

THE UNIVERSITY OF CHICAGO

Comparison of Dishonesty Interventions: A Conceptual Replication Study

By

Nikal S. Toor

December 2022

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Emma Levine
Preceptor: Rubén Rodríguez Barrón

Introduction

Dishonesty has been of considerable research interest due to its pervasive and generally costly nature. Dishonesty is a broad categorization that includes lying and other expressions of falsehoods. A large body of research has examined the causes of dishonest behavior (for an overview, see: Gerlach et al., 2019). For example, individual qualities, such as creativity (Gino & Ariely, 2012); triggers, such as energy depletion (Gino et al., 2011); and motivations, such as competition (Cohen et al., 2009), have all been linked with increased dishonesty. However, if we hope to limit dishonesty behavior, an understanding of effective interventions is needed.

Which interventions are most effective in reducing dishonest behavior? Currently, this question is difficult to answer with confidence. There are multiple reasons for this. First, the ongoing replication crisis has found one-off investigations less likely to be accurate and consequently less trusted (Baker, 2015). Second, there has been no direct comparison of interventions using a standard procedure or measure. Due to several dishonesty interventions failing to replicate (e.g., Cai et al., 2015), the use of nonuniform procedures (Hertwig & Mazar, 2022), and other recent credibility incidents (Kristal et al., 2020), there is a need for increased reliability of dishonesty reducing interventions.

The present study aims to provide an increased understanding of dishonesty interventions. In order to do so, we replicate several previous studies. Instead of using exact replications, we opted for conceptual replication and a standardized procedure with limited differences between interventions. Conceptual replication has the added benefit of testing the intervention's generalizability, making it possible to compare interventions against one another.

Literature Review

Treatment Selection

The present study's treatment list was inspired by studies included in Gerlach et al.'s (2019) meta-analysis on dishonest behavior. Gerlach et al. (2019) focused on one-shot interactions within four experimental paradigms: sender-receiver games (e.g., Gneezy, 2005), die-roll tasks (e.g., Fischbacher & Franziska, 2013), coin-flip tasks (Houser et al., 2012), and matrix tasks (e.g., Mazar et al., 2008). In order to generate our treatment list, each study included in Tables A1 through A4 in Appendix A of the meta-analysis was systematically evaluated and sorted into seven categories based on their manipulations (Gerlach et al., 2019). These categories included group size (e.g., Cohen et al., 2009), incentive size (e.g., Kajackaite & Gneezy, 2017), motivations (e.g., Gneezy, 2005), nudges (e.g., Mazar et al., 2008), triggers (e.g., Gino et al., 2011), individual differences (e.g., Piff et al., 2012), and other (e.g., Wibral et al., 2012). These categories were not mutually exclusive, as several studies fit into multiple categories.

We excluded studies falling into the individual differences category as we were not interested in individual participant attributes (e.g., Piff et al., 2012). The initial list was then tailored for an online participant pool. Many studies were not methodologically feasible with online participants and were eliminated, such as those requiring participants to make significant external arrangements (e.g., Kouchaki et al., 2014). Due to the nature of an online participant pool and the present study's interest in individual-level lying, studies falling within the group size category were also excluded. As the present study utilized a die-roll task, several studies were eliminated on their infeasible nature under the

restrictions of this paradigm, such as those manipulating strategy in sender-receiver games (e.g., Cohen et al., 2009).

With the removal of these manipulations, the individual differences, group sizes, and other categories were eliminated. Upon review of the remaining studies, those examining emotional manipulation (e.g., Motro et al., 2018) were differentiated enough to be included as a category, bringing the total to five categories: incentive size, motivations, nudges for honesty, triggers for dishonesty, and emotional manipulations. The remaining studies in these five categories were then further sorted into 31 specific subcategories representing potential experimental treatment groups. Of the 31 categories, those chosen included moral primes, signing, publicity, and descriptive norms because they fall under the nudges for honesty category, which was of the most interest in evaluating, and because of their popularity among other researchers (Garlach et al., 2019).

From these categories the final treatments were selected. They included: reading a descriptive norm, observability via a virtual die-roll (included 2 treatments, one explicitly stating the ability for the experimenters to view the die-roll outcomes), being primed with a religious moral prime (recalling and typing the Ten Commandments), reading an honesty pledge, and signing an agreement to be honest. A presentation by Peer et al. (2022) also motivated the inclusion of copying an honesty pledge (by typing it out) due to its fit in the intervention list and potential for high effectiveness.

Interventions for curbing dishonesty

The interventions tested in the present study (and the papers they were inspired by) are some of the most widely tested methods of reducing dishonesty (Hertwig & Mazar, 2022). They include descriptive norms (Robert & Arnab, 2013), observability (Ploner &

Regner, 2013), moral primes (common, see: Hertwig & Mazar, 2022), reading a pledge (Shu et al., 2011), typing a pledge (Peer et al., 2022), and signing (Shu et al., 2011 & Chou 2015). The following sections will provide background details about each intervention selected.

