Dishonesty and Charitable Behavior *

Doru Cojoc[1] and Adrian Stoian[2]

[1]University of Iowa
[2]California State University, East Bay

December 26, 2012

Abstract

We examine in the laboratory how having the opportunity to donate to a charity in the future affects the likelihood of currently engaging in dishonest behavior. We also examine how charitable donations are affected by past ethical choices. Subjects first complete a task which determines their payoff. They then self-report their performance, which provides them with an opportunity for undetected cheating. In the second stage they are given the opportunity to donate some of the money earned in the first stage to a charity. Only subjects in the treatment group know about the opportunity to donate in the second stage. We find that more subjects cheat if they know they can donate some of the money to charity. We compare different mechanisms that lead to these results by looking at subjects' donations. We find that subjects in treatment end up donating less to charity and that both honest and dishonest subjects donate less in treatment.

*We are grateful to the editor and two anonymous referees for their extremely helpful comments and suggestions. We benefited from discussions with Dan Ariely, Michael Bailey, Doug Bernheim, Jed DeVaro, Dan Fragiadakis, Martin Gervais, Asen Ivanov, Uzma Khan, Muriel Niederle, Jonathan Meer, Tomás Rodriguez Barraquer. We would like to also thank participants at the University of Iowa, California State University, East Bay, Bay Area Behavioral and Experimental Economics workshops as well as the ESA meetings in New York and Tucson, the ESA European conference in Cologne, the WEAI conference in San Francisco and the SEA conference in New Orleans for their helpful comments. We would also like to acknowledge the financial support we received from California State University, East Bay. All errors are ours.

1 Introduction

Guilty individuals can typically engage in a wide range of good deeds in order to ease their conscience. For example, sinners seeking penance can volunteer for soup kitchens, donate blood, or raise money for the needy. Attorneys who overbill their clients can soothe their conscience by offering pro-bono legal representation. A CEO whose company pollutes can allocate some funds to green causes. Some questions arise regarding these types of situations. First, do opportunities to do good in the future affect an individuals' current ethical choices? If that is the case, do individuals end up behaving less ethically? Second, do individuals follow up and behave ethically when the opportunities to do good arrive? What is the mechanism that drives their ethical decisions? Is it the case that people who made different initial moral choices end up behaving differently in the future?

We conduct an experiment in the laboratory to shed light on the questions above. Specifically, we examine whether people cheat more when they have an opportunity to donate some of their earnings to a charity. In the first stage of our experiment we give subjects the opportunity to behave dishonestly without the possibility of being detected using a design similar with Mazar and Ariely (2006) and Gill et al. (2012), among others. Subjects self report the outcome of a random event that determines their expected payoff. They can misreport their performance in order to increase the likelihood of receiving a greater monetary payoff. In the second stage, subjects decide how much of their earnings to give to the American Red Cross. Subjects in the Treatment group know about the second stage from the beginning. Subjects in the Control group only find out they have an opportunity to donate to charity after they have completed the first stage.

Two main results emerge from our experiment. First, we find that more subjects in Treatment behave dishonestly than subjects in Control. The proportion of dishonest subjects in Control implied by their self-reports is 31.7%. In Treatment, the proportion of cheaters is 49.7%. We thus find that immoral behavior increases when people can do good in the future. Our second set of results is that on average subjects donate less when they know about the opportunity to give from the beginning. Subjects in Control donate an average of $4 out of a prize of $10; subjects in Treatment donate only $2. About one quarter of the difference in average giving is due to more subjects donating positive amounts to charity in Control than in Treatment. The difference in donations between treatments remains significant after controlling for all the available exogenous variables and for the subjects' behavior in the first stage of the experiment. Both the honest and the dishonest subjects in Treatment give significantly less than the honest and, respectively, the dishonest subjects in Control. Both in Control and in Treatment, the honest subjects give slightly more than the dishonest subjects, although these differences are not statistically significant.

We use the results on charitable giving to distinguish between two competing mechanisms that could drive our results on cheating. The fact that subjects in the Treatment group behave less honestly than subjects in Control may be a consequence of the "transgression-compliance" hypothesis (Cialdini et al. (1973)). In this framework, past transgressions enhance one's conscience leading to a desire to alleviate the guilt through more ethical behavior. If good deeds can "cleanse" past unethical choices, then having an opportunity to soothe one's conscience through charitable giving might result in less honest behavior.

