Journal of Economic Psychology 73 (2019) 136–151

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

The effect of pledges on the distribution of lying behavior: An online experiment

Franziska Heinicke[a,*], Stephanie Rosenkranz[a], Utz Weitzel[a,b]

[a] Utrecht University School of Economics, 3584 EC Utrecht, the Netherlands
[b] Radboud University Nijmegen, Institute for Management Research, Department of Economics, 6525 AJ Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Lying
Truth pledge
Extreme liars
Loss aversion
Online experiment

JEL:
D03
D63
C90

PsycINFO:
2360
2260

ABSTRACT

Reminding people to behave honestly or asking them to actively commit to honest behavior is an easily implementable intervention to reduce dishonesty. Earlier research has shown that such truth pledges affect lying behavior on a group level. In this study we are analyzing how a truth pledge changes the distribution of lying types which have been established in the literature, i.e. truth tellers, partial liars and extreme liars, to better understand whether truth pledges can affect the decision to lie or merely the extent of lies. For this purpose, we conduct a 2 × 2 experiment with 484 participants in which we apply a truth pledge in a gain and a loss frame. We introduce a novel "Even-Odd task" for online lying experiments, which is based on the well-established coin-toss design. The Even-Odd task takes into account that unbiased, physical randomization devices are not always available in online settings, which can be a problem for truth-tellers if they are bad mental randomizers. We therefore ask participants to think of privately known numbers (house numbers, phone numbers) and then determine randomly whether even or uneven numbers result in the higher payment. We find that the truth pledge significantly reduces lying but also that this effect is strongest for extreme liars. The uneven shift in the distribution of liars suggests that truth pledges are effective in decreasing the size of lies but not the number of lies told. This result is robust for both frames.

1. Introduction

In many economic interactions one party possesses private information and lying about this information can generate additional benefits. A financial adviser can hide his own financial interests when recommending his company's funds, or a lawyer can overstate the chances of winning a trial to get a contract with a client. In many of these interactions, it is socially-optimal that private information is revealed truthfully and society has an interest in discouraging lying altogether. Since dishonest behavior is often not directly observable, one convenient intervention to implement, which does not require knowledge of the private information, is to remind people to act honestly or to make them actively commit to honest behavior. Examples of the application of such truth pledges are oath taking in the Anglo-Saxon legal system (for an overview on the development of this concept see Willen, 1983), academic honor codes (McCabe & Trevino, 1993, 1997), and the Dutch bankers' oath (NVB, 2018).

The application of these interventions are supported by a growing literature on honesty preferences. While in conventional economic theory people lie whenever the monetary gain from lying outweighs the monetary costs involved with behaving

* Corresponding author at: Utrecht University School of Economics, Kriekenpitplein 21, 3584 EC Utrecht, the Netherlands and University of Mannheim, L7 3-5, 68161 Mannheim, Germany.

E-mail addresses: f.heinicke@uni-mannheim.de (F. Heinicke), s.rosenkranz@uu.nl (S. Rosenkranz), u.weitzel@uu.nl (U. Weitzel).

https://doi.org/10.1016/j.joep.2019.05.006
Received 31 July 2018; Received in revised form 12 April 2019; Accepted 27 May 2019
Available online 05 June 2019
0167-4870/ © 2019 Elsevier B.V. All rights reserved.

dishonestly, and the net gain from lying exceeds the payoff from behaving honestly (Becker, 1968), the overwhelmingly robust finding in the recent literature is that people lie to increase their payoff but not in a profit maximizing way (for a review see Jacobsen, Fosgaard, & Pascual-Ezama, 2017; Rosenbaum, Billinger, & Stieglitz, 2014). While people strive to increase their payoff, there seem to be strong preferences for honesty which keep people from fully exploiting situations by lying. These honesty preferences are driven by people's desire to maintain a positive self-image which would be threatened by violating moral values of honesty (Gneezy, Kajackaite, & Sobel, 2018; Mazar, Amir, & Ariely, 2008). From these insights follows that truth pledges are effective in decreasing dishonesty because they increase the threat to the self-image and make maintaining a positive self-image after telling a lie more difficult than in a situation without a pledge. The effectiveness of truth pledges is further supported by experimental findings which show that a truth pledge or an active reminder to answer honestly decreases the average level of over-reporting (Bucciol & Piovesan, 2011; Mazar et al., 2008; Shu, Gino, & Bazerman, 2011).

While earlier research has shown that truth pledges affect reports in lying experiments, it remains unclear what the effect is on the distribution of lying. Significantly lower reports under an honesty reminder or truth pledge could be the result of less people lying or the same amount of liars lying less. Literature on honesty and lying has established that there exists heterogeneity in lying behavior and that people can be grouped into three different groups (Fischbacher & Föllmi-Heusi, 2013; Gneezy, Rockenbach, & Serra-Garcia, 2013). First, every participant sample shows a group of truth tellers who refrain from lying completely and report truthfully. Second, the majority of people seem to fall into the group of partial liars who distort the truth slightly in their favor but do not maximize their payoff. Third, there also exists a small group of extreme liars who lie to the full extent and maximize their payoff as would be expected under maximization of monetary utility.

The question addressed by this study is whether truth pledges affect the extent of lying or merely the size of lies told. In the context of truth tellers, partial liars and extreme liars, this translates to the question of how the different types of liars react to a truth pledge. If this intervention influences the decision to lie, the effect should be a shift of the total distribution which implies a change in the share of truth tellers. If only the size of lies is affected, the share of partial and extreme liars will change. Insights into this question will help understand the potential of truth pledges and design effective interventions to discourage lying in a range of situations.

We test the effect of a truth pledge in a 2×2 treatment design in an experiment with 484 participants. We analyze changes in the distribution of reports under a truth pledge as compared to a base treatment for two frames - a gain frame and a loss frame. The two frames allow us to test how robust observed behavior is to different initial levels of lying since loss frames have been shown to increase lying (Grolleau, Kocher, & Sutan, 2016; Schindler & Pfattheicher, 2017). We conduct the experiment online to maximize anonymity and to minimize the effect that changes in the believes over others have on the individual answer.[1] For this purpose, we introduce a new task to elicit honesty preferences called the "Even-Odd task". Participants are instructed on how to generate a random number and their payoff depends on whether the generated number is even or odd. The advantage of the Even-Odd task is that it does not require additional tools like the well-established coin-toss or roll-a-die tasks (Fischbacher & Föllmi-Heusi, 2013). Because in an online experiment it cannot be guaranteed that participants have such tools readily available and actually use them, using a tool-dependent task might invite participants to randomize in their head. Human randomization is known to be biased against repetition and to impose high cognitive load (Figurska, Stańczyk, & Kulesza, 2008; Jokar & Mikaili, 2012; Wagenaar, 1972). These drawbacks of human randomization become increasingly problematic when randomization is done over several periods. By introducing the Even-Odd task we aim to circumvent these issues while staying as close as possible to the structure of established tasks.

Our results of applying the Even-Odd task in four between-subject treatments over 10 rounds show that participants claim to have earned a high payoff significantly more frequently than expected under truth-telling. A simple truth pledge administered at the beginning of the experiment reduces the probability of claiming a high payoff significantly. Loss framing has no significant effect on the overall level of lying. When considering changes in the distribution of lying types, we find that the truth pledge causes a shift from extreme liars to partial liars but not from partial liars to truth tellers under both frames. This implies that this manipulation mainly affects lying in the extreme but does not discourage lying altogether. Loss framing causes no additional difference in the distribution of types in our sample.

The remainder of this paper is organized as follows. In the next section we will discuss related literature on truth pledges and loss framing. This is followed by a detailed description of the experimental design which introduces a new task to elicit honesty preferences. Finally, we will report results and discuss the implications of our findings.

2. Literature and hypotheses

In this study we apply a truth pledge which asks people to promise to tell the truth during an experimental task. Such an intervention has been shown to significantly reduce over-reporting in lying experiments (Bryan, Adams, & Monin, 2013; Bucciol & Piovesan, 2011; Mazar et al., 2008; Shu et al., 2011). Mazar et al. (2008) conduct a cheating experiment where participants self-report the outcome of a calculation task. They find that reminding participants of the honor code of their university, which focuses on honest behavior, causes participants to make reports which are not significantly different to a control group which could not cheat. Similarly, Shu et al. (2011) find that making people sign an honor code significantly reduces overstating of participants' performance in a math task. Bucciol and Piovesan (2011) conducted a coin toss experiment with children. Their experimental design does not include an explicit truth pledge but children in the treatment group were told not to lie. This request to answer truthfully significantly

[1] Moreover, as we are interested to distinguish different types of liars, we need a larger number of lies to start with. An online environment seems most conducive for this purpose, and at the same time minimizes possible experimenter effects with respect to the effect of a truth pledge.

reduced the number of favorable outcomes reported but the effect was more prevalent for girls. Similarly, Bryan et al. (2013) show that participants in a mind-game are less likely to cheat if they were told not to cheat.