Social Norms

The influence of social forces, such as social norms, on human behavior, is well documented (Cialdini & Goldstein, 2004). One hypothesized reason for the use of social norms in decision making is to conform to others and respond accurately in social situations (Cialdini & Goldstein, 2004). Responding accurately in social situations is thought to be motivated by people's general objective of achieving their goals effectively (Cialdini & Goldstein, 2004). The social norms of interest in the present study are descriptive norms. Descriptive norms give us information about typically behavior and can be contrasted with injunctive norms, which give us information about typical expectations (Cialdini & Goldstein, 2004).

The descriptive norms intervention was inspired by Robert & Arnab (2013). In their study a classic sender-receiver game was used to test if different descriptive norms effected lying behavior. Roberts & Arnab (2013) took one sample from Calcutta, India and one sample from Arizona, USA. In the control groups they found that baseline levels of lying were 55.6% in Calcutta and 41.2% in Arizona. Telling participants that 15% of previous participants had lied in a similar situation, which was the lowest level presented, significantly decreased lying in Calcutta to 30.8%. The same message did not significantly decrease lying in Arizona (36%). Due to these mixed results and influence of social forces on human behavior, descriptive norms were included in the present study's intervention

list to be replicated. The presentation of descriptive norms also fit nicely with the present study's methodology.

Observation

Reputational forces have long been thought to change moral behavior and are theorized to have played a part in the evolution of human morality (Tomasello et al., 2013). Due to the importance of reputation, being observed has an effect on (im)moral behavior (Fries et al. 2021). The effect is so strong that being subconsciously primed to believe you are being watched has been seen to effect lying behavior (Haley & Fessler, 2005), though these findings have been questioned (Cai et al., 2015).

Fischbacher & Föllmi-Heusi's (2013) study, which guided the creation of both of our observability treatments, did not explicitly test observability to curb dishonesty, but did find that anonymity significantly increased dishonest behavior. Using a similar procedure, Fries et al. (2021) failed to replicate Fischbacher & Föllmi-Heusi's (2013) findings. However, Fries et al. (2021) did find significant evidence that observability decreases lying behavior. Due to the history of strong evidence that observability decreases immoral behavior, it was included in the present studies interventions list.

Moral Primes

Priming is the exposure to external stimuli that subconsciously activate mental constructs and influence responses to subsequent stimuli (Weingarten et al., 2016). Priming has been a controversial topic in the social sciences (for an overview: Weingarten et al., 2016). There have been successful (e.g., Hull et al., 2002) and failed (e.g., Doyen et al., 2012) attempts to replicate priming effects. In a meta-analysis, Weingarten et al. (2016) found that priming had a small impact on behavior, and claimed that although the

impact was small, there was still hope that priming could be useful in shaping behavior. In the present study, primes with moral connotation (i.e., moral primes), were used to evaluate their effectiveness as an intervention to decrease dishonesty.

It has been hypothesized that religion aided the human development of large and relatively peaceful communities of unrelated members (Atran & Norenzayan, 2004). As such, moral primes with religious connotation have often been used due to the impact religion has had on human life (e.g., Shariff & Norenzayan, 2007). For example, it was found that priming participants by having them solve word scramble puzzles with religious terms (i.e., spirit, divine, God, sacred, and prophet), significantly increased the value of donations made by participants (Shariff & Norenzayan, 2007). There is also an extensive literature on the neurocognitive effects of religion (for an overview see: Boyer, 2008).

The moral prime used in the present study was to have participants recall and type out the Ten Commandments. According to Hertwig & Mazar (2022), prompting participants to recall the Ten Commandments is the most tested intervention to decrease dishonesty. In a seminal paper, Mazar et al. (2008) found that those who were asked to recall and write down the Ten Commandments reported solving significantly less problems, thus receiving less payoff, on a problem-solving task than those who did not. Mazar et al. (2008) attributed the reduction in solved problems to less cheating by those who were given the priming task. Verschuere et al. (2018) performed a large ($n = 5,786$) replication study of the Mazar et al. (2008) paper and found that no such effect was present. Due to the prevalence of having participants recall the Ten Commandments and the presence of conflicting results, it was included as an intervention in the present study.

Moral pledges

Maintaining a positive self-concept is hypothesized to play a critical role in dishonest behavior (Mazar et al., 2008). Pledges can change the salience of ethical standards and therefore alter an individual's awareness of their self-concept (Shu et al., 2011). By increasing engagement with a pledge, the amount of moral disengagement possible decreases as the self-concept is made more salient (Shu et al. 2011). Examples of different levels of engagement with pledges include reading, signing, and directly copying. For example, increasing engagement with a signature, which is thought to be a representation of the self, decreases the level of moral disengagement possible (Chou, 2015).

Shu et al. (2011) had one group of participants read honor pledges and those in another group read and sign the same pledge. They found that signing was more effective, but both interventions significantly decreased dishonesty. In a paper presentation, Peer et al. (2022) claimed that to make pledges most effective, participants need to interact with the pledge, such as by copying them word-for-word. Signing on the other hand has recently failed to replicate as an intervention (Kristal et al., 2020). There have also been recent incidents involving the legitimacy of the data that failed to replicate (Simonsohn et al., 2021). Due to the high interest in studies involving pledges (Hertwig & Mazar, 2022), the recent events on signing interventions (Simonsohn et al., 2021), and the ability for pledges to be easily administered in our online procedure, reading a pledge, typing a pledge, and signing a statement were included in our interventions list.