We propose an alternative mechanism that can also drive our results on cheating. Instead of enhancing guilt, it may be the case that initial transgressions numbs one's conscience, resulting in smaller increases in guilt from additional violations of social norms. We call this hypothesis *conscience numbing*. Conscience numbing can take place in a few situations. For example, individuals may make moral decisions in part to maintain their self-regard,

like in Mazar et al. (2008)'s framework. If they have already done a bad deed, they may feel discouraged and give up on maintaining their self-regard, thus diminishing the consequences of further bad behavior. Conscience numbing could also appear in self-signaling models such as Benabou and Tirole (2011). Individuals may be uncertain about their own moral type and try to infer it from their past ethical choices. One's immoral behavior can have such a large impact on one's posterior beliefs that no further good deeds can change one's moral self-assessment.

While both hypotheses are consistent with more dishonest behavior in Treatment, they have very different predictions about giving in the second stage of our experiment. The transgression-compliance hypothesis predicts that cheaters would give more than honest subjects, in both Treatment and Control. It also predicts that cheaters would give more in Treatment than in Control, and that average giving would increase in Treatment compared to Control. If conscience numbing drives the increase in dishonesty, one expects to see precisely the opposite outcomes. Our results on giving reject the predictions based on the transgression compliance hypothesis, and are consistent with the predictions of conscience numbing. These results have important consequences: it appears that subjects do not cheat more in the first stage of the experiment when they know about charitable donations because they plan to give. Instead, they expect to violate social norms in both moral dilemmas. There is no trade-off between good and bad deeds in our experiment, but rather one bad deed induces people to engage in others.

Not all observed giving behavior is consistent with the conscience numbing hypothesis, however. The fact that honest subjects give less in Treatment than in Control remains puzzling. Conscience numbing, as well as the transgression compliance hypothesis, would predict that the honest subjects should be equally stingy towards the charity across treatments. We conjecture that this result is driven by a modified version of "moral licensing" (Monin and Miller (2001)) that only operates when subjects make multiple moral decisions at the same time, like in Treatment.

The results in this paper suggest that analyzing ethical behaviors separately may miss relevant links between them. In particular, institutions which diminish the guilt associated with previous immoral behavior through good deeds may lead to undesirable outcomes. They may give an excuse for engaging in immoral behavior and at the same time decrease the likelihood of future ethical behavior both among moral and immoral individuals. Kotchen (2006) finds a similar result: in his model, the introduction of green goods meant to facilitate contributions to the environment may induce some economic agents to decrease their overall public good contribution.

Our paper contributes to both the literature attempting to understand the causes and consequences of dishonest behavior and to the one analyzing charitable giving. Both dishonesty and giving have been extensively analyzed separately in the past.[1] Besides our paper we are aware of only one other study in the economics literature that looks at these behaviors jointly. Gneezy et al. (2012) ask how the opportunity to give to charity affects the likelihood of subjects deceiving other subjects for personal gain. Like us, they find statistically significant differences in dishonesty between subjects who were informed about charitable giving and uninformed subjects. In contrast to our paper, the subjects in their experiment knew the experimenter was able to observe their lying. The main focus of their study is on how the amount of time that has elapsed from the initial moral decision affects subsequent donations, so the results on charitable giving are not directly comparable to ours.

The rest of the paper is organized as follows. In the next section we describe the experimental design. In Section 3 we present the results on the cheating behavior. In Section 4 we contrast two models of moral behavior that

[1] Papers in economics that look at the dishonest behavior include Gneezy (2005), Hurkens and Kartik (2009), Lundquist et al. (2002). For a survey of the literature on charitable giving from an economic perspective see Andreoni (2006).

both predict the cheating behavior we observe but which have opposite predictions regarding charitable behavior. We empirically distinguish between these models in Section 5 where we discuss the results on donations. The last section concludes.

2 Experimental Design

The experiment took place at the end of 2011 and during 2012 at Stanford University and California State University, East Bay. Participants at Stanford were recruited using the available subject pool. Participants at California State were recruited using mass emails sent to undergraduate students. We ran 10 sessions at Stanford and 16 at California State. On average there were 6 participants per session, including the one who rolled the dice 80 times as described below. In each session there was an approximately equal number of subjects randomly assigned to each of the two treatments. In total there were 66 subjects in Treatment and 64 subjects in Control. The average age of participants was 23.11, 46.92% were female, and 61.53% pursued a business-related major.