The literature discusses two possible explanations of why truth pledges are effective in reducing dishonest behavior. Both explanations start from the insight that people prefer to consider themselves and to present themselves to others as moral and honest decision makers (Aquino et al., 2002; Johansson-Stenman & Martinsson, 2006). Any action that contradicts this image is costly because it negatively affects the self-image. Thus, lying is a threat to the self-image. When people tell a lie to secure benefits for themselves, they face a trade off between the involved gains and maintaining a positive self-image (Gneezy et al., 2018; Mazar et al., 2008).

The first approach to explain the honesty-increasing effect of a truth pledge is that this intervention could hinder the ease with which lies can be justified. The ease with which people can justify their behavior is an important factor in the decision to lie (Shalvi, Dana, Handgraaf, & De Dreu, 2011). A justified action seems less immoral and allows the self-image to remain unaffected (Jacobsen et al., 2017). Thereby, justification strategies allow people to lie even if they have a strong moral value for honesty. Truth pledges negatively affect the ease with which a lie can be justified by explicitly defining wrong-doing (McCabe & Trevino, 1993, 1997). Shalvi, Gino, Barkan, and Ayal (2015) argue that if a situation is ambiguous or rules are not fully defined, people can justify their actions by strategically perceiving the rules in their favor. Thus, without a truth pledge people can justify a lie by considering the specific setting outside the realm in which honesty is morally required. Shu et al. (2011) further find that participants who cheated in an experiment remembered fewer rules of an honor code and the authors interpret this behavior as strategic forgetting where people fail to recall moral rules in situations of an ethical dilemma. Explicit truth pledges might interfere with such a justification strategy because the moral rule is actively recalled right before the action is taken.

The second explanations focuses on priming and suggests that truth pledges prime people on their moral identity (Mazar et al., 2008). When people commit to a truth pledge before the decision to lie, attention to their own moral standards is increased which makes these standards more accessible. According to the theory of objective self-awareness, which is related to this argument, actions have a greater impact on the self-image if people are more self-aware when taking the action (Diener & Wallbom, 1976; Duval & Wicklund, 1972). Thus, priming people on their moral identity increases the negative effect that a lie has on the self-image and in the trade off between gains and costs of lying the cost side increases.

This paper does not aim to distinguish between the two approaches to explain the effect of truth pledges but focuses on a common feature of both. Both explanations conclude that truth pledges make it more difficult to maintain a positive self-image while telling a lie. Regardless of whether personal moral standards are made salient or justifications strategies are restrained, people will have more difficulty maintaining a positive attitude towards themselves when telling a lie after having made a truth pledge.

While earlier research has shown that truth pledges can decrease lying at the group level, it remains unclear how they affect the distribution of lying types. The heterogeneity in behavior between the three types can be explained by differences in the costs of lying. Abeler, Nosenzo, and Raymond, 2018 as well as Gneezy et al. (2018) suggest that people incur a cost of lying that is a function of the difference between their observation and their report, as well as of the probability that a given report is interpreted as a lie. Differences between types come from differences in intrinsic costs for acting immorally and image concerns for being seen by the self or others as a liar. While the monetary gain of any positive deviation between report and observation is linearly increasing, costs are likely to be convex, as the intrinsic costs are increasing and the probability that a given report is interpreted as a lie is increasing overproportionately. Truth tellers refrain from lying because the costs caused by violating values of honesty outweigh the monetary gain of the lie, while for extreme liars costs from lying are low enough to make extreme lying the utility maximizing option. Partial liars confront the costs of lying with the monetary gain in order to maintain a positive self-image (Gneezy et al., 2018; Mazar et al., 2008). While partial liars engage in immoral behavior, they do not take full advantage of the situation and only increase their own gain moderately. Thus, partial lies present a compromise between increasing monetary gains and appearing honest before oneself and others by reporting less monetarily rewarding outcomes which are more credible (Fischbacher & Föllmi-Heusi, 2013).

In such an interpretation, differences in the level of lying need to be related to differences in the cost function of lying, where extreme liars most likely incur lower marginal costs, and possibly even lower absolute levels of costs. An intervention which increases the self-image costs of lying will then increase not only absolute costs for all types but, because of this, also increase the probability that a given report is interpreted as a lie. Hence, because of the convexity of the cost function their effect is larger on the extreme liars than on partial liars. This is due to the fact that the lower slope of the cost function of extreme liars causes greater changes following an upward shift of the cost function as a whole. Based on these arguments we formulate our first hypothesis as follows:

Hypothesis 1. A truth pledge causes a greater shift from extreme liars to partial liars than from partial liars to truth tellers.

This study tests the effect of truth pledges in two frames, a gain frame and a loss frame. Earlier research on framing and unethical behavior suggests that people are more likely to engage in unethical behavior to avoid losses than to secure gains. As an early example, Newberry, Reckers, and Wyndelts (1993) show that tax practitioners are more willing to sign off on questionable tax statements in order to maintain a client relationship than in order to win a new client. Schweitzer, Ordóñez, and Douma (2004) show that people who have set but not yet met specific goals engage in more unethical behavior than people without goals. The relationship between framing and unethical behavior has also been specifically explored for dishonesty. Cameron, Miller, and Monin (2008) and Grolleau et al. (2016) applied gain and loss frames to self-grading tasks and find that participants are more likely to overstate their performance if the payoff is presented in a loss frame. This result is supported by Schindler and Pfattheicher (2017) who analyze the effect of a loss frame in the commonly used coin toss and die under the cup task and find lower average reports for both tasks under a loss frame as compared to a gain frame. Further adding to the evidence that people lie to avoid losses, Shalvi (2012) finds that people lie in order to turn negative gambles into positive gambles and Balasubramanian, Bennett, and Pierce (2017) show that participants in an online labor market lie more if they have not yet reached their average earnings. In addition to these individual tasks, Kern and Chugh (2009) find that participants in a negotiation experiment are more prone to misrepresent information in a loss frame than in a gain frame. As an exception to the studies cited here, Charness, Blanco-Jimenez, Ezquerra, and Rodriguez-Lara (2018) report no significant difference between a gain and a loss frame in a die roll experiment with a ten-sided die.

The suggested channel through which loss framing enables dishonest behavior is that it installs a feeling of deservingness in people (Cameron et al., 2008). Being put in a situation where they might lose something, people feel entitled not to lose and thus, they are better able to justify their unethical behavior. This notion is supported by the finding of Cameron et al. (2008) that participants in a loss frame condition reported a higher amount of money when asked how much they deserve to be paid for their participation in the experiment than participants in the gain frame condition. The feeling of deservingness allows people to rephrase the lie as merely protecting what is deserved, which reduces the self-image effect of a lie and thus the costs of lying.

In line with the arguments presented above, it is reasonable to assume that the loss frame thus shifts the cost functions of all different types of liars downwards. Note that a truth pledge intervention has the opposite effect. Adding a truth pledge in a situation with lower cost levels, however, reduces the effect of a pledge for all types due to the convexity of cost functions. As in the gain frame, the shape of the cost function causes the change to be most prominent for extreme liars. Hence, we formulate our second hypothesis as follows:

Hypothesis 2. The shift from extreme liars to partial liars caused by a truth pledge is lower under loss framing than under gain framing.

3. Design and procedure

3.1. The Even-Odd task

For this study we apply a novel design to elicit lying behavior which requires no additional tools to generate a random outcome. In the Even-Odd task participants first receive instructions to think of familiar numbers (see the appendix for the instructions). The familiar numbers include birthday and age of relatives and friends, house numbers, digits of telephone numbers or the number of letters in a word. Second, participants are asked to sum up the numbers they just thought of (see Table 10 for all questions).

This generating process is intended to give participants as little as possible control over the final number that is calculated and to avoid a personal connection participants might have to a certain number. Even if the originally number that participants had in mind (e.g. a birthday) holds a strong personal connection, the final sum will be disconnected from this. Similarly, even if participants think about familiar numbers with the payoff in mind, it is more demanding to recall familiar numbers that also add up to a desired number.[2]

At the end of the number generating process participants are asked to keep in mind whether the generated number was even or odd. On a new screen participants are informed on whether an even or odd number leads to a higher payoff and, on the same screen, participants then have to report if they had calculated an even or an odd number. Whether even or odd numbers result in a higher payoff is determined randomly for each participant. Therefore, during the number generating process participants are not aware whether their number will yield a higher payoff but at the moment of reporting their number they are fully aware of the payoffs related to even and odd numbers and thus, the monetary consequences of their report.

