the outcome. The other participant will not be informed of the real outcome of the die roll. However, he/she will be told that you have been informed about the outcome of the die roll.

We now ask you to send a message to the other participant. The message will correspond to a number from 1 to 6. There are six possible messages:

Message 1: “The outcome from the roll of the six-sided die is 1.”
Message 2: “The outcome from the roll of the six-sided die is 2.”
Message 3: “The outcome from the roll of the six-sided die is 3.”
Message 4: “The outcome from the roll of the six-sided die is 4.”
Message 5: “The outcome from the roll of the six-sided die is 5.”
Message 6: “The outcome from the roll of the six-sided die is 6.”

Your message will be shown to the other participant, and then he/she will be asked to choose a number between 1 and 6. The choice of the number by the other participant will determine the payments in the experiment. Specifically, your acquisition points are determined according to the following two options.

[Option A]: If he/she chooses the actual outcome of the roll of the dice, then you will receive 300 points and he/she will receive 300 points.
[Option B]: If he/she chooses a number different than the actual outcome, you will receive X points and he/she will receive Y points.

The values of X and Y are different for each round. The specific values of X and Y in each round are provided when selecting the message to send to the receiver. Finally, only you will be informed of the particular monetary value connected to each message. The other participant will not be informed of these monetary values. However, he/she will be told that you have been informed of the monetary value connected to each message.

Instructions for role B (Receiver)

Welcome to our experiment. Please read these instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

You will be matched randomly with another participant in this experiment. Neither of you will know the identity of the other.

Before starting this experiment, a computer has rolled a six-sided die, and the outcome has been informed to the other participant, but the computer is not going to reveal this information to you.

After being informed of the roll of the die, the other participant has sent a message to you. The message corresponds to a number from 1 to 6. There are six possible messages:
Message 1: “The outcome from the roll of the die is 1.”
Message 2: “The outcome from the roll of the die is 2.”
Message 3: “The outcome from the roll of the die is 3.”
Message 4: “The outcome from the roll of the die is 4.”
Message 5: “The outcome from the roll of the die is 5.”
Message 6: “The outcome from the roll of the die is 6.”
Now we ask you to choose a number between 1 and 6. The message you receive is the only information you will have regarding the roll of the die. Your choice of a number will determine the payments in the experiment according to two different options (option A and option B), known only to the other participant.

If you choose the same number as the number that came up in the roll of the die, both of you will be paid according to option A. However, if you choose a number different from the actual number, both of you will be paid according to option B.

Supplement 2. Post questionnaire to test the understanding level of participants

Your age:___years old

Your gender: Male (0), Female (1)

Your work experience in years:________

1. Understanding level of the experiment: Please rate your level of understanding from 1 to 7 (1: "I could not understand at all"; 7: "I understood very well")

1 2 3 4 5 6 7

2. Understanding level in the identification of the partner of the experiment: Have you been able to identify the partner during the experiment? Please rate your level of understanding from 1 to 7 (1: "I did not know at all who was my partner"; 7: "I completely identified my partner")

1 2 3 4 5 6 7

3. Understanding level of experimental rewards: Did you participate in the experiment with the understanding that you could receive higher experimental rewards when you got higher points in the experiment? Please rate your level of understanding from 1 to 7 (1: "I could not understand at all"; 7: "I understood very well")

1 2 3 4 5 6 7

4. Understanding level of experimental rules: Did you understand the rule that role A will know the number chosen by the computer each time, but role B will not know it? Please rate your level of understanding from 1 to 7 (1: "I could not understand at all; 7: "I understood very well")

1 2 3 4 5 6 7

Supplement 3. Robustness check for the main experiment

S3-1. Regression analysis of hypothesis 1 (H1)

To check the robustness of the results in the main experiment described in sections III and IV, we perform a regression analysis because it enables us to control for all other individual factors (e.g., gender, morality, and lie acceptability) that might affect lying behavior.

First, to check the validity of H1, a probit regression analysis is performed on the samples of senders' behavior in the first and last rounds under the no escalation condition. The model to be tested is as follows:

$$Y_i = \alpha + \beta_1 Last + \beta_2 Gen_i + \beta_3 Dark_i + \beta_4 Accept_i + \beta_5 MFQ_i + \varepsilon_i \qquad (1)$$

where *Yi* represents a dummy variable of subject i, that takes 1 if the sender tells a lie and 0 if the sender tells the truth under the message choice. *Last* is a dummy variable that takes 1 if the round is last ("harmful lie") and 0 if first ("harmless lie"). According to H1 (lying aversion), we hypothesize that the fraction levels of lie in the last round ("harmful lie") is higher than that in the first round ("harmless lie"); that is, we predict β_1 is negative.