The experiment lasted in total approximately 50 minutes and had two stages. Besides the participation fee of $5, subjects could earn in the first stage a payment of either $10 or $25, out of which they could donate in the second stage any integer amount to American Red Cross. Subjects earned on average $16.57, excluding the participation fee, and donated on average $4.14. Subjects were paid with Amazon gift cards.

In the preliminary part of the experiment, subjects were randomly assigned to either Control or Treatment groups and were given an ID number based on which the payment was made. Subjects were informed that the data will be collected completely anonymously and they were instructed to carefully and completely read the instructions (see Appendix). The instructions for the two groups were identical except for one detail: the instructions for the Control group did not disclose the fact that subjects will have the opportunity to later on donate part of their earnings in the first round to a charity. Both instructions for Treatment and for Control groups started with some summary information about American Red Cross. This was done to prime subjects in both groups towards prosocial behavior. After all subjects finished reading the instructions, the experimenter answered the questions they had and then instructed them to start the software application on their computers. Throughout the experiment there was no interaction between subjects or between subjects and the administrator.

In the first stage of the experiment subjects were asked to roll three dice and to input the outcome of the roll in the computer. Subjects rolled the dice in a box with cloth on the bottom to attenuate noise. At each round the computer informed them of the probabilities to obtain $25 and respectively $10 if they reported rolling three fives in that round. The probability to be paid $25 if reported 5-5-5 was 90% in the first round and decreased linearly with each round. If a subject recorded a roll of three fives, the first stage of the experiment ended and the second stage started. Subjects were neither monitored, nor were they asked to prove that they rolled 5-5-5. If a subject reported a different outcome, the computer prompted them after a delay of 12 seconds to roll the dice again. The first stage of the experiment ended when a subject either reported rolling 5-5-5 or rolled the dice 80 times.

In the second stage, subjects were informed (those in the treatment group were only reminded) that they have the opportunity to donate a part of the reward from the first stage to American Red Cross. Using the strategic method, they were asked how much they would donate if they were to receive $25 and how much they would donate if they were to receive $10. After subjects entered the two amounts, the computer randomly chose their reward with a probability depending on the round at which they stopped rolling, and implemented their corresponding donation

decision. After that the computer displayed the compensation in the first stage (either \$25 or \$10) and the final compensation, which was equal to their compensation in the first stage, minus their corresponding donation, plus \$5 show up fee. Subjects were instructed to sit quietly at their computers until the administrator would announce that the experiment is over. They were allowed to use the internet at this point.

One randomly chosen subject, which we omit from the analysis, had to roll the three dice 80 times no matter what she reported. Subjects were informed in the instructions that other subjects might have to roll the dice 80 times. This guaranteed more uniformity across sessions. No matter their individual decisions, subjects spent roughly the same amount of time in the lab and they heard at least somebody rolling the dice 80 times in all sessions.

After everyone finished, the administrator informed the participants of their total donations and randomly picked one subject to monitor the issuing and mailing of the check to the American Red Cross. For this extra task, the monitor was compensated with an additional \$5. The experiment ended with the payment of the subjects.

3 Results on Dishonest Behavior

The first set of results we present regard subjects' dishonest behavior. We wanted to observe undetected cheating; at the same time, we did not want to deceive the subjects who thought they were cheating without the possibility of being detected. Indeed we do not know with certainty which subjects cheated and which did not, so in our analysis we rely on average group behavior. A few details of the first stage were meant to encourage subjects to cheat. In particular, subjects were not monitored and were not required to prove they rolled three fives. We delayed moving to the next round after failing to roll three fives by 12 seconds in order to render the task onerous and to diminish any excitement from rolling the dice, as well as to guarantee that subjects rolled roughly at the same pace. The "performance" to be reported by subjects is entirely random, so that one who is unlucky and fails to roll 5-5-5 may find it easier to rationalize reporting a success. In these conditions, we should not be surprised if some subjects chose to cheat. We are however interested in whether there is any difference in dishonesty across treatments.