The advantage of the Even-Odd task is that it requires no additional tools to generate a random outcome. This makes it appropriate for online experiments where the availability and use of a tool cannot be guaranteed. We consider the online context beneficial to the present study because we are interested in changes in the distribution of liars. For this purpose we need to ensure a positive amount of lying in our sample and earlier research has shown that lying increases with anonymity which is greater in an online setting than in an laboratory experiment (Conrads & Lotz, 2015; Naquin, Kurtzberg, & Belkin, 2010). While the high levels of lying in an online setting are an advantage to our research question, we do not expect this to affect the treatment effect as earlier research has shown that online and laboratory experiments yield consistent results for various different treatments (Anderhub, Müller, & Schmidt, 2001; Birnbaum, 1999; Goodman, Cryder, & Cheema, 2013; Gosling, Vazire, Srivastava, & John, 2004; Horton, Rand, & Zeckhauser, 2011). Specifically in the context of dishonesty, Bryan et al. (2013) find that their treatment effect is qualitatively the same in both, online and offline, settings.[3]

Given the increasing popularity of online experiments, the Even-Odd task presents an important extension to existing lying tasks. The economic literature on lying includes two tool-dependent tasks to elicit honesty preferences in isolation without other interfering preferences which are widely used to study lying behavior. In coin toss experiments participants are asked to flip a coin and to report the side that landed on top (e.g. Cohn, Maréchal, & Noll, 2014; Cohn, Fehr, & Maréchal, 2014; Houser, Vetter, & Winter, 2012). The payoff of participants depends on the side of the coin that participants reported to lie on top and as the toss is done privately, participants can misreport the outcome. A similar approach is taken in die roll tasks where the random outcome is generated by a

[2] A robustness check shows that the type of question does not affect lying (see Section 4.1).

[3] Bryan et al. (2013) applied a word manipulation in the study instructions of a lying task. In contrast to our study, the authors use the classical mind game setup, which, as explained further below, may distort results when lying is elicited over several rounds. Bryan et al. (2013) administered their task in a one shot setting, which eliminates the risk for distortion because of bad mental randomization. Hence, the correspondence of online and offline treatment effects found in Bryan et al. (2013) can, in principle, also apply to our setting.

privately rolled die and the payoff of participants is determined by the number the die is showing (Fischbacher & Föllmi-Heusi, 2013).

Both methods provide certain challenges when conducting them online. As participants are all by themselves at the time of the experiment and not under supervision of an experimenter, it cannot be ensured that they indeed make use of the randomization tool. If no coin or die is easily available, participants could resort to deciding on a possible outcome in their mind. In a one-shot decision this is less problematic but over ten rounds the attempt of participants to randomize by themselves can lead to considerable distortions. First, people are known to be biased randomizers who tend to report a deficit of repeats of the same number (Bains, 2008; Figurska et al., 2008; Jokar & Mikaili, 2012; Wagenaar, 1972). Second, randomizing in the head also imposes considerable cognitive load on participants because when generating a random sequence humans try to make the sequence appear random and thus, recall all prior draws before making another one (Bains, 2008; Jokar & Mikaili, 2012). Imposing cognitive load over the course of the whole experiment has been shown to influence lying behavior (Shalvi, Eldar, & Bereby-Meyer, 2012).

Other proposed tasks to elicit honesty preferences are mind-games, in which participants are ask to make a decision in their mind which will determine their payoff (Jiang, 2013), and self-grading tasks which give participants the option to cheat on the outcome of a real-effort task by letting them grade their own work in private (Mazar et al., 2008). While the former bears the same challenges concerning randomization as explained above, the latter does not elicit honesty preferences in isolation but in connection to own effort. Houser et al. (2012) show that frustration in a preceding task increases lying in a subsequent lying experiment.

3.2. Implementation and procedure

For this study, we let participants play ten rounds of the Even-Odd task with ten different number generating questions which are shown in Table 10 in the Appendix. Of the ten rounds, one was randomly selected for payoff at the end of the experiment. At the end of each round, participants were informed about what their payoff would be if that round was selected for the final payoff. Hence, participants could maximize their payoff by maximizing the probability of receiving the higher amount and thus, by reporting the favorable outcome every round. We applied the Even-Odd task to a 2 × 2 experimental design which included a gain frame with and without a truth pledge, PLEDGE_gain and BASE_gain, and a loss frame with and without a pledge, PLEDGE_loss and BASE_loss. For the instructions, please see the Appendix.

Treatment BASE_gain was structured as the standard lying task: participants can earn additional money (on top of their show-up fee) for every reported favorable outcome. Participants received an initial budget of €1.50. In case of a winning number participants received €1.50 extra (see Supplementary Fig. 4). Hence, winning numbers were framed as a gain on top of the endowment, which was also reiterated on the payment screen (see Supplementary Fig. 6).

Treatment PLEDGE_gain was identical to BASE_gain with the only difference that, after the welcome screen, we showed participants the following sentence: "By continuing from this page with the "NEXT"-button, I agree to tell the truth during this experiment". To make this pledge very salient we (i) presented the sentence on a separate screen, (ii) presented the last part of the sentence in bold face, (iii) titled the screen with "truth telling", and (iv) also linked the statement of telling the truth to a particular action, namely pressing the "Next"-button in order to continue (see Supplementary Fig. 2). This pledge was only shown once at the beginning of the experiment. The remaining part of the experiment was identical in the pledge and no pledge treatments.

Treatment BASE_loss was identical to treatment BASE_gain with the exception that payments were displayed with a loss frame: in the loss frame, participants were informed in each round that they have an initial budget of €3. In case of a winning number they could keep this amount. Otherwise, €1.50 were deducted from their initial budget (see Supplementary Fig. 5). Hence, losing numbers were framed as a loss that was deducted from the endowment, which was also reiterated on the payment screen (see Supplementary Fig. 7).[4] Prior literature has shown that this framing generates treatment effects that are in line with interpreting the initial budget as a reference point, i.e., deductions from initial endowments are perceived as losses, while receiving extra earnings on low endowments are perceived as gains (e.g. Hannan, Hoffman, & Moser, 2005; Sonnemans, Schram, & Offerman, 1998). Cameron et al. (2008) use this way of framing with regards to lying and found significant framing effects.

Treatment PLEDGE_loss was identical to treatment BASE_loss with the only difference that, after the welcome screen, participants saw an additional screen with the truth pledge and had to click "next" in order to proceed. This was exactly the same screen as in PLEDGE_gain (see Supplementary Fig. 2).

After completing the ten rounds, participants filled out a questionnaire on demographics which have previously been identified as affecting lying behavior. Namely, we elicit gender (Friesen & Gangadharan, 2012; Jacobsen et al., 2017; Rosenbaum et al., 2014), age (Friesen & Gangadharan, 2013; Fosgaard, in press) and the field of study (López-Pérez & Spiegelman, 2012). At the end of the experiment, participants decided whether they wanted to receive their payment via a bank transfer or via the online payment provider Skrill. The delay in payment was the same in both methods and participants were paid within two days after the experiment. The experimental study was conducted with students recruited from the subject base of the online experimental platform GlobaleXperimentalPlatform (GXP) and implemented with the open-source and online software oTree (Chen, Schonger, & Wickens, 2016). For a period of four months all newly recruited participants received an invitation to this task as their first experiment on the GXP platform. In total nine sessions were scheduled and each session was available for a period of several hours to give participants the opportunity to participate at a time that was convenient to them.

[4] Note that the total payoff was identical under the gain and the loss frame. In both cases, reporting a winning (losing) number generated €3 (€1.50).

Participants where able to open the experiment via their personal profile on GXP. The software assigned participants in an alternating manner to each treatment. In total 484 students participated in the experiment of which 88% finished all ten rounds. The following analysis is limited to those 432 participants who finished all ten rounds of the experiment and the subsequent survey. Participants who did not finish all rounds did not receive any payoff and we exclude their data to avoid confounding effects from participants who were not incentivized. The experiment took approximately 10 min and, on average, participants earned €2.47 (equivalent to €14.82 per hour). The sample included 70% Dutch students and 59% female students with an average age of 21.3 years. The students came from a range of study programs with the largest share being enrolled in an economics or business program (34%).

3.3. Estimation strategy

By applying the Even-Odd task over ten rounds we can identify different types of liars and changes in the distribution of types. Extreme liars are identified by having reported 10 favorable outcomes which maximizes the probability of earning €3.00 in the experiment. Partial liars increase the probability of receiving the higher amount by misreporting some answers but not all ten rounds. This group will have reported six to nine favorable outcomes. Truth tellers will be identified by having reported five or less favorable outcomes. We define a categorical variable with these three types of liars as our primary dependent variable. In order to test our hypotheses we will analyze the probability of being an extreme liar or a truth teller as compared to being a partial liar in a multinomial logit regression with partial liars as base. We evaluate a model with the following specification:

$$Y_j = \beta_0 + \beta_1 Pledge_i + \beta_2 Loss_i + \beta_3 Pledge_i * Loss_i + \beta_4 C_i + \epsilon$$

where $Y_j = Ln\frac{Pr(y=j)}{Pr(y = \text{partial liar})}$ represents the probability that individual j is a truth teller, or extreme liar respectively, $Pledge_i$ is a dummy for the treatments PLEDGE_gain and PLEDGE_loss, $Loss_i$ is a dummy for treatments PLEDGE_loss and BASE_loss, $Pledge_i * Loss_i$ is the interaction of the two dummies, β_0, β_1, β_3, β_4 are parameters to be estimated, and ϵ are unobserved factors. In addition, we control for the possibility of non-random assignment of treatments to subjects by including a vector of subject-level control variables C_i: gender, age, and field of study. For testing Hypothesis 1 we estimate the above specification without interaction term, expecting a negative and statistically significant coefficient of $Pledge_i$ for extreme liars but not for partial liars. For testing Hypothesis 2 we estimate the specification as stated above, expecting a positive and statistically significant coefficient of $Pledge_i * Loss_i$ for extreme liars but not for partial liars, which indicates that the negative base effect of a pledge on extreme liars is smaller in the loss frame than in the gain frame.