Conceptual Replication & direct comparison of interventions

Due to the ongoing replication crisis in the social sciences (Baker, 2015), it is of paramount importance for researchers to demonstrate the validity of previous findings. Conceptual replications are one way of demonstrating validity. The present study is a conceptual replication study as each of its interventions were inspired by previous work but did not directly replicate their procedures. By conceptually replicating the aforementioned interventions we aim to provide further evidence for the effectiveness or ineffectiveness of each intervention in decreasing dishonest behavior. Conceptual replication also allows each intervention to be tested in a different scenario from its original study, allowing for the generalizability of the interventions to be assessed.

By doing these conceptual replications with a standardized procedure and consistent paradigm, head-to-head comparisons of these interventions was also possible. The idea to use a standardized a procedure and to assess multiple interventions against one another was inspired by DellaVigna & Pope (2018). In their study a variety of incentives were evaluated for their effectiveness in motivating effort from participants. These comparisons allowed for DellaVigna & Pope (2018) to rank the incentives and provide evidence of their relative effectiveness.

One of the reasons for the lack of comparison between dishonesty decreasing interventions has been the broad array of methods used (Gerlach et al., 2019). Generally, four main paradigms have been used: sender-receiver games, coin-flip tasks, matrix tasks, and die-roll tasks (Gerlach et al., 2019). Each of these paradigms has its advantages, however using different paradigms has resulted in the inability to directly compare the effectiveness of the interventions. One of the goals of this research is to

provide clarity when comparing different interventions by using a standardized procedure with a consistent paradigm.

The paradigm chosen for the present study was the die-roll paradigm. The first benefit of using the die-roll paradigm is the ability to observe partial lies. Using a similar die-roll paradigm, Fischbacher & Föllmi-Heusi (2013) found that approximately 20% of participants lied to the full extent, while approximately 41% lied partially. Another benefit of using the die-roll paradigm is its ability to provide high rates of dishonesty. Of the four main paradigms covered by Gerlach et al. (2019) in their meta-analysis of 558 studies, the die-roll paradigm provided the highest rate of dishonesty at 52%. A third benefit of the die-roll task is its easy portability to the online environment used in the present study.

The use of an online survey also allowed for a highly streamlined procedure. The presentation of each survey was consistent other than the differences necessary for each intervention. Data was collected from each intervention at the exact same time, allowing for temporal consistency. The online sample provided the same participant pool for each intervention, namely users from Prolific Academic. Lastly, Prolific Academic has been found to produce high quality data (Peer et al, 2017).

Due to the costly nature of lying, it is important to determine which interventions are the most effective at reducing it. With the use of a standardized procedure, our objective was to help make this determination. By utilizing a die-roll paradigm and an online survey it was possible to conceptually replicate and compare previously studied interventions.

Method

Participants

In total, 401 participants were recruited on Prolific Academic. One participant was dropped from the study for failing an attention check. This left 400 participants (199 male, $M_{age} = 41.33$ years, age range: 18 – 78 years) who completed the study. The sample was gender balanced due to the possibility of gender differences in lying (Drever & Johannesson, 2008). Participants were compensated \$1.05 for completing the entire survey and a bonus of up to \$0.81, depending on their performance.

Procedure

Participants were recruited to complete an unrelated initial survey about advertisements containing Coronavirus Disease 2019 (COVID-19) related information. The COVID-19 survey was used in order to display the main die-roll task as a surprise bonus.

Control Condition

As this is a conceptual replication study, each of the interventions and the overall format itself was guided by previous works. The basics of the die-roll intervention layout was directed by the use of Fischbacher & Franziska (2013) as well as Kajackaite & Gneezy (2017). As both of those studies took place in a laboratory setting, the present research had to be adapted to an online environment.

On the primary instructions page (shown in Figure A. 1) the control group participants were told they would have to locate a physical or virtual (i.e., online) 6-sided die and to roll the die three times when asked to. They were told that only the third number they rolled should be reported. Control group participants then saw the die-roll page (shown in Figure A. 2). On this page, participants were instructed to locate and roll their die three times. Participants were also reminded that only the third roll determines their

pay. Finally, on the report page (shown in Figure A. 3), participants were asked to report their third roll by typing it into an entry box. On all three pages of the die-roll task (instructions, die-roll, and report) participants were presented a table presenting the bonus payment scheme. The scheme was constructed such that participants received \$0.01 for reporting a 1, \$0.03 for reporting a 2, \$0.09 for reporting a 3, \$0.27 for reporting a 4, \$0.81 for reporting a 5, and \$0 for reporting a 6.

Several methods were used to make it more likely for participants to be dishonest. First, on the instruction page participants were asked to roll the die three times and were explicitly told that only their third roll counted and was to be reported. By rolling the die three times participants produced counterfactuals, which have been shown to increase the likelihood of lying (Shalvi et al., 2011). Second, a payment of \$0 for reporting a six was used to tempt those participants who rolled the highest number, but were paid less, to cheat (Fischbacher & Föllmi-Heusi, 2013). Third, an exponentially increasing payment scheme was used to increase the appeal of reporting the higher payout numbers. Fourth, a die-roll task was used in order to measure the extent to which participants are willing to lie, as it has been found that many participants only partially lie when performing this task (Fischbacher & Föllmi-Heusi, 2013).