Figure 1 plots the cumulative distribution function of the round in which a subject has reported rolling 5-5-5 for subjects in Control (solid red) and subjects in Treatment (dashed blue). The cumulative distribution function for subjects in Treatment first-order stochastically dominates the one for subjects in Control, suggesting that on average subjects in Treatment are reporting a successful roll earlier than subjects in the control group. Four complementary formal statistical tests support the hypothesis that the subjects in Treatment are less honest than the subjects in Control.

On average, subjects in Control report rolling 5-5-5 after 52.25 rounds of rolling the dice; subjects in Treatment roll the dice 44.56 times on average before reporting a success. The difference of 7.68 rounds is statistically significant in a non-parametric two-sample Mann-Whitney test ($z = 1.672$, p-value of 0.094). We also run a Cox proportional hazard model of the round in which a subject reported rolling 5-5-5. In addition to whether or not a subject was in the treatment group, we control for the subjects' gender, age, location (Stanford or California State), major (business-related or other), and the number of subjects in the session. The coefficient on Treatment is 1.47 with a p-value of 0.10, so conditional on the available covariates, we reject the hypothesis that subjects in Treatment are at the same "risk" of stopping rolling the dice as subjects in the control group.

Figure 2 shows histograms of the rounds to success in Control (in red, on the left) and in Treatment (in blue, on the right). Figure 2 reveals that the starkest difference between treatments take place at the very small and the very

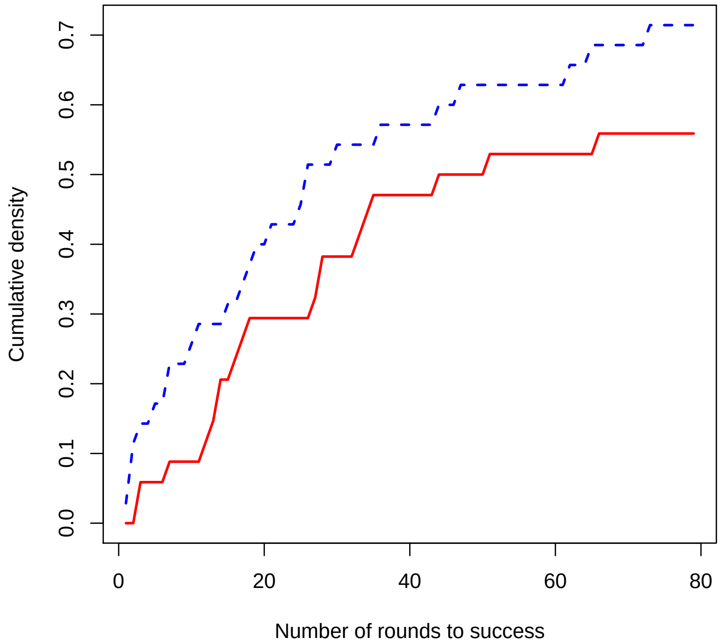

Figure 1: The cumulative distribution functions of "rounds to success" in control (solid red) and treatment (dashed blue)

large values of the round in which a success is reported. Subjects who never report rolling 5-5-5 are certainly honest, albeit unlucky. The proportion of these certainly honest subjects is larger in Control than in Treatment. On the other hand, the subjects who report a success very early are most likely dishonest. There are more of this type of subjects in Treatment than in Control.

We formalize these impressions with two tests of equality of proportions. If subjects were equally dishonest across treatments, the proportion of subjects who never report a success and of subjects who report a success very early should be similar in Control and Treatment. The proportion of subjects who stopped rolling before the 80th round is 50% in Control and 62% in Treatment. A Pearson Chi-squared test rejects the hypothesis of this difference being caused by mere chance at 10-percent significance level ($z = -1.39$ with a one-sided p-value of 0.08). 9% of subjects who did not know about the opportunity to give to charity finished rolling in the first 10 rounds; in Treatment, 18% of subjects reported a success within the first 10 rounds ($z = -1.47$ with a one-sided p-value of 0.07).

Since not all sessions had subjects reporting an early success, we can supplement the previous analysis by looking only at the subject who reported the first success in each session. Out of 26 sessions, subjects in Control finish first in 8 of them, with subjects in Treatment finishing first in the remaining 18 sessions. If subjects in Control were no more honest than subjects in Treatment, this result would obtain around 5% of times, so we can reject the hypothesis of identical dishonest behavior across treatments.