The following four variables represent the subjects' personalities that are generally controlled in previous studies (see Supplement 4). *Gen* is a dummy variable that takes 1 if the subject is female and 0 if the subject is male. *Dark* expresses the Dark Triad Dirty Dozen (DTDD) measure that Jonason and Webster (2010) provide. The dark triad measure consists of three factors: narcissism, Machiavellianism, and psychopathy. These three characteristics are representative of antisocial personality.[18] *Accept* shows the total score of the lie acceptability scale, which is a measure used to evaluate an individual's attitude about deceptive communication, as shown by Oliveira and Levine (2008).[19] *MFQ* expresses the moral foundations questionnaire (MFQ) that is designed to assess the degree to which people prioritize five foundational domains in moral decision-making: Harm/Care, Fairness/Reciprocity, Ingroup/Loyalty, Authority/Respect, and Purity/Sanctity (Graham et al. 2011).[20] The results of the Pearson correlation analysis among the independent variables of the model of H1 are summarized in the Table S3-1, and the results of the model are summarized in Table S3-2.

[Insert Table S3-1 and S3-2 about here.]

We focus on the coefficients of *Last*. The coefficients of *Last* are negative and significant at $p < 0.05$. We predict that they have a negative sign because of H1 (lying aversion): specifically, we predict that the fraction levels of lies in the last round ("harmful lie") under the no escalation condition is lower than that in

[18] Majors (2016) experimentally examines the interaction of communicating measurement uncertainty and the dark triad score on managers' reporting decisions.

[19] Wright et al. (2015) experimentally show that good liars tend to get a high score on the lie acceptability measure.

[20] These five domains are described in the Moral Foundations Theory, which is grounded in evolutionary theory and supported by ethnographic evidence worldwide (Haidt and Graham 2007).

the first round (“harmless lie”). Thus, even after controlling for the personality of the subjects, under the no escalation condition, the lying aversion effect exists. This result supports H1 (lying aversion).[21]

S3-2 Regression analysis of H2

Second, to check the validity of H2, a probit regression analysis is performed on the subsamples of senders’ behavior in the first and last rounds under the escalation condition. The model to be tested follows:

$$Y_i = \alpha + \beta_1 Last + \beta_2 Gen_i + \beta_3 Dark_i + \beta_4 Accept_i + \beta_5 MFQ_i + \varepsilon_i \qquad (2)$$

Yi represents a dummy variable of subject i, which takes 1 if the sender tells a lie and 0 if the sender tells the truth under the message choice. *Last* is a dummy variable that takes 1 if the round is last (“harmful lie”) and 0 if first (“harmless lie”). According to H2 (lying escalation), we hypothesize that the fraction levels of lie in the last round (“harmful lie”) is not higher than that in the first round (“harmless lie”); that is, we predict β_1 is not negative. The following four variables represent the same personalities that are controlled in the previous subsection. The results of the Pearson correlation analysis among the independent variables of the model of H2 correlation are summarized in Table S3-3, and the results of the model are summarized in Table S3-4.

[Insert Table S3-3 and S3-4 about here.]

We focus on the coefficients of *Last*. The coefficients of *Last* are not significant at $p < 0.05$. We predict that they have a non-negative sign because of H2 (lying escalation): specifically, we predict that the fraction level of lies in the last round (“harmful lie”) under the escalation condition is not lower than that in the first round (“harmless lie”). Thus, even after controlling for the personality of the subjects, under the escalation condition, the lying escalation effect exists. This result supports H2 (lying escalation).

[21] The sign of the MFQ_Fairness term in Model 5 was negative (significant at $p < 0.01$), indicating that the higher the fairness, the less likely one is to lie in the case of a harmful lie, which may hurt the others.

Supplement 4. Personality measures of the experiment

We used some personality measures to control individual characteristics of the subjects, and participants in study 1 answered questions related to these measures: Dark triad, Lie Acceptability Scale, and Moral Foundations Questionnaire (MFQ).