Treatment Conditions

The procedure for the descriptive norms group was nearly identical to the control condition. The only difference was on the primary instructions page the text, “In previous situations, approximately 8% of participants misreported the results of similar tasks” was included (shown in Figure A. 4). This intervention was guided by Robert & Arnab’s (2013) procedure, in which they told participants different levels of participants had lied in the

past, ranging from 15% to 85%. The percentage presented in the present study's descriptive norm intervention was produced by an early pilot of the observability intervention, in which 8% of participants lied. It should be noted that the 8% message used in the present study's intervention was lower than the 15% used by Robert & Arnab (2013) in their "low" lying percentage treatment group.

For the two observability groups, participants read nearly identical instructions to the control condition, except participants were told they would be provided a virtual die, instead of having to find their own (shown in Figure A. 5). The instructions for the explicitly observed group also included the line of text which read, "However, you are able to report any number" altered to read, "Although we are able to observe your rolls, you are able to report any number" (shown in Figure A. 6). On the die-roll page participants in the observability groups were then presented with a virtual die and reminded that only their third roll determined their pay (shown in Figure A. 7). Participants had to roll their virtual die three times before continuing. They were then presented with the same report page as the other groups. The observability interventions were inspired by Ploner & Regner (2013) and loosely inspired by experiments which provide observability primes (e.g., Haley & Fessler, 2005). In Ploner & Regner's (2013) procedure an assistant was physically present observing the rolls. However, as our procedure was online, a virtual die built into the survey was used, consequently leading participants to presume that the true result of their die-roll was observable.

The moral primes, signing, read-pledge, and type-pledge treatments were very similar to the control condition. Each of them first presented the same primary instructions page as the control condition. The main difference came between the primary instructions

page and the die-roll page instructions. In this gap these groups were subjected to alternative pages described below. Participants were forced to remain on these pages for a minimum of 10-seconds before they could continue.

Participants in the moral prime group were requested to take a maximum of 2-minutes to recall and type out as many of the Ten Commandments as they could in the space provided (shown in Figure A. 8). Asking participants to recall the Ten Commandments as a moral prime is a common practice (Hertwig & Mazar, 2022).

Participants in the signing group were asked to sign their initials and agree to a statement which read, “By signing, you certify that the information you submit is complete and correct, to the best of your knowledge”, and then presented a space to virtually sign their initials (shown in Figure A. 9). The main guide for this intervention was Shu et al. (2011), and the use of a digital signature was inspired by Chou (2015).

Participants in the read-pledge group were asked to read the following pledge: “I promise that I will only report the number that appears on the dice, as rolled. I know that I will be paid based on my reporting and hence will take it very seriously to be accurate in my reporting” (shown in Figure A. 10). The main guide for this intervention was Shu et al. (2011). It should be noted that the pledge presented to participants in the present study was significantly shorter than that used by Shu et al. (2011). This was because the same pledge was used by the type-pledge group, and it would not have been practical for participants to have typed a longer pledge.

The type-pledge procedure was the only intervention not guided by a paper from the Garlach et al., (2019) meta-analysis. This intervention was inspired by a paper presentation by Peer et al. (2022) and was included due to its relevance and potential for

high effectiveness. Participants in the type-pledge group were presented the same instructions as the read-pledge group, and then requested to type out the pledge word-for-word in a space provided (shown in Figure A. 11). The text of the pledge was in the form of a screenshot, so participants could not cheat by copying and pasting the pledge. While it was not ensured that the text matched letter for letter, it was ensured that participants typed at least as many characters as in the pledge. Upon inspection, it was seen that all participants in this group typed the pledge with high accuracy.

Results

In this section, the interventions will first be compared to the percent each number is expected to be rolled and the expected mean number rolled. For all results statistics reporting a 6 was converted to 0 due to the nature of the payment scheme (reporting a 6 paid \$0). Due to this the expected mean is 2.5. The die is fair and so a uniform distribution is expected. A six-sided die was used throughout, so the uniform distribution would produce an expected frequency of $1/6$ or 16.7% for each number rolled. Next, the interventions will be compared head-to-head against one another. Across participants in all groups there was no difference between sexes ($p > 0.1$).

Comparison of groups to baseline

The results for two-sided binomial test comparing each roll to the expected frequency of 16.7% are presented in Table 1. With the exception of the observability and explicit observability groups ($p > 0.05$), all groups were found to significantly underreport the value of 1 ($p < 0.05$). All groups, with the exception of the explicit observability and signing groups ($p > 0.05$), were found to overreport the value of 5 ($p < 0.05$).