Using three distinct types of liars has the advantage of clear cut definitions, but suppresses some information about the individual extent of lying within and between categories. For example, although partial liars reported different numbers of favorable outcomes, they all enter the multinomial logit equally. Also, the difference in the extent of lying between partial liars and, e.g., extreme liars, is not accounted for. Following Abeler et al. (2018), we therefore also compute a continuous measure of lying for participants as a secondary dependent variable. For this we subtract, per treatment and per reported outcome category, the true binomial pdf from the actually observed pdf. For example, if we observe in BASE_gain an actual probability for participants to report 10 favorable outcomes ($Prob_{actual} = 8.16\%$), then we subtract the true probability for this specific outcome ($Prob_{true} = 0.01\%$). The result is an indicator for potential lying for participants who report a specific outcome in a specific treatment (also see Fig. 2 in Section 4.2). For our analyses, we normalize potential lying with the actually observed probability.[5] We refer to this ratio as *propensity to lie*, which we compute for each participant as $(Prob_{actual} - Prob_{true})/Prob_{actual}$, with a propensity to lie of zero for truth tellers.

For our econometric estimations we use the 90% and 50% quantile (Q90, Q50) of the propensity to lie as dependent variable in simultaneous-quantile regressions with 500 bootstrap replications. Specifically, we simultaneously estimate the econometric specification shown above with $Y_j = Q50$ and with $Y_j = Q90$. An inspection of the data shows that the 90% quantile of the propensity to lie corresponds with 9 favorable outcomes and therefore lies around the cut point to extreme lying. The 50% quantile corresponds to 6 favorable outcomes and lies around the cut point to truth telling.[6] To test Hypothesis 1 we estimate the above specification for Q50 and Q90 without interaction term, expecting a pledge effect on Q90 but not on Q50. For Hypothesis 2, we apply the same specification to the gain frame and loss frame subsamples, expecting a pledge effect on Q90 (not on Q50) in the gain frame, but not in the loss frame.

4. Results

4.1. General level of lying

Before we test our hypotheses in the next Section 4.2, we first provide some background analyses by considering the effect of truth pledges and loss frames on the general level of lying. The Even-Odd task revealed significant lying across all treatments as shown by the results summarized in Table 1. In all four treatments, participants reported, on average, between 6.25 and 6.69 favorable outcomes. The medians of the number of rounds for which a participant declared a favorable outcome are all significantly higher than 5

[5] In our example this would be: $8.15/8.16 = 0.99 \in [0, 1]$ for reporting 10 favorable outcomes in BASE_gain.

[6] We also run robustness checks with other quantile combinations involving Q45 and Q40 (corresponding to 6 and 5 favorable outcomes, respectively). The results do not change qualitatively.

Table 1
Summary statistics for all four treatments.

	BASE_gain	PLEDGE_gain	BASE_loss	PLEDGE_loss
Total participants (N)	98	106	116	112
Mean favorable outcomes/participant	6.6939	6.3396	6.6034	6.25
	(1.9126)	(1.756)	(2.0211)	(1.7732)
Misreported answers	34%	27%	32%	25%
Estimated mean percent lying (EV_6)	60%	51%	51%	45%
Lower bound of 99% CI (for EV_6)	51%	40%	41%	33%
Upper bound of 99% CI (for EV_6)	67%	60%	59%	54%
Female	52%	63%	61%	60%
Age	20.45	21.66	21.67	21.37
	(2.1999)	(4.0307)	(3.438)	(4.2085)
Economics	31%	44%	27%	35%

Note: Standard deviation in parentheses.

(p-values $<$ 0.0001, two-sided signed-rank test).[7] Following the procedure of Houser et al. (2012) we can also calculate the share of misreported answers from all answers given. Under the assumption that participants only misreport to their advantage, the share of reported high payoffs h is given by:

$$h = m * 1 + (1 - m) * 0.5 \tag{1}$$

where m is the ratio of misreported answers which means that participants observed a number leading to a low payoff but claimed a high payoff instead. Thus, the share of misreported answers is the share of lying: If people lie they report the high payoff with probability 1 and if they are truthful they get the high payoff with probability 0.5. Our treatments show between 34% and 25% of misreported answers which is comparable to results from the literature (Hugh-Jones, 2016; Rosenbaum et al., 2014). Following Garbarino, Slonim, and Villeval (2017) we also estimate the mean percent of individuals who lie when reporting 6 or more favorable outcomes (EV_6), conditional on having received 5 or less favorable outcomes. In contrast to the computation of misreported answers, the estimation of EV_6 takes the distribution of reports of all subjects into account. This allows for the calculation of confidence intervals, which we report in Table 1. The estimated mean percent of all individuals lying by reporting 6 or more winning outcomes varies between 45 and 60 percent, depending on the treatment. This share is statistically significant and clearly above zero. With 99% confidence we can expect to have at least 33% to 51% dishonest subjects in our treatments (see lower bound of CI for EV_6 in Table 1).[8] A first indication from this analysis is that participants lied significantly in all treatments to increase their payoff.

Pairwise comparisons of the distributions in the four treatments show no significant differences in the general level of lying for either applying a pledge or a loss frame. The pledge does not lead to a significant decrease of favorable reports in the gain frame (KS p = 0.563, two-sided Mann-Whitney test (MW) z = 1.290, p = 0.1972) or in the loss frame (KS p = 0.618, MW z = 1.235, p = 0.2167). Loss and gain frame cause no significant difference in reporting without a pledge (KS p = 0.995, MW z = 0.438, p = 0.6615) or with a pledge (KS p = 1.000, MW z = 0.439, p = 0.661). Finally, also the estimated mean percent of dishonest participants (EV_6) does not differ between treatments as all confidence intervals for all treatments overlap with each other.

These tests, however, do not take into account the distribution of demographic variables across treatments. As seen in Table 1, the treatments show differences in the variables of gender, age and field of study. For gender and age the differences are not significant (respectively, Pearson $\chi^2(3) = 2.7418$, p = 0.433 and Kruskal-Wallis $\chi^2(3) =$ 5.199, p = 0.1513).[9] The proportion of economics students differed significantly (Pearson $\chi^2(1) = 5.011$, p = 0.025) between treatments including the pledge (39% economics students) and those not including a pledge (29% economics students).

Therefore, we further analyze the effect of the pledge in linear regression models where we can control for demographic variables. Table 2 reports the results of OLS estimations with the reported number of favorable outcomes per participant as the dependent variable.[10] The independent variables of interest are dummies for treatments with a "Loss frame", for treatments with a "Pledge", and the interaction thereof. Control variables are dummy variables for gender (Female), a count variable for "Age" (in years), and students with economics as a major (Economics).[11]

[7] Moreover, in all treatments, the distribution of reported favorable outcomes is significantly different from a binomial distribution with 10 draws and a success probability of 0.5 which would have been expected under truth-telling (p-values $<$ 0.001, Kolmogorov-Smirnov test (KS)).

[8] For the computation, we use the total number of participants (N), the number of participants who report 6 or more favorable outcomes (R_6), the cumulative probability of experiencing 5 or fewer winning outcomes (P_5) in a binomial distribution (10, 0.5), and the confidence interval *CI*. For illustrative purposes, in BASE_gain, the values $N = 98$, $R_6 = 74$, $P_5 = 62\%$, and $CI = 99$, can be used to compute EV_6 (see http://lyingcalculator.gate.cnrs.fr/).

[9] The results reported here refer to the comparison across all four treatments. Differences between the pledge and no pledge treatments or between gain and loss frame are not significant either.

[10] Differences in the sample sizes are due to four participants completing all ten rounds but not the demographics questionnaire.

[11] Other fields of study were balanced across treatments (Social science: Pearson $\chi^2(3) = 3.8463$, p = 0.279, natural science: Pearson $\chi^2(3) = 1.5091$, p = 0.680, humanities: Pearson $\chi^2(3) = 2.3119$, p = 0.510, other fields: Pearson $\chi^2(3) = 1.1994$, p = 0.753). Robustness checks controlling for other fields of studies are included in the Appendix (Table 6).

Table 2
OLS regression analysis of treatment effects.