Dark triad

1. I tend to manipulate others to get my way.
2. I have used deceit or lied to get my way.
3. I have used flattery to get my way.
4. I tend to exploit others toward my own end.
5. I tend to lack remorse.
6. I tend to be unconcerned with the morality of my actions.
7. I tend to be callous or insensitive.
8. I tend to be cynical.
9. I tend to want others to admire me.
10. I tend to want others to pay attention to me.
11. I tend to seek prestige or status.
12. I tend to expect special favors from others.

Note: We use the DTDD measure, which is a brief version of the Dark triad measure by Jonason and Webster (2010).

Lie acceptability scale

1. Never tell anyone the real reason you do anything unless it is useful to do so.
2. Lying is immoral. (R)
3. It is okay to lie in order to achieve one's goals.
4. What people do not know cannot hurt them.
5. The best way to handle people is to tell them what they want to hear.
6. There is no excuse for lying to someone else. (R)
7. Honesty is always the best policy. (R)
8. It is often better to lie than to hurt someone's feelings.
9. Lying is just wrong. (R)
10. Lying is no big deal.
11. There is nothing wrong with bending the truth now and then.

Note: We use the lie acceptability scale, which is the measure used to evaluate an individual's attitude about deceptive communication, as designed by Oliveira and Levine (2008). (R) indicates reverse-scored items.

MFQ

1. Whether or not someone suffered emotionally.
2. Whether or not some people were treated differently than others.

3. Whether or not someone's action showed love for his or her country.

4. Whether or not someone showed a lack of respect for authority.

5. Whether or not someone violated standards of purity and decency.

6. Whether or not someone was good at math.

7. Whether or not someone cared for someone weak or vulnerable.

8. Whether or not someone acted unfairly.

9. Whether or not someone did something to betray his or her group.

10. Whether or not someone conformed to the traditions of society.

11. Whether or not someone did something disgusting.

12. Whether or not someone was cruel.

13. Whether or not someone was denied his or her rights.

14. Whether or not someone showed a lack of loyalty.

15. Whether or not an action caused chaos or disorder.

16. Whether or not someone acted in a way that God would approve of.

17. Compassion for those who are suffering is the most crucial virtue.

18. When the government makes laws, the number-one principle should be ensuring that everyone is treated fairly.

19. I am proud of my country's history.

20. Respect for authority is something all children need to learn.

21. People should not do things that are disgusting, even if no one is harmed.

22. It is better to do good than to do bad.

23. One of the worst things a person could do is hurt a defenseless animal.

24. Justice is the most important requirement for a society.

25. People should be loyal to their family members, even when they have done something wrong.

26. Men and women each have different roles to play in society.

27. I would call some acts wrong on the grounds that they are unnatural.

28. It can never be right to kill a human being.

29. I think it is morally wrong that rich children inherit sufficient money while poor children inherit nothing.

30. It is more important to be a team player than to express oneself.

31. If I were a soldier and disagreed with my commanding officer's orders, I would still obey the orders because that is my duty.

32. Chastity is an important and valuable virtue.

Note: We use the MFQ, which is designed to assess the degree to which people prioritize five foundational domains in moral decision-making: Harm/Care, Fairness/Reciprocity, Ingroup/Loyalty, Authority/Respect, and Purity/Sanctity (Graham et al. 2011).

Supplement 5. Experimental Design of the main experiment

	Number of			Number of		
Condition	participants	S	R	rounds	S-obs.	R-obs.
Escalation	44	22	22	5	110	110
No escalation	44	22	22	2	44	44
Total	88	44	44		154	154

Note: S, senders; R, receivers; S-obs., total number of observations for senders; R-obs., total number of observations for receivers. In the escalation condition, the payoff gradually changes from “Harmless lie” in the first round to “Harmful lie” in the last round with repetition. The escalation condition involves 44 participants forming 22 pairs across five rounds, thus providing 110 sender (receiver) observations. In the no escalation condition, the payoff structure of the first and last rounds is the same as that under the escalation condition but there are only two rounds in the game. This condition includes 44 participants forming 22 pairs across two rounds, thus resulting in 44 sender (receiver) observations.

Supplement 6. Statistical Test of Differences of the Lie Distance Ratio by Round and Each Condition

				Tests of differences			
		Round		*t*-test		Mann–Whitney U	
Condition		First	Last	*t*	*p*-value (one-tailed)	U	*p*-value (one-tailed)
Escalation	Mean	0.59	0.45	1.25	0.110	195.50	0.032
	Median	0.50	0.25				
No escalation	Mean	0.59	0.62	0.22	0.589	116.00	0.622
	Median	0.60	0.67				

Note: This table shows the statistical test of differences of the lie distance ratio by round and each condition. Please see the notes appended to Figure 2 for a description of the lie distance ratio.