Table 1. Percentage of subjects who reported corresponding numbers.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------------------------|-------|-------|-------|----------|----------|--------|
| Control (n = 50) | 4.00* | 8.00 | 10.00 | 38.00*** | 38.00*** | 2.00** |
| Descriptive Norms (n = 50) | 4.00* | 10.00 | 24.00 | 24.00 | 38.00*** | 0.00** |
| Observability-Explicit (n = 51) | 9.80 | 23.53 | 13.73 | 13.73 | 21.57 | 17.65 |
| Observability (n = 51) | 17.65 | 19.61 | 9.80 | 9.80 | 33.33** | 9.80 |
| Prime (n = 49) | 4.08* | 4.08* | 6.12 | 30.61* | 44.90*** | 10.20 |
| Read Pledge (n = 49) | 4.08* | 4.08* | 16.33 | 32.65** | 36.73*** | 6.12 |
| Type Pledge (n = 50) | 6.00* | 6.00* | 22.00 | 18.00 | 44.00*** | 4.00* |
| Sign (n = 50) | 6.00* | 12.00 | 22.00 | 30.00* | 20.00 | 10.00 |

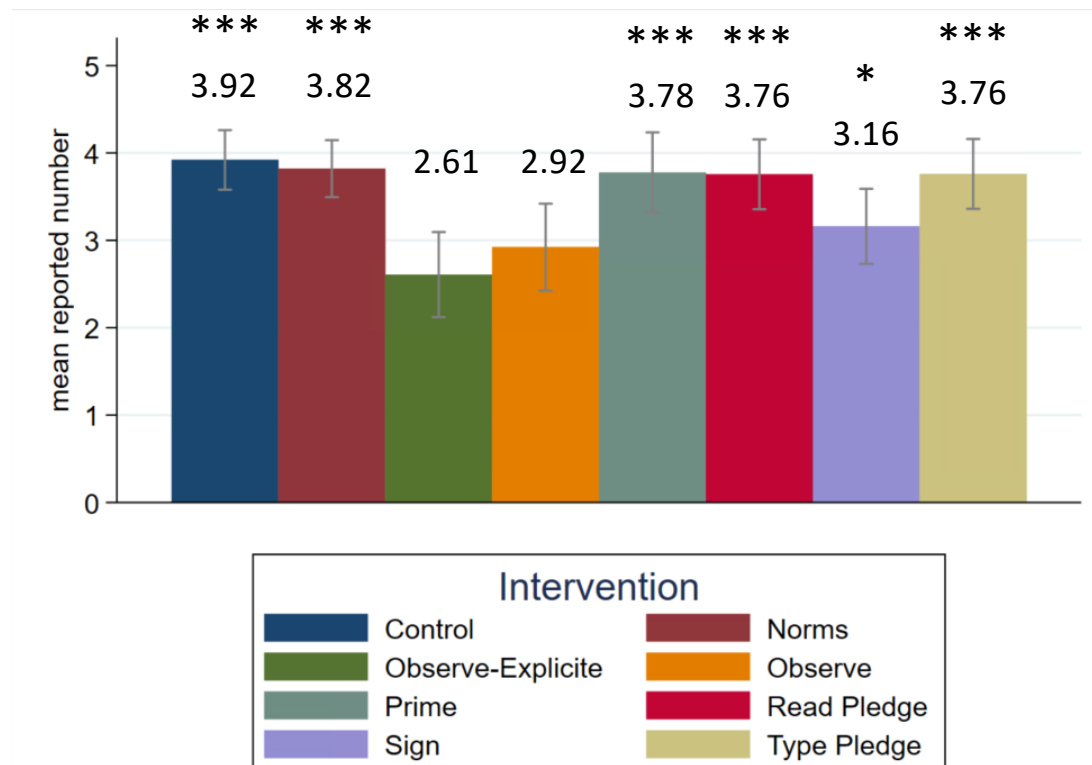
Note: *, ** and *** denote 5%, 1%, and .1% significance levels respectively of a two-sided binomial test of the null hypothesis that the observed frequency is equal to 16.7%.

Figure 1 shows the means of each intervention compared to the expected mean of 2.5. The following group's means were all significantly greater than expected with $p < 0.001$: control ($M = 3.92$), descriptive norms ($M = 3.82$), moral prime ($M = 3.78$), read-pledge ($M = 3.76$), and type-pledge ($M = 3.76$). Signing ($M = 3.16$) was significantly greater than expected with $p < 0.01$. The observability group's ($M = 2.92$) and explicit observability group's ($M = 2.92$) means did not significantly differ from the expected mean.

Comparison between groups

A one-way ANOVA was performed to compare the effect of intervention group on mean number reported. The ANOVA showed a significant difference ($F(7, 392) = 5.39$, $p < 0.001$). A Tukey's HSD Test found the mean die-roll number reported was significantly lower for the explicit observability group than the mean of each of the following groups: control ($p < 0.001$), descriptive norms ($p < 0.01$), moral prime ($p < 0.01$), read-pledge ($p < 0.01$), and typepledge ($p < 0.01$). The same test found the mean of the die-roll number reported was significantly lower for the observability group than the control group ($p < 0.05$). Figure 2 gives a visual of each intervention's distribution.