Dependent variable: reported number of favorable outcomes			
	Model 1a	Model 1b	Model 1c
Loss frame	−0.0900	−0.0268	−0.0744
	(0.1796)	(0.1763)	(0.2620)
Pledge	−0.3538	−0.4917**	−0.5417*
	(0.1799)	(0.1777)	(0.2632)
Loss*Pledge			0.0929
			(0.3547)
Female		0.2333	0.2372
		(0.1861)	(0.1847)
Age		−0.0026	−0.0019
		(0.0222)	(0.0225)
Economics		1.0570***	1.0591***
		(0.1965)	(0.1969)
Constant	6.6937***	6.2951***	6.3032***
	(0.1629)	(0.5200)	(0.5223)
N	432	428	428
R2	0.0094	0.0749	0.0751

Note: Robust standard errors in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Model 1a shows no significant effect for the pledge or the loss frame which confirms the initial pairwise comparison made above. However, when controlling for demographic variables that have been shown to affect lying behavior (Rosenbaum et al., 2014), the pledge significantly reduces the number of favorable outcomes which a participant reports (see Model 1b). This indicates that a truth pledge reduces average lying in our experiment which is in line with earlier research on truth pledges.

The loss frame does not show any significant effect on the amount of favorable outcomes. Recall, that Hypothesis 2 predicts differences in the effect of a truth pledge between gain and loss framing, *conditional* on loss framing having an effect on lying. Given that this condition is not satisfied, we can interpret the statistically insignificant coefficients of loss framing in Table 2 as a first indication that Hypothesis 2 may not be supported. Moreover, if we assume no effect of loss framing on lying, we would, based on our argumentation, conclude that the pledge should have the same effect under gain and loss framing. This is indeed what we observe in Model 1c, where the coefficient of the interaction term between the pledge and the loss frame is not significant.

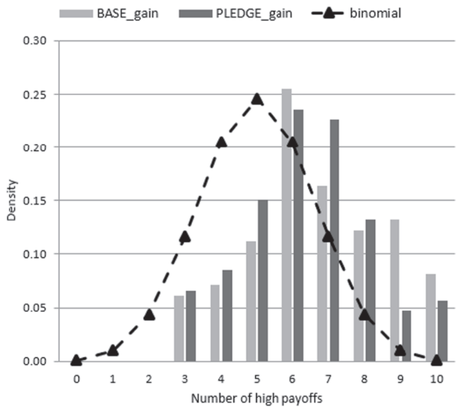
With regard to the control variables, only one demographic variable, i.e. being in an economics or business program, significantly increases the number of favorable outcomes reported. This is in line with earlier literature that found economics students to lie more than students of other programs (López-Pérez & Spiegelman, 2012) and thus, omitting this variable biased the effect of the pledge upwards in Model 1a.[12]

As a robustness check we analyze whether the pledge effect, which was administered at the start of the experiment, is stable across all rounds or "wears off" over time. For this, we estimate a random effects probit model with all individual answers of all participants (panel variable) over all periods (time variable). The dependent variable indicates, for each round, whether a participant reported a favorable outcome (1) or not (0). The independent variable of interest is a count variable, *Round*, which captures a possible trend over rounds. The results in Table 8 show that lying does not differ significantly between earlier and later rounds, i.e. there is no trend in lying behavior over time. This also applies to the interaction of *Pledge* with *Round*. Effron, Bryan, and Murnighan (2015) show that in a repeated cheating task participants show no trend in cheating over rounds but cheat significantly more in the last round. Thus, we further analyze whether lying behavior in round 10 differs from the other rounds. The results are included in Table 8. In the last round the favorable outcome is reported significantly more often. This difference is driven by the pledge treatment because the pledge reduces lying in the other rounds but has no significant effect in the last round.[13]

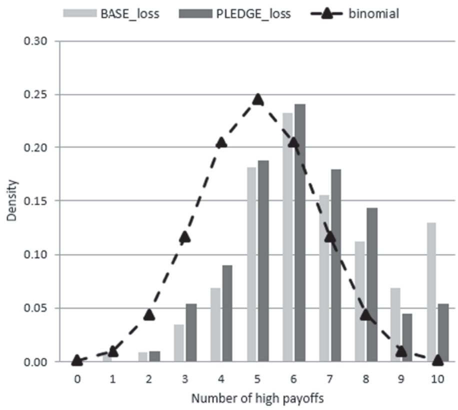
A final robustness check pertains to the type of questions involved. The Even-Odd task asks participants to calculate a random

[12] All reported results are qualitatively equivalent to the results of a Tobit regression which accounts for ceiling effects: participants who maximized their payoff reported a favorable outcome in ten rounds but they might have lied more often if they could have increased their payoff by doing so. Results of the Tobit regressions, including different fields of study, are reported in the Appendix (Table 6).

[13] Different to Effron et al. (2015) where participants were paid for all rounds, participants in this study were only paid for one randomly selected round which likely accounts for the limited support we find for the last round effect. Effron et al. (2015) argue that participants face a trade off between the guilt caused by lying and the regret caused by giving up monetary payoff and that regret is stronger in the last period where no further money can be earned. If only one round is selected for payoff, the amount of money that can still be earned is not decreasing over rounds but kept constant.



(a) Distribution of number of favorable outcomes reported for truth pledge in gain frame



(b) Distribution of number of favorable outcomes reported for truth pledge in loss frame

Fig. 1. Distribution of reported favorable outcomes.

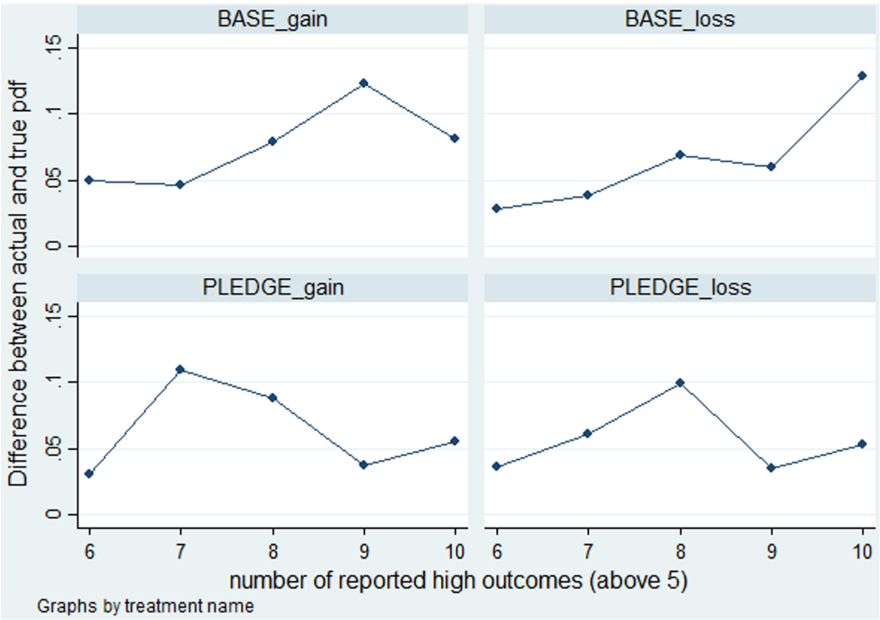


Fig. 2. Distribution of potential lying (actual minus true pdf).

number based on personal information.[14] This has the advantage that participant do not have to randomize in their heads. However, relying on personal information and invoking memories of relatives and friends might prime participants on honest behavior.[15] As the questions vary in the degree to which they rely on personal information some questions might have involved biased behavior. As a robustness check, we therefore run pairwise comparisons of all ten questions (across all treatments) and test for the equality of proportions in reported favorable outcomes. Table 7 (in the Appendix) shows the overall share of favorable outcomes for each question and provides pairwise comparisons between all questions across all treatments. As this analysis shows no systematic differences due to the question type, we can be confident that the question type did not influence truth-telling.

4.2. Distribution of lying

Both of our hypotheses focus on shifts in the distribution of lying behavior. We therefore start with a graphical representation of the distributions in Fig. 1a, which contrasts the BASE_gain treatment with the PLEDGE_gain treatment. While the pledge causes a

[14] All questions are included in the Appendix.
[15] We thank an anonymous reviewer for drawing our attention to this issue.

Table 3
Share of liar types and confidence intervals for extreme liars.

	Binomial(10, 0.5)	BASE_gain	PLEDGE_gain	BASE_loss	PLEDGE_loss
Truth tellers	62.3%	24.49%	30.19%	30.17%	33.93%
Partial liars	37.6%	67.35%	64.15%	56.90%	60.71%
Extreme liars	0.1%	8.16%	5.66%	12.93%	5.36%
EV_{10}		8.08%	5.57%	12.85%	5.27%
Lower CI 99%		7.23%	4.77%	12.1%	4.48%
Upper CI 99%		8.16%	5.66%	12.93%	5.36%

decrease in 9 and 10 favorable outcomes, it seems to increase reports of 7 and 8 favorable outcomes. The effect of the truth pledge in the loss frame is captured in Fig. 1b where we compare the BASE_loss treatment with the PLEDGE_loss treatment. This comparison shows the same pattern as in Fig. 1a with a decrease in very high reports for the truth pledge but an increase for partial lying. These distributions are a first indication that changes in lying become most apparent in the extremes of the distribution as predicted in Hypothesis 1. Fig. 2 shows this effect by subtracting the true pdf from the actual pdf (dotted line and bars in Fig. 1, respectively).[16] For reasons of exposition we focus on positive values which indicate potential lying. In accordance with Hypothesis 1, the pledge treatments (lower panel of Fig. 2) show lower values for extreme lying (9 or 10 favorable outcomes) than the base treatments (upper panel) and mostly higher values for less extreme lying (6 to 8 favorable outcomes), indicating a shift from extreme to partial lying.