Supplement 7. Statistical Test of Differences in Decision Time by Round and Each Condition

				Tests of differences			
		Round		*t*-test		Mann–Whitney U	
Condition		First	Last	*t*	*p*-value (one-tailed)	U	*p*-value (one-tailed)
Escalation	Mean	22.05	19.09	1.58	0.060	324.50	0.026
	Median	20.50	18.00				
No escalation	Mean	23.77	22.86	0.32	0.371	264.00	0.306
	Median	23.50	21.50				

Note: This table shows the statistical test of differences in decision time by round and each condition.

Supplement 8. Experimental Design of the Follow-up Experiment

Condition	Number of participants	S	R	Number of rounds	S-obs.	R-obs.
Inc-Con	44	22	22	5	110	110
Inc-Dec	44	22	22	5	110	110
Total	88	44	44		220	220

Note: S, senders; R, receivers; S-obs., total number of observations for senders; R-obs., total number of observations for receivers. In the Inc-Con condition, the payoff gradually changes from *Little selfish and harmless lie* to *Selfish and harmless lie* with repetition, where the self-payoff increases but the others-payoff constants. This condition involves 44 participants forming 22 pairs across five rounds, thus providing 110 sender (receiver) observations. In the Inc-Dec condition, the payoff structure of the deception game gradually changes from *Little selfish and harmless lie* to *Selfish and harmful lie* with repetition, where the self-payoff increases and the others-payoff decreases. This condition includes 44 participants forming 22 pairs across five rounds, thus resulting in 110 sender (receiver) observations.

Supplement 9. Statistical Test of Differences of the Lie Distance Ratio by Round and Each Condition of the Follow-up Experiment

		Round		Tests of differences			
				t-test		Mann–Whitney U	
Condition		First	Last	*t*	*p*-value (one-tailed)	U	*p*-value (one-tailed)
	Mean	0.488	0.633	1.54	0.068	120.00	0.000
Inc-Con	Median	0.400	0.670				
Inc-Dec	Mean	0.560	0.766	2.58	0.008	91.00	0.000
	Median	0.670	0.750				

Note: This table shows the statistical test of differences of the lie distance ratio by round and each condition of the follow-up experiment. Please see the notes appended to Figure 2 for a description of the lie distance ratio.

Supplement 10. Statistical Test of Differences in Decision Time by Round and Each Condition of the Follow-up Experiment

Panel A. Result of the full sample

		Round		Tests of differences			
				t-test		Mann–Whitney U	
Condition		First	Last	*t*	*p*-value (one-tailed)	U	*p*-value (one-tailed)
Inc-Con	Mean	26.04	22.59	1.18	0.123	316.50	0.040
	Median	22.00	19.00				
Inc-Dec	Mean	25.00	21.68	2.09	0.021	327.50	0.022
	Median	25.00	21.50				

Panel B. Result of the sub-sample restricted to the sender reporting a lie

		Round		Tests of differences			
				t-test		Mann–Whitney U	
Condition		First	Last	*t*	*p*-value (one-tailed)	U	*p*-value (one-tailed)
Inc-Con	Mean	27.86	24.42	0.81	0.213	128.00	0.162
	Median	26.00	20.00				
Inc-Dec	Mean	26.30	22.23	1.88	0.035	123.00	0.025
	Median	27.00	21.00				

Note: This table shows the statistical test of differences of the decision time by round and each condition of the follow-up experiment. Panel A presents the result of the full sample; Panel B presents the result of the sub-sample restricted to the sender reporting a lie.

References for the supplement files

Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, Ditto PH (2011) Mapping the moral domain. *J. Pers. Soc. Psychol.* 101:366–385.

Haidt J, Graham J (2007) When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Soc. Just. Res.* 20:90–116.

Jonason PK, Webster GD (2010) The dirty dozen: a concise measure of the dark triad. *Psychol. Assess.* 22:420–432.

Majors TM (2016) The interaction of communicating measurement uncertainty and the dark triad on managers' reporting decisions. *Account. Rev.* 91(3):973–992.

Oliveira CM, Levine TR (2008) Lie acceptability: a construct and measure. *Commun. Res. Rep.* 25(4):282–288.

Wright GR, Berry CJ, Catmur C, Bird G (2015) Good liars are neither 'dark' nor self-deceptive. *PLoS ONE* 10(6):1–11.