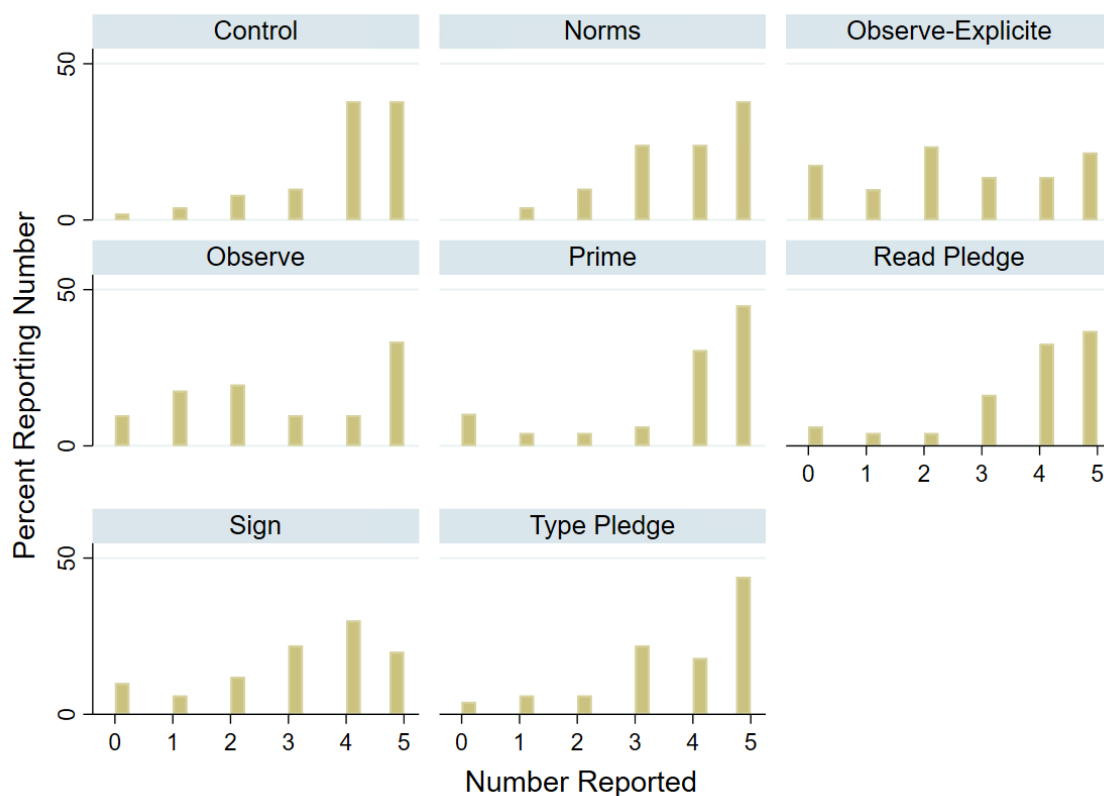
Figure 1. Mean comparison of interventions



Note: *, ** and *** denote 5%, 1%, and .1% significance levels respectively of *t*-tests of the null hypothesis that the mean is equal to 2.5. Reported values of 6 converted to 0.

Looking at Figure 1, it can be seen that the mean number reported was lower for the two observability interventions. This was followed by the signing intervention, however this intervention resulted in a significantly higher mean number of 3.16 reported than the expected 2.5 value ($p < 0.05$), suggesting there was significant lying in this group.

Figure 2. *Distribution of numbers reported by intervention*



Note: Reported values of 6 converted to 0.

Discussion

With the present study, we aimed to replicate and compare previously studied interventions' ability to decrease dishonest behavior. In order to achieve these goals, we utilized a standardized procedure, a die-roll paradigm, and an online participant pool. In

order to evaluate the conceptual replication of these interventions, we compared them to the expected uniform distribution of numbers rolled. This comparison showed that only the observability intervention with an explicit statement of observability was not significantly different from a uniform distribution for every value reported. The observability intervention without the explicit statement was also not significantly different from a uniform distribution, except for the reports of the number five, which resulted in the highest payoff. The observability interventions' means were also not significantly different from the expected value from random die-rolls.

Based on these findings, the two observability interventions were the only ones to replicate their ability to significantly reduce lying. The next most effective intervention was signing, but it did not significantly reduce dishonesty. However, digital signatures, such as that used in the present study, have been found to be less effective in curbing dishonest behavior than hand signatures (Chou, 2015). The rest of the interventions showed little or no effect in reducing lying.

Interestingly, the reported value of three was not significantly different from the expected 16.7% value for any of the interventions. This finding suggests that being in the top half of payoffs, regardless of the intervention, was enough to encourage participants to be honest. Another possible interpretation is that some participants who rolled in the bottom half of the distribution lied and reported a three, while others who rolled three reported even higher numbers. Due to most of the interventions' unobservable nature, it is impossible to know which interpretation is accurate. The observability interventions were the only interventions where this difference could be evaluated. However, they are

not helpful for this comparison because they were the only two interventions to reduce lying successfully.

The second interpretation is in line with findings that suggest there are psychological motives to not lie to the fullest extent (Fischbacher & Föllmi-Heusi, 2013). The two motives presented by Fischbacher & Föllmi-Heusi (2013) include appearing honest to others and the ability to maintain a favorable self-concept. Alternatively, Kajackaite & Gneezy (2017) suggest that higher incentives do not entice lying behavior due to perceived increased risk of being exposed as a liar.