In order to test our hypotheses, as explained in Section 3.3, we consider changes in types of liars: participants who reported 10 favorable outcomes are categorized as *extreme liars*, participants with 6 to 9 favorable outcomes as *partial liars* and participants with less than 6 favorable outcomes as *truth tellers*. Due to the nature of the task we are not able to determine lying on the individual level and each of these groups may also include truth tellers who reached a high number of favorable outcomes without lying. However, the over-reporting of 6 or more favorable outcomes as compared to the binomial distribution indicates that not all these outcomes can result from truth tellers. Thus, the partial liars in the sample will be within the group of 6 to 9 favorable outcomes while extreme liars will have 10 favorable outcomes. Table 3 summarizes the share of each type in the four treatments and shows the expected shares for each group for the binomial distribution which would be expected under truth-telling.[17]

As a first test of Hypothesis 1, we follow Garbarino et al. (2017) and estimate the mean percent and 99% confidence intervals for extreme liars, EV_{10}, i.e. the share of participants who lie in order to receive the maximum payment for 10 favorable outcomes. The results are reported in the lower panel of Table 3. Not surprisingly, as there is only a 0.1% chance for experiencing 10 favorable outcomes, the estimated share of extreme liars is very close to the actual share of participants in this category. Note that our definition of extreme liars (in the upper panel) is identical with the upper bound of the confidence interval of the estimated share of extreme liars, EV_{10}. Hence, we can be 99% confident that our definition of extreme liars captures real lying.

The 99% confidence interval in Table 3 shows a clear and statistically significant pledge effect on extreme lying, both in the gain and in the loss frame.[18] Recall, that we did not find any significant effect of the pledge on the estimated percent of individuals who lie when reporting 6 or more favorable outcomes, EV_6, which mainly includes partial liars (see Table 1 in Section 4.1). Hence, in support of Hypothesis 1, we find that the pledge causes a (greater) shift to less extreme lying, but not to more truth-telling. Moreover, we find a significant shift to more extreme lying due to loss framing when pledges are missing (BASE_loss v BASE_gain). Once participants pledge to report the truth, loss framing does not seem to have any significant influence on extreme liars anymore (PLEDGE_loss v PLEDGE_gain).

For a more comprehensive analysis of Hypothesis 1, we conduct regression analyses which controls for demographic variables. As outlined in Section 3.3, we take two approaches: a multinomial regression with a categorical variable for the three types of liars (extreme liars (EL), partial liars, truth tellers (TT)) as dependent variable and partial liars as the base category and a quantile regression on the propensity to lie. The choice of quantiles in the quantile regression follows from the distribution of reports. The 90% quantile only includes reports of 9 or 10 favorable outcomes in all treatments and below the 50% quantile 6 or less favorable outcomes are reported. Table 4 shows the results. Model 2a follows the analysis above and defines extreme liars as participants who reported 10 favorable outcomes. Model 2b includes 9 and 10 favorable outcomes into the category of extreme liars to test whether our results are sensitive to the exact specification of types. The results of the quantile regression are shown in Model 2c. In support of Hypothesis 1, we find that the truth pledge reduces the probability of being an extreme liar as compared to a partial liar but does not

[16] Also see Section 3.3.

[17] Fischbacher and Föllmi-Heusi (2013) suggest that the actual share of extreme liars can be calculated by assuming that no one who observed 10 favorable outcomes reported less than that. In order to get the share of participants who misreported having 10 favorable outcomes we need to subtract the expected probability of observing 10 favorable outcomes and to take into account participants who would be extreme liars but actually observed 10 favorable outcomes. For the BASE_gain treatment this adjustment yields $\frac{0.0816-0.001}{1-0.001}=0.0807$, for the PLEDGE_gain treatment $\frac{0.0566-0.001}{1-0.001}=0.0557$, for the BASE_loss treatment $\frac{0.1293-0.001}{1-0.001}=0.1284$ and for the PLEDGE_loss treatment $\frac{0.0536-0.001}{1-0.001}=0.0526$. The small deviations from shares reported in Table 3 signify that the group of extreme liars indeed mostly includes participants who misreported some share of the rounds they observed.

[18] This result also holds when we pool the gain and loss frames.

Table 4
Regression analysis Hypothesis 1.

	Model 2a[a,b]		Model 2b[ac]		Model 2c[d]	
	TT	EL	TT	EL	Q50	Q90
Loss frame	0.2147	0.5108	0.174	0.0141	−0.0292	−0.068
	(0.2197)	(0.3900)	(0.2241)	(0.2880)	(0.0512)	(0.0480)
Pledge	0.2581	−1.0355**	0.1624	−0.9709**	−0.0176	-0.1660**
	(0.2218)	(0.4032)	(0.2263)	(0.3020)	(0.0532)	(0.0633)
Female	−0.2024	0.8745*	−0.2209	0.3039	0.048	0.0000
	(0.2334)	(0.4152)	(0.2371)	(0.3083)	(0.0446)	(0.0563)
Age	0.0178	−0.0039	0.0182	0.0018	0.0000	0.0000
	(0.0295)	(0.0561)	(0.0301)	(0.0426)	(0.0054)	(0.0060)
Economics	−0.6005**	2.218***	−0.6391*	0.9206**	0.2424*	0.1340*
	(0.2629)	(0.4475)	(0.2648)	(0.3082)	(0.1020)	(0.0621)
Constant	−1.08107	−3.5251**	−0.8698	−1.4747	0.1481	0.9264***
	(0.6751)	(1.3010)	(0.6892)	(0.9558)	(0.1162)	(0.1585)
N	428		428		428	
Pseudo R2	0.0652		0.0377		0.0339	0.0844
Log Likelihood	-344.4393		-402.8067			
AIC	712.8786		829.6134			

Note: standard errors in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
[a] Partial liars are base category.
[b] Extreme liars reported 10 favorable outcomes.
[c] Extreme liars reported 9 or 10 favorable outcomes.
[d] Simultaneous quantile regression for 0.5 and 0.9 quantiles.

increase the probability of being a truth teller. From this follows that the shift in distribution is taking place in the extreme. The pledge significantly discourages extreme lying but does not cause a significant shift from partial liars to truth teller. This finding is robust for a broader specification of extreme liars (reporting 9 or 10 favorable outcomes) in Model 2b and is also reflected in the quantile regression where the pledge negatively affects the 90% quantile but has no significant effect on the median. The coefficients of the control variables are not significant or are consistent with earlier findings (Abeler et al., 2018; Rosenbaum et al., 2014).[19]

In order to test Hypothesis 2 we analyze whether the effects of the pledge observed in Table 4 differ between the gain and the loss frame. Therefore, we introduce an interaction term between the pledge and loss framing in the multinominal regression with a categorical variable for the three types of liars as dependent variable and partial liars as the base category and run quantile regressions on the propensity to lie separately for the two frames. Table 5 shows the results. As before, Model 3a assumes extreme liars to have reported 10 favorable outcomes and Model 3b includes 9 and 10 favorable reports in the category of extreme liars. Models 3a and 3b show no support for Hypothesis 2. The interaction term is not significant and thus, there is no indication of the pledge effect differing between frames. The quantile regressions in Models 3c and 3d show that the pledge significantly reduces the 90% quantile in the gain frame but not in the loss frame while the 50% quantile is unaffected by the pledge in both frames. Hence, we can conclude that, although Models 3a and 3b provide no support for Hypothesis 2, the quantile regressions in Models 3c and 3d provide some indications that the effect of loss framing on the pledge is going in the same direction as expected in Hypothesis 2.

5. Conclusion

In this study we analyze the effect of a truth pledge on the distribution of lying behavior in an online experiment with 484 participants. We find that a truth pledge administered at the beginning of the experiment significantly reduces over-reporting which is in line with earlier studies on the effectiveness of such an intervention (Bucciol & Piovesan, 2011; Mazar et al., 2008; Shu et al., 2011). However, the effect of a truth pledge is mainly driven by changes in the upper extreme of the distribution. The uneven shift in the distribution of lying behavior indicates that truth pledges reduce the size of lies but do not affect the decision to lie as predicted by Hypothesis 1. We only find weak support of Hypothesis 2 that the shift in the distribution of liars is lower under loss framing than under gain framing. We encourage further research on the topic, as we would expect Hypothesis 2 to hold in a possibly modified setting that reduces the costs of lying more effectively.