The four complementary statistical tests establish the following main result:

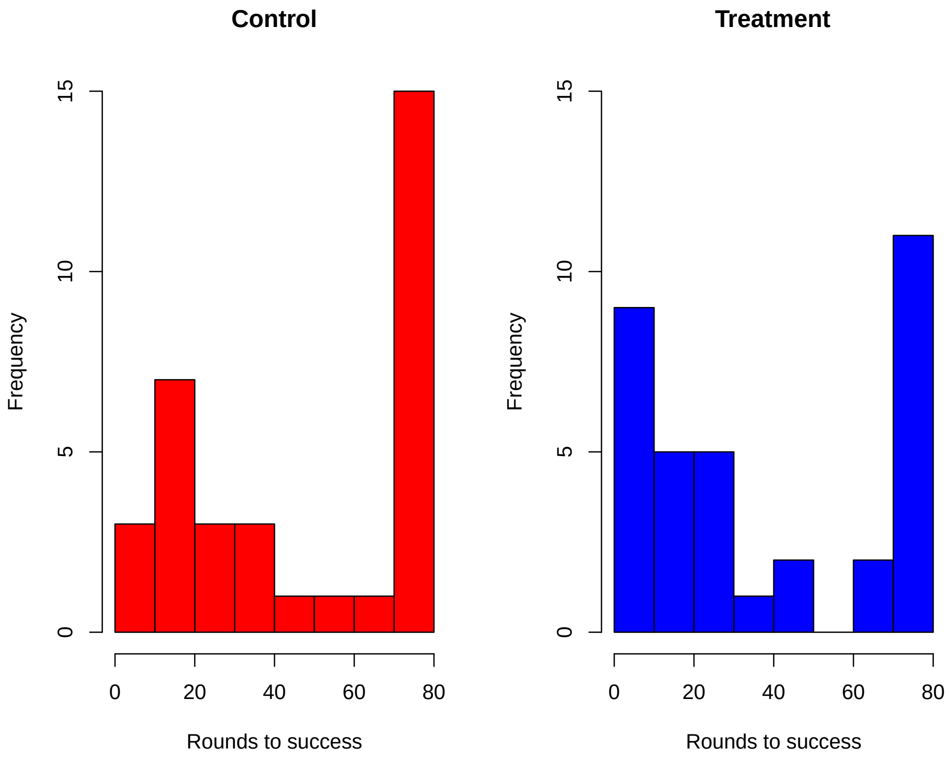

Figure 2: Histograms of the round to success in Control (red, left) and in Treatment (blue, right)

Result 1: On average, the subjects who knew about the opportunity to give to charity behave less honestly than the subjects who did not know about it.

We decompose the overall increase in dishonesty in Treatment into an "extensive" and an "intensive" margin. There are clearly fewer subjects who behave honestly in Treatment than in Control. But once they decided to cheat, are subjects in Treatment systematically different from subjects in Control? Although there are significantly more subjects in Treatment who report rolling 5-5-5 within the first 10 rounds, on average the difference in the round in which subjects reported rolling a 5-5-5 is not statistically significant. "Successful" subjects in Treatment report rolling three fives around 1.6 rounds earlier than the "successful" subjects in Control. This difference is not significant using a Mann-Whitney non-parametric test ($z = 0.973$ with p-value$= 0.33$). Therefore we can establish a secondary result:

Result 2: The difference between cheating in Treatment and Control is mainly due to the "extensive" margin: fewer people remain honest in Treatment, but once they decided to cheat, subjects are on average no less honest in Treatment than in Control.

4 Model

In this section we begin analyzing why the opportunity to donate to charity induces some people to behave less honestly. We develop a stylized model of behavior in which an individual's ethical choices depend on the guilt she feels when violating social norms. We formalize the two hypotheses that are consistent with our results on dishonesty, the transgression-compliance hypothesis and *conscience numbing*. These hypotheses generate opposite predictions regarding subjects' charitable donations. We test these predictions in the next section.