Another finding of interest was that the value of one was reported significantly less than expected in all interventions except for the observability conditions. However, the reporting of six, which resulted in no pay, was only significantly lower than expected in the type-pledge, descriptive norms, and control conditions. This finding could also be explained by the psychological motives described by Fischbacher & Föllmi-Heusi (2013). If participants did not win a bonus, they might feel less tempted to lie and take money they did not win. However, if participants won at least some money, they might be comfortable lying to receive increased pay. Another explication for this might be that participants did not notice that reporting a six resulted in no pay, resulting in some participants believing they would be paid the maximum amount. This explanation is less plausible as participants were shown the payout table on the instructions, die-roll, and report pages.

By comparing the interventions to one another, we found that the explicit observation intervention significantly reduced the mean reported score more than the descriptive norms, moral prime, read-pledge, and type-pledge interventions. The second

most effective intervention was the observability intervention without the explicit statement. These corroborate with previous findings that suggest that moral behavior is strongly affected by being observed by others (e.g., Fries et al., 2021). The next most effective treatment was the signing group, though its results were not significantly different from any of the other interventions. Based on the findings of the present study, having participants recall the Ten Commandments, read a pledge, type out a pledge, or read a descriptive norm had little to no effect on lying behavior.

It is also important to note that this research was primarily interested in behavioral outcomes. It did not aim to evaluate the psychological mechanisms that they may use. According to Hertwig and Mazar (2022), it may be helpful to consider these psychological mechanisms, as different interventions are likely to use different psychological mechanisms. In the present study, different interventions may reduce lying behavior more effectively depending on the procedure. For example, Chou (2015) found that digital signatures are less effective in reducing dishonesty than hand signatures. They attributed this effect to digital signatures inducing less “self-presence.” As this type of behavioral research progresses, the potential effects of different psychological mechanisms are something to keep in mind.

Given the findings of the present study, there is still much to be done in order to determine which interventions work and to rank which interventions work best at curbing dishonest behavior. However, we did find sound evidence that observability decreases immoral behavior. We also provide a working methodology that will allow future research to examine different sets of interventions. Such research has the potential to provide an increasingly broad sense of the effectiveness of dishonesty-preventing interventions.

The present study was conducted as a pilot for a larger experiment. As such, several changes should be made before further investigation. Due to the nature of a pilot, there was a small sample size. A power analysis should be done on this pilot data such that subsequent studies can adequately display the effective intervention. The signing intervention is on the verge of significantly reducing dishonest behavior. With a larger sample size, this may become observable. Another minor change is that the norms treatment could be made more uniform by including it between the instructions page and the die-roll page, as most of the interventions were (except for the observability interventions). The more similar the procedures are between the interventions, the more reliable their comparisons can be.

Conclusion

The objective of the present study was to evaluate the efficacy of previously studied interventions in their ability to decrease lying behavior. In doing so, we conducted a conceptual replication study and directly compared these interventions. In order to do this, we used a standardized procedure, a consistent paradigm (die-roll), and an online survey and participant pool. The interventions evaluated included descriptive norms, observability, moral primes, reading a pledge, typing a pledge, and signing.

The findings revealed that only the observability interventions replicated their ability to reduce dishonest behavior significantly. When the interventions were compared head-to-head, the observability interventions were the most effective, followed by signing. The rest of the interventions had little to no effect on lying behavior in the given procedure.

Further research needs to be conducted with a larger participant pool and further enhancements to the procedure. With these changes, other interventions seen here, namely signing, may also be seen to significantly reduce dishonest behavior.

References

- Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and brain sciences*, 27(6), 713-730.
- Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature News*, 27.
- Boyer, P. (2008). Religion: Bound to believe?. *Nature*, 455(7216), 1038-1039.
- Cai, W., Huang, X., Wu, S., & Kou, Y. (2015). Dishonest behavior is not affected by an image of watching eyes. *Evolution and Human Behavior*, 36(2), 110-116.
- Chou, E. Y. (2015). What's in a name? The toll e-signatures take on individual honesty. *Journal of Experimental Social Psychology*, 61, 84-95.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual review of psychology*, 55(1), 591-621.
- Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, 45(6), 1321-1324.
- DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, 85(2), 1029-1069.
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind?. *PloS one*, 7(1), e29081.
- Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99(1), 197-199.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, 58(4), 723-733.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.
- Fries, T., Gneezy, U., Kajackaite, A., & Parra, D. (2021). Observability and lying. *Journal of Economic Behavior & Organization*, 189, 132-149.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin*, 145(1), 1.
- Gino, F., & Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *Journal of personality and social psychology*, 102(3), 445.

- Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational behavior and human decision processes*, 115(2), 191-203.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384-394.
- Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, 26(3), 245-256.
- Hertwig, R., & Mazar, N. (2022). Toward a taxonomy and review of honesty interventions. *Current Opinion in Psychology*, 101410.
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.
- Hull, J. G., Slone, L. B., Meteyer, K. B., & Matthews, A. R. (2002). The nonconsciousness of self-consciousness. *Journal of personality and social psychology*, 83(2), 406.
- Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102, 433-444.
- Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences*, 117(13), 7103-7107.
- 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.
- Peer, E., Mazar, N., Feldman, Y., Ariely, D. (2022, February 9-12). Honesty Pledges Reduce Cheating Through Involvement and Identification. [Paper presentation]. Society for Judgment and Decision Making The 42nd Annual Conference, Virtual.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153-163.
- Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, 109(11), 4086-4091.
- Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 374-383.