Our findings have implications for the application of truth pledges in various settings. While applications such as oath taking in the Anglo-Saxon legal system, in academic honor codes and in the Dutch bankers' oath might prevent extreme violations of the rules, they could be less effective in reducing smaller violations. Thus, future research should consider carefully how an intervention similar

[19] Being an economics student causes a shift towards more lying in general, but robustness checks controlling for other fields of studies do not alter the results qualitatively (see Table 9). Model 2a shows a gender effect on extreme liars but this effect disappears in the other model specifications.

Table 5
Regression analysis Hypothesis 2.

	Model 3a[a,b]		Model 3b[a,c]		Model 3c[d] Gain frame		Model 3d[d] Loss frame	
	TT	EL	TT	EL	Q50	Q90	Q50	Q90
Loss frame	0.3413	0.7127	0.2614	−0.0128				
	(0.3262)	(0.5069)	(0.3354)	(0.3681)				
Pledge	0.3874	−0.7599	0.2525	−1.0192*	−0.0656	−0.1951*	0.0304	−0.166
	(0.3325)	(0.6002)	(0.3406)	(0.4322)	(0.1144)	(0.0780)	(0.0715)	(0.0876)
Loss*Pledge	−0.2354	−0.4957	−0.1588	0.0946				
	(0.4431)	(0.8024)	(0.4531)	(0.5966)				
Female	−0.2115	0.8520*	−0.2271	0.3083	0.0861	0.0000	0.0000	0.0000
	(0.2341)	(0.4172)	(0.2379)	(0.3095)	(0.0983)	(0.0822)	(0.0610)	(0.1002)
Age	0.0161	−0.0078	0.0171	0.0026	0.0000	0.0000	0.0000	0.0000
	(0.0296)	(0.0569)	(0.0303)	(0.0428)	(0.0089)	(0.0089)	(0.0059)	(0.0133)
Economics	−0.6049*	2.2152***	−0.643*	0.9227**	0	0.0617	0.1944	0.134
	(0.2631)	(0.4489)	(0.2651)	(0.3085)	(0.1018)	(0.0716)	(0.1410)	(0.1019)
Constant	−1.1084	−3.5475**	−0.8920	−1.4807	0.1961	0.9264***	0.1189	0.8584*
	(0.6775)	(1.3130)	(0.6926)	(0.9532)	(0.1899)	(0.1812)	(0.1362)	(0.3486)
N	428		428		201		227	
Pseudo R2	0.0659		0.0379		0.0229	0.0678	0.0656	0.1089
log-lik	-344.3486		-402.7138					
AIC	716.2971		833.4275					

Note: standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
[a] Partial liars are base category.
[b] Extreme liars reported 10 favorable outcomes.
[c] Extreme liars reported 9 or 10 favorable outcomes.
[d] Simultaneous quantile regression for 0.5 and 0.9 quantiles by frame.

to a pledge affects lying in the specific situations.

Contrary to most studies on loss framing and unethical behavior, we find no difference in lying behaviour across frames and therefore, no effect of loss aversion on lying behaviour. Schindler and Pfattheicher (2017) and Cameron et al. (2008) find medium effect sizes for loss framing in different lying tasks. Our sample without the pledge had a power of 0.952 to detect a medium effect (d = 0.5). Grolleau et al. (2016) even found a large effect size (d = 0.956) of loss framing for which our sample has a power of 1.[20] Thus, an effect comparable to those studies is not present in our data. Our results concerning loss framing are in line with Charness et al. (2018) who find no significant difference for behavior in different frames. When comparing our study to those earlier studies, one possible explanation for the observed differences could be stake sizes. Stakes in Grolleau et al. (2016) and Schindler and Pfattheicher (2017) were considerably higher than in the present study, while Charness et al. (2018) use stakes more comparable to ours. However, the focus of the present study was on getting a more detailed understanding of the effect of truth pledges on lying and drawing conclusions on the mechanism of loss framing is beyond the scope of the presented work. Future research should systematically approach loss framing and lying behavior to explain observed differences in the effect.

Earlier literature on how lying behaviour varies with certain demographics has mostly found women to engage less in lying or to not behave differently from men (Rosenbaum et al., 2014, for a review). In our sample there was no effect of gender in the analysis on the participant level. However, when analyzing the distribution of types of liar, we found women to be more likely to be extreme liars than partial liars. Further analysis of this relationship will improve the understanding of gender effects in lying behavior.

[20] Earlier studies not only found considerable effect sizes but also highly significant effects. P-values range from p = 0.0229 in Cameron et al. (2008) to $p < 0.001$ in Grolleau et al. (2016).

Appendix A. Robustness checks

See Tables 6–9.

Table 6
OLS and Tobit regression analysis of treatment effects with main fields of study included.

Dependent variable: reported number of favorable outcomes						
	OLS regression[a]			TOBIT regression[b]		
Loss frame	−0.0249	−0.0288	−0.022	0.0031	−0.0018	0.0049
	(0.1768)	(0.1764)	(0.1764)	(0.1905)	(0.1905)	(0.1904)
Pledge	−0.4953**	−0.4927**	−0.489**	−0.5679**	-0.5633**	−0.5593**
	(0.1782)	(0.1780)	(0.1779)	(0.1916)	(0.1914)	(0.1913)
Female	0.2639	0.2243	0.2160	0.3401	0.2852	0.2772
	(0.2084)	(0.1911)	(0.1884)	(0.2167)	(0.2043)	(0.2035)
Age	−0.0039	−0.0027	−0.0016	−0.0048	−0.0031	−0.0019
	(0.0223)	(0.0222)	(0.0223)	(0.0266)	(0.0264)	(0.0264)
Economics	1.0219***	1.0376***	1.0180***	1.1714***	1.2029***	1.1845***
	(0.2186)	(0.2050)	(0.2030)	(0.2293)	(0.2212)	(0.2187)
Social sciences	−0.1000			−0.1472		
	(0.2410)			(0.2522)		
Natural sciences		−0.0928			−0.0943	
		(0.3111)			(0.3045)	
Humanities			−0.2416			−0.2344
			(0.3375)			(0.3391)
Constant	6.3462***	6.3222***	6.3142***	6.3775***	6.3298***	6.3210***
	(0.5312)	(0.5288)	(0.5193)	(0.6135)	(0.6064)	(0.6002)
sigma				1.9427	1.9429	1.9419
				(0.0710)	(0.0710)	(0.0710)
N	428	428	428	428	428	428
(pseudo) R2	0.0753	0.0751	0.0762	0.0214	0.0213	0.0215

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
[a] Robust standard errors in parentheses.
[b] Standard errors in parentheses.

Table 7
Share of favorable outcomes in all question.

Question	Share favorable reports	Pairwise tests of proportions[a] 1	2	3	4	5	6	7	8	9
1	63%***									
2	64%***	0.9436								
3	66%***	0.4769	0.5217							
4	66%***	0.3540	0.3919	0.8293						
5	68%***	0.1516	0.1728	0.4697	0.6119					
6	61%***	0.5273	0.4822	0.1793	0.1192	0.0390				
7	67%***	0.2241	0.2521	0.6137	0.7724	0.8272	0.0648			
8	60%***	0.3269	0.2933	0.0909	0.0567	0.0159	0.7276	0.0282		
9	61%***	0.5273	0.4822	0.1793	0.1192	0.0390	–[b]	0.0648	0.7276	
10	69%***	0.0719	0.0838	0.276	0.3822	0.7139	0.0151	0.5586	0.0055	0.0151

*** $p < 0.001$ for a two-sided binomial test against expected proportion of 0.5.
[a] Reports the p-value of test on the equality of proportion between column and row question.
[b] Share of favorable outcomes is identical in question 6 and 9.

Table 8
Random effects probit regression on favorable outcome in all rounds.

Dependent variable: favorable outcome reported				
Round	0.0015	−0.0105		
	(0.0066)	(0.0095)		
Last round			0.1434*	0.0701
			(0.0678)	(0.0947)
Loss frame	−0.004	−0.0039	−0.0041	−0.0042
	(0.0508)	(0.0508)	(0.0508)	(0.0508)
Pledge	−0.1471**	−0.2748**	−0.1471**	−0.1608**
	(0.0518)	(0.0896)	(0.0518)	(0.0531)
Pledge*Round		0.0232		
		(0.0132)		
Pledge*Last round				0.1430
				(0.1354)
Female	0.0722	0.0723	0.0723	0.0724
	(0.0541)	(0.0541)	(0.0542)	(0.5419)
Age	−0.0008	−0.0008	−0.0008	−0.0008
	(0.0063)	(0.0063)	(0.0064)	(0.0064)
Economics	0.3141***	0.3143***	0.3145***	0.3145***
	(0.0609)	(0.0609)	(0.0610)	(0.0610)
Constant	0.3383*	0.4046*	0.3324*	0.3396*
	(0.1544)	(0.1596)	(0.1486)	(0.1489)
sigma	0.3133	0.3136	0.3141	0.3141
	(0.0343)	(0.0343)	(0.0343)	(0.0343)
rho	0.0894	0.0895	0.0898	0.0898
	(0.0178)	(0.0178)	(0.0179)	(0.0179)
N	4280	4280	4280	4280
Wald chi2(6)	31.61	35.3	35.84	37.13
Log Pseudolikelihood	−2732.13	−2730.76	−2729.93	−2729.39

Coefficients of random effects probit regression.
Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 9
Results of Multinomial Logit regression with main fields of study included.