Subjects are indexed by their type $\theta > 0$ which can be interpreted as their general intolerance to violations of social norms. The value of this parameter is unobserved, but we assume that we obtain the same distribution of θs in both treatments. Subjects derive disutility (guilt) from violating social norms. The guilt of a type θ subject can be written as $G(c, s; \theta) = \theta \cdot c + g(\theta \cdot c, s)$, with $c \in \{0, 1\}$ representing cheating and $s < T$ denoting stinginess towards the charity relative to some social norm T. The first term represents the guilt from cheating in the first stage of the experiment. The second represents the guilt from stinginess in the second stage of the experiment. The second term depends on whether the subject cheated and, if she cheated, on her type. Thus, we implicitly assume that all subjects who did not cheat experience the same guilt from any degree of stinginess. We assume that g increases in both arguments and that it is convex in s.

Consider first the behavior of subjects in Control. Subjects are not aware of the opportunity to give to charity, so in the first stage their guilt function only contains the first term. A subject chooses $c \in \{0, 1\}$ to maximize

$$w + c \cdot (W - w) - \theta \cdot c, \tag{1}$$

where $W - w$ is the additional expected payoff one obtains by cheating.[2] Thus, subjects with $\theta < \theta_C = W - w$ cheat and the rest do not. In the second stage, subjects learn about the opportunity to give to a charity. This opportunity carries an implicit requirement to satisfy social norms on charitable behavior. If a subject donates the amount T she feels she is required, her stinginess towards charity is $s = T - d = 0$, but any smaller d increases her stinginess and thus her guilt. When deciding how much to donate, subjects implicitly chooses the level of stinginess s in order to maximize

$$w + c \cdot (W - w) - (T - s) - \theta \cdot c - g(\theta \cdot c, s). \tag{2}$$

Thus the type θ subject who chose c in the first stage of the experiment chooses stinginess $s^*(c; \theta)$ given by $1 - \frac{\partial g(\theta \cdot c, s)}{\partial s} = 0$.

By implicit function theorem and convexity of g in s, two results emerge:

$$\frac{\partial s^*}{\partial \theta}\Big|_{c=0} = 0 \tag{3}$$

$$\text{sign}\left(\frac{\partial s^*}{\partial \theta}\Big|_{c=1}\right) = -\text{sign}\left(\frac{\partial^2 g}{\partial \theta \, \partial s}\right) \tag{4}$$

Subjects who did not cheat are equally stingy regardless of their type. Whether a larger θ makes the subjects who cheated stingier or less stingy depends on what one assumes about the properties of function $g(\theta \cdot c, s)$. We examine the consequences of two assumptions:

[2] In light of Result 2, we ignore the decision when to cheat for subjects who decide to cheat.

Assumption 1 (Transgression-Compliance Hypothesis): The function $g(\theta \cdot c, s)$ exhibits increasing differences; that is, $\frac{\partial g(b,s)}{\partial s} > \frac{\partial g(a,s)}{\partial s}$ for all $b > a \geq 0$ for all degrees of stinginess s.

To see why Assumption 1 captures the transgression-compliance hypothesis, consider the type θ subject in Control. According to the hypothesis, had she cheated, she would have had additional reasons to donate to charity than if she had not cheat. We formalize this by assuming that reducing one's stinginess reduces guilt at a faster rate along the curve $g(\theta, s)$ than along the curve $g(0, s)$. In other words, one's guilt from cheating enhances the additional guilt one feels when one is stingy towards the charity. The hypothesis also implies that the larger one's intolerance to violations to social norms, the more reasons one would have to give in order to alleviate the initial guilt from dishonest behavior.

We consider an alternative hypothesis we call *conscience numbing*. It may be that initial cheating numbs one's conscience, resulting in smaller increases in guilt from additional violations of social norms. A conscience numbing effect may take place in a few frameworks of moral behavior. If, for example, individuals make ethical decisions to maintain their self-regard, like in (Mazar et al. (2008)), then past unethical behavior may make them feel discouraged. They may give up on maintaining their self-regard, lowering the consequences of further bad behavior. Conscience numbing could also appear in various self-signaling models, such as Benabou and Tirole (2011). The posterior beliefs about one's moral type could be heavily influenced by one's initial immoral behavior. If individuals believe that no further good deeds can change them much, additional unethical behavior would have a diminished impact on one's welfare.

Assumption 2 (Conscience Numbing): The function $g(\theta \cdot c, s)$ exhibits decreasing differences; that is, $\frac{\partial g(b,s)}{\partial s} < \frac{\partial g(a,s)}{\partial s}$ for all $b > a \geq 0$ for all degrees of stinginess s.