- Robert, I., & Arnab, M. (2013). Is dishonesty contagious?. *Economic Inquiry*, 51(1), 722-734.
- Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181-190.
- Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological science*, 18(9), 803-809.
- 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.
- Simonsohn, U., Nelson, L., & Simmons, J. (2021). Evidence of fraud in an influential field experiment about dishonesty. *Data Colada*, 98.
- Tomasello, Michael, and Amrisha Vaish. "Origins of human cooperation and morality." *Annual review of psychology* 64, no. 1 (2013): 231-255.
- Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., ... & Yıldız, E. (2018). Registered replication report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, 1(3), 299-317.
- Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological bulletin*, 142(5), 472.
- Wibral, M., Dohmen, T., Klingmüller, D., Weber, B., & Falk, A. (2012). Testosterone administration reduces lying in men. *PloS one*, 7(10), e46774.

Appendix A: Instructions

Surprise!

You now have the opportunity to receive an additional payoff.

However, this payoff is not the same for every participant.

You determine your own payoff by rolling a virtual die.

We ask you to please find a 6-sided die to complete this task. The die can be physical or online (you can find one by searching Google for "virtual dice").

Please, roll the die three times. Only the third roll determines your pay.

We will then ask you to report the third number that you rolled. However, you are able to report any number.

Your bonus only depends on the number you report; and you will be approved as long as you complete the study.

You can see the exact payoff for the number you report, below.

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.1. Screenshot of the instructions page for the control, moral prime, signing, read pledge, and type pledge groups.

Please locate and roll the die now.

You should roll the die three times. Only the third roll determines your pay.

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.2. Screenshot of the die-roll page for all treatment groups with the exception of the observability groups.

Please enter the third number you rolled:

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.3. Screenshot of the report page for all treatment groups.

Surprise!

You now have the opportunity to receive an additional payoff.

However, this payoff is not the same for every participant.

You determine your own payoff by rolling a die.

We ask you to please find a 6-sided die to complete this task. The die can be physical or online (you can find one by searching Google for "virtual dice").

Please, roll the die three times. Only the third roll determines your pay.

We will then ask you to report the third number that you rolled. However, you are able to report any number.

Your bonus only depends on the number you report; and you will be approved as long as you complete the study.

In previous situations, approximately 8% of participants misreported the results of similar tasks.

You can see the exact payoff for the number you report, below.

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.4. Screenshot of the instructions page for the social norms group

Surprise!

You now have the opportunity to receive an additional payoff.

However, this payoff is not the same for every participant.

You determine your own payoff by rolling a virtual die.

Please, roll the die three times. Only the third roll determines your pay.

We will then ask you to report the third number that you rolled. However, you are able to report any number.

Your bonus only depends on the number you report; and you will be approved as long as you complete the study.

You can see the exact payoff for the number you report, below.

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.5. Screenshot of the instructions page for the observability without statement group.

Surprise!

You now have the opportunity to receive an additional payoff.

However, this payoff is not the same for every participant.

You determine your own payoff by rolling a virtual die.

Please, roll the die three times. Only the third roll determines your pay.

We will then ask you to report the third number that you rolled. Although we are able to observe your rolls, you are able to report any number.

Your bonus only depends on the number you report; and you will be approved as long as you complete the study.

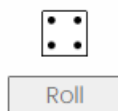
You can see the exact payoff for the number you report, below.

| | | | | | | |
|------------------|--------|--------|--------|--------|--------|--------|
| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.6. Screenshot of the instructions page for the observability with statement group.

Please roll the die now.

You should roll the die three times. Only the third roll determines your pay.



| Number rolled | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------|--------|--------|--------|--------|--------|--------|
| Resulting Payoff | \$0.01 | \$0.03 | \$0.09 | \$0.27 | \$0.81 | \$0.00 |

Figure A.7. Screenshot of the die-roll page for the observability groups.

Before completing this task please briefly (within 2 minutes) attempt to type out the Ten Commandments in the box below. Please, only use your memory to complete this task.

Figure A.8. Screenshot of the treatment page for the moral prime group.

To continue, you need to sign your initials with your mouse cursor.

Sign your initials in the box below. By signing, you certify that the information you submit in this study is complete and correct, to the best of your knowledge.



Figure A.9. Screenshot of the treatment page for the signing group.

Before completing this task please read the statement below:

I promise that I will only report the number that appears on the dice, as rolled.
I know that I will be paid based on my reporting and hence will take it very seriously to be accurate in my reporting.

Figure A.10. Screenshot of the treatment page for the read-pledge group.

Before completing this task please read and type out the statement below:

I promise that I will only report the number that appears on the dice, as rolled.
I know that I will be paid based on my reporting and hence will take it very seriously to be accurate in my reporting.

Please type out the text exactly as it appears above in the box below to move on.

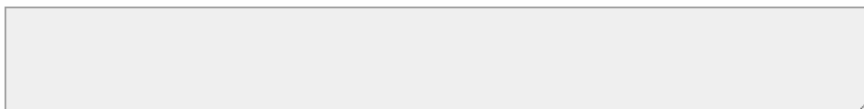


Figure A.11. Screenshot of the treatment page for the type-pledge group.