	Model (i)		Model (ii)		Model (iii)	
	Truth tellers	Extreme liars	Truth tellers	Extreme liars	Truth tellers	Extreme liars
Loss frame	0.2183	0.5147	0.2149	0.5072	0.2028	0.5098
	(0.2199)	(0.3920)	(0.2198)	(0.3900)	(0.2208)	(0.3904)
Pledge	0.2524	−1.0605	0.2582	−1.0346*	0.2516	−1.0369*
	(0.2223)	(0.4051)	(0.2219)	(0.4029)	(0.2229)	(0.4032)
Female	−0.1586	0.976	−0.2012	0.8619*	−0.1481	0.8859*
	(0.2534)	(0.4143)	(0.2362)	(0.4175)	(0.2362)	(0.4158)
Age	0.0161	−0.0085	0.0178	−0.0038	0.0147	−0.0049
	(0.0298)	(0.0556)	(0.0295)	(0.0562)	(0.0297)	(0.0560)
Economics	−0.6438*	1.8137***	−0.5987*	2.1620***	−0.4847	2.2364***
	(0.2767)	(0.4973)	(0.2728)	(0.4689)	(0.2711)	(0.4651)
Social sciences	−0.1317	−0.9974				
	(0.2813)	(0.7383)				
Natural sciences			0.007	−0.3854		
			(0.3321)	(1.0866)		
Humanities					0.642	0.1181
					(0.3580)	(1.0979)
Constant	−1.0141	−3.0643*	−1.0833	-3.4625**	−1.1363	−3.5248**
	(0.6906)	(1.3115)	(0.6824)	(1.3137)	(0.6802)	(1.3058)
N	428		428		428	
Log Likelihood	-343.425		-344.368		-342.837	
AIC	714.8509		716.7369		713.675	
Pseudo R2	0.0679		0.0653		0.0695	

Note: Partial liars are base category.
Coefficients of multinomial logit regression with standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix B. Experimental instructions

See Table 10.

Table 10
Questions in the Even–Odd task.

Question 1	For this round please recall the day and month of the birthday of a relative or friend. Next, please sum up the 4 digits of the birthday, hence d + d + m + m
Question 2	For this round please recall the last 4 digits of a phone number you know by heart. Next, please sum up the 4 digits
Question 3	For this round please recall two pets of a relative or friend. How many letters are in each of the pets' name? Next, please sum up the number of letters of the first name and the number of letters of the second name
Question 4	For this round please recall the house number of a close friend and the house number of a relative. Next, please sum up all digits of the two numbers
Question 5	For this round please recall the age of two friends. Next, please sum up all digits of the two numbers
Question 6	For this round please recall the year of birth of a relative. Next, please sum up the 4 digits of the year
Question 7	For this round please recall the first and last name of a friend or relative. How many letters are in the first name and how many in the last name? Next, please sum up the number of letters in the first name and the number of letters in the last name
Question 8	For this round please check the time at this exact moment. Next, please sum up the 4 digits of the time, hence h + h + m + m
Question 9	For this round please recall the last name of a relative and the place this relative lives in. How many letters are in the last name and in the name of the place? Next, please sum up the number of letters in the name and the number of letters in the name of the place
Question 10	For this round please recall the first names of two friends. How many letters are in each name? Next, please sum up the number of letters of these two names

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.joep.2019.05.006.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2018). Preferences for truth-telling. *Econometrica* forthcoming.
Anderhub, V., Müller, R., & Schmidt, C. (2001). Design and evaluation of an economic experiment via the internet. *Journal of Economic Behavior & Organization, 46*(2), 227–247.
Aquino, K., Reed, I., et al. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440.
Bains, W. (2008). Random number generation and creativity. *Medical Hypotheses, 70*(1), 186–190.
Balasubramanian, P., Bennett, V. M., & Pierce, L. (2017). The wages of dishonesty: The supply of cheating under high-powered incentives. *Journal of Economic Behavior & Organization, 137*, 428–444.
Becker, G. S. (1968). Crime and punishment: An economic approach. In N. G. Fielding, A. Clarke, & R. Witt (Eds.). *The economic dimensions of crime* (pp. 13–68). UK: Palgrave Macmillan.
Birnbaum, M. H. (1999). Testing critical properties of decision making on the internet. *Psychological Science, 10*(5), 399–407.
Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General, 142*(4), 1001–1005.
Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology, 32*(1), 73–78.
Cameron, J. S., Miller, D., & Monin, B. (2008). *Deservingness and unethical behavior in loss and gain frames. Working paper.* Berkeley: University of California.
Charness, G., Blanco-Jimenez, C., Ezquerra, L., & Rodriguez-Lara, I. (2018). Cheating, incentives, and money manipulation. *Experimental Economics,* 1–23.
Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree-an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97.
Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature, 516*, 86–89.
Cohn, A., Maréchal, M. A., & Noll, T. (2014). Bad boys: how criminal identity salience affects rule violation. *The Review of Economic Studies, 82*(4), 1289–1308.
Conrads, J., & Lotz, S. (2015). The effect of communication channels on dishonest behavior. *Journal of Behavioral and Experimental Economics, 58*, 88–93.
Diener, E., & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality, 10*(1), 107–111.
Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness.* Academic Press.
Effron, D. A., Bryan, C. J., & Murnighan, J. K. (2015). Cheating at the end to avoid regret. *Journal of Personality and Social Psychology, 109*(3), 395.
Figurska, M., Stańczyk, M., & Kulesza, K. (2008). Humans cannot consciously generate random numbers sequences: Polemic study. *Medical Hypotheses, 70*(1), 182–185.
Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise-an experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525–547.
Fosgaard, R. T. (2019). Students cheat more: Comparing dishonesty of a student and a representative sample in the laboratory. *The Scandinavian Journal of Economics* (in press).
Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters, 117*(3), 624–626.
Friesen, L., & Gangadharan, L. (2013). Designing self-reporting regimes to encourage truth telling: An experimental study. *Journal of Economic Behavior & Organization, 94*, 90–102.
Garbarino, E., Slonim, R., & Villeval, M. C. (2017). Loss aversion and lying behavior: Theory, estimation and empirical evidence. GATE Working Paper 1631.
Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review, 108*(2), 419–453.
Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization, 93*, 293–300.
Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making, 26*(3), 213–224.
Gosling, S., Vazire, S., Srivastava, S., & John, O. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93–104.
Grolleau, G., Kocher, M. G., & Sutan, A. (2016). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science, 62*, 3428–3438.
Hannan, R. L., Hoffman, V. B., & Moser, D. V. (2005). Bonus versus penalty: Does contract frame affect employee effort? *Experimental business research* (pp. 151–169).

Springer.
Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics, 14*(3), 399–425.
Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review, 56*(8), 1645–1655.
Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization, 127*, 99–114.
Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2017). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys, 32*, 357–387.
Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization, 93*, 328–336.
Johansson-Stenman, O., & Martinsson, P. (2006). Honestly, why are you driving a BMW? *Journal of Economic Behavior & Organization, 60*(2), 129–146.
Jokar, E., & Mikaili, M. (2012). Assessment of human random number generation for biometric verification. *Journal of Medical Signals and Sensors, 2*(2), 82–87.
Kern, M. C., & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science, 20*(3), 378–384.
López-Pérez, R., & Spiegelman, E. (2012). Do Economists Lie More? Working Papers in Economic Theory, Universidad Autónoma de Madrid (Spain).
Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644.
McCabe, D. L., & Trevino, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education, 64*(5), 522–538.
McCabe, D. L., & Trevino, L. K. (1997). Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education, 38*(3), 379–396.
Naquin, C. E., Kurtzberg, T. R., & Belkin, L. Y. (2010). The finer points of lying online: E-mail versus pen and paper. *Journal of Applied Psychology, 95*(2), 387–394.
Newberry, K. J., Reckers, P. M. J., & Wyndelts, R. W. (1993). An examination of tax practitioner decisions: The role of preparer sanctions and framing effects associated with client condition. *Journal of Economic Psychology, 14*(2), 439–452.
NVB (2018). Code banken. https://www.nvb.nl/publicaties/gedragscodes/1241/code-banken.html.
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.
Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology, 69*, 172–177.
Schweitzer, M. E., Ordóñez, L., & Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Academy of Management Journal, 47*(3), 422–432.
Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making, 7*(3), 292–303.
Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*(2), 181–190.
Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23*(10), 1264–1270.
Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science, 24*(2), 125–130.
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.
Sonnemans, J., Schram, A., & Offerman, T. (1998). Public good provision and public bad prevention: The effect of framing. *Journal of Economic Behavior & Organization, 34*(1), 143–161.
Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin, 77*(1), 65–72.
Willen, R. S. (1983). Rationalization of anglo-legal culture: The testimonial oath. *British Journal of Sociology,* 109–128.