The two assumptions have distinct predictions about stinginess is affected by a subjects' type. Under transgression-compliance hypothesis, the types $\theta \in (0, \theta_C)$ who cheated in the first stage are less stingy as θ increases. The types who decided to be honest in the first stage are more stingy than the types that cheated. The opposite happens if conscience numbing operates: the types $\theta \in (0, \theta_C)$ who cheated in the first stage are stingier as θ increases, and the honest subjects are less stingy than the dishonest ones.

Consider next the subjects in Treatment. The only difference is that they know about charitable giving from the beginning, so they choose c and s to maximize Equation 2. Two results emerge: first, if subject θ is equally honest in Control and in Treatment, she will be equally stingy in both treatments. Second, the difference in stinginess across treatments is entirely driven by the marginal subjects, the subjects that would cheat in one treatment but wouldn't in the other.

Result 1 in the previous section showed that the number of subjects who cheat is larger in Treatment than in Control. This could take place under both assumptions. Consider the θ_C subject who is indifferent between cheating and not cheating in Control. In Treatment, she would choose c and s to maximize $w + c \cdot (W - w) - (T - s) - \theta_C \cdot c - g(\theta_C \cdot c, s) = w - T + s - g(\theta_C \cdot c, s)$. She will choose to cheat if $s^*(1; \theta_C) - g(\theta_C, s^*(1; \theta_C)) > s^*(0; \theta_C) - g(0, s^*(0; \theta_C))$. This relation could hold under both assumptions, although it is somewhat more plausible under Assumption 2, because under Assumption 1 it would only take place if $g(0, s^*(0; \theta_C)) > g(\theta_C, s^*(1; \theta_C))$. If the relations holds, then in Treatment types $\theta < \theta_T$ cheat and the rest do not, where $\theta_T > \theta_C$. The marginal cheaters

are less stingy under Assumption 1 and stingier under Assumption 2, so we can derive the following predictions about subjects' charitable giving under each hypothesis.

Predictions of Transgression-Compliance Hypothesis:

1. Subjects donate more in Treatment than in Control.
2. Dishonest subjects donate more in Treatment than in Control.
3. Dishonest subjects donate more than honest subjects, both in Control and in Treatment.

The conscience numbing hypothesis has exactly the opposite predictions.

5 Results on Charitable Donations

We report our results on subjects' donations and relate the findings with the predictions of the two hypotheses discussed in the previous section. Figure 3 suggest that subjects in Control donated more out of both amounts than subjects in Treatment. Some subjects chose to give their entire earnings to charity, with more of this type of subjects in Control. Also, there are more subjects who chose not to donate any positive amount to charity in Treatment than in Control. We first ask whether the difference in giving across treatments is statistically significant and to what extent it is driven by differences in extreme behavior between groups.

Table 1 shows the intercept and the coefficient corresponding to Treatment in ordinary least squares (odd columns) and Tobit censored from below and above (even columns) regressions of giving out of 10 dollars (top part of the table) and out of 25 dollars (bottom part). Because they are easier to interpret, we refer in the text to the coefficients in the odd columns.[3]

Column (1) shows the average donation in Control and Treatment unconditional on the rest of covariates. A subject who was not aware of the opportunity to donate ends up donating on average 4 out 10 dollars, and 8 out of 25 dollars.[4] A subject who knew about the opportunity to donate ends up donating significantly less: 2 dollars out of 10 and 5 dollars out of 25. The coefficients in column (3) show that the difference remains statistically significant and is not caused by subjects having different exogenous characteristics in the two samples. In columns 5-12 of Table 1 we also control for whether a subject never reported a success in the first stage of the experiment (indicating an honest subject) or whether she reported a success very early (indicating a most likely cheater). The coefficients on Treatment remain large and statistically significant in all these specifications. We thus establish the following main result:

Result 3: On average, subjects in Treatment donate less to charity than subjects in Control.

To what extent is the difference in giving driven by extreme values for donations? Some subjects donate the entire available amount to charity, with more subjects of this type in Control than in Treatment. Running separate

[3]All the relevant coefficients from Tobit regressions except one have the same sign and same statistical significance as the coefficients from OLS regressions, so they would support the same data analysis.

[4]This is roughly double than the average donation in a dictator game reported by Camerer (2003), but it is similar to donations to charity in Carpenter et al. (2008).