Table S3-1. Results of the Pearson correlation analysis among the independent variables of the model of H1

	Gender	Dark_Machi	Dark_Psy	Dark_Na	Lie	MFQ_	MFQ_	MFQ_	MFQ_	MFQ_
		-avellianism	-chopathy	-rcissism	Accept	Harm	Fairness	Loyalty	Authority	Purity
Gender										
Dark_										
Machiavellianism	0.129									
Dark_Psychopathy	−0.046	0.595								
Dark_Narcissism	0.209	0.494	0.151							
Lie Accept	−0.037	0.392	0.187	0.307						
MFQ_Harm	0.459	−0.008	−0.242	0.156	0.071					
MFQ_Fairness	0.326	−0.077	−0.252	−0.019	−0.061	0.443				
MFQ_Loyalty	0.117	−0.307	−0.550	−0.383	−0.180	0.410	0.166			
MFQ_Authority	−0.116	−0.038	−0.116	−0.135	−0.198	0.267	0.116	0.463		
MFQ_Purity	0.631	0.061	−0.437	0.294	0.137	0.573	0.389	0.411	0.082	

Note: Please see Supplement 3 for a description of each term.

Table S3-2. Results of the probit regression analysis of H1

	Model 1		Model 2		Model 3		Model 4		Model 5	
(Intercept)	1.063		0.254		2.341		2.077		2.039	
Last dummy	−0.870	**	−0.893	**	−0.895	**	−1.035	**	−1.100	**
Gender	0.061								−0.408	
Dark_Machiavellianism			0.009						0.092	
Dark_Psychopathy			0.006						0.016	
Dark_Narcissism			0.056						0.060	
Lie_accept					−0.026				−0.054	
MFQ_Harm							0.157	*	0.150	*
MFQ_Fairness							−0.186	**	−0.168	*
MFQ_Loyalty							−0.046		−0.007	
MFQ_Authority							0.060		0.031	
Obs.	44		44		44		44		44	
AIC	53.272		56.285		52.637		54.149		59.949	

Note: Please see Supplement 3 for a description of each term. In the analysis, MFQ_Purity, which had high correlation coefficients, was excluded. **, $p < 0.05$. *. $p < 0.10$.

Table S3-3. Results of the Pearson correlation analysis among the independent variables of the model of H2

	Gender	Dark_Machi -avellianism	Dark_Psy - chopathy	Dark_Na -rcissism	Lie Accept	MFQ_ Harm	MFQ_ Fairness	MFQ_ Loyalty	MFQ_ Authority	MFQ_ Purity
Gender										
Dark_ Machiavellianism	−0.381									
Dark_Psychopathy	0.133	−0.170								
Dark_Narcissism	−0.414	0.575	−0.061							
Lie Accept	−0.574	0.370	0.149	−0.008						
MFQ_Harm	0.193	−0.094	−0.484	0.018	−0.608					
MFQ_Fairness	0.335	−0.187	−0.335	0.040	−0.753	0.710				
MFQ_Loyalty	−0.014	−0.102	0.041	0.397	−0.516	0.327	0.392			
MFQ_Authority	0.035	−0.029	0.031	0.260	−0.347	0.365	0.159	0.614		
MFQ_Purity	0.230	−0.136	−0.216	−0.084	−0.425	0.536	0.539	0.434	0.260	

Note: Please see the supplement 3 for a description of each term.

Table S3-4. Results of the probit regression analysis of H2

	Model 1		Model 2	Model 3		Model 4	Model 5	
(Intercept)	1.342	***	2.187	0.659		0.773	8.306	*
Last dummy	−0.644		−0.647	−0.706		−0.634	−0.779	
Gender	−0.469						−0.013	
Dark_Machiavellianism			−0.119				−0.182	
Dark_Psychopathy			−0.135				−0.419	*
Dark_Narcissism			0.123				0.147	
Lie_accept						0.007		
MFQ_Harm				−0.046			−0.203	*
MFQ_Loyalty				0.162	**		0.150	
MFQ_Authority				−0.046			−0.055	
MFQ_Purity				−0.013			0.019	
Obs.	44		44	44		44	44	
AIC	49.867		51.151	51.190		50.965	53.496	

Note: Please see Supplement 3 for a description of each term. In the analysis, MFQ_Fairness term and Lie_accept term, which had high correlation coefficients, were excluded. **, $p < 0.05$. *. $p < 0.10$.