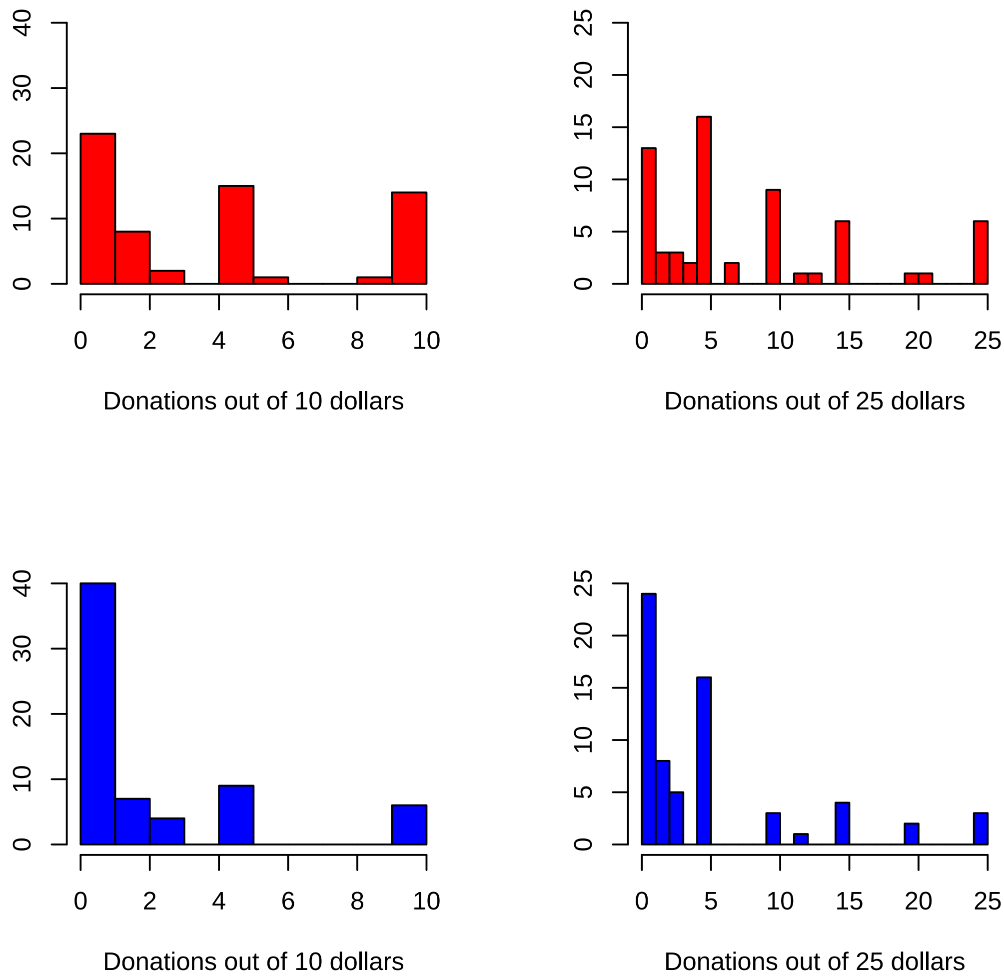

Figure 3: Amount donated to charity in Control (red) and in Treatment (blue)

OLS regressions for the subset of subjects who did not donate everything to charity still reveals large and significant coefficients on Treatment (-1.0 with p-value 0.02 for donations out of 10 dollars and -2.3 with p-value 0.02 for donations out of 25 dollars).[5] Comparing these coefficients with the ones in Column (3) suggests that Result 3 is not driven entirely by a few subjects exhibiting an unusual desire to give. We also decompose the overall difference in giving into an "extensive" and an "intensive" effect by looking at the behavior of subjects who donate amounts larger than zero. Running separate OLS regressions for this subset of subjects yields large and significant coefficients on Treatment (-1.47 with p-value 0.05 for donations out of 10 dollars and -2.75 with p-value 0.05 for donations out of 25 dollars).

Result 4: Approximately one quarter of the difference in donations between Treatment and Control is due to the "extensive" margin: more subjects fail to donate any positive amount to charity in Treatment than in Control. The

[5]We use the same set of covariates as in Table 1. We do not report these and the next set of regressions, but they are available upon request.

rest is due to the “intensive” margin: subjects donating positive amounts, but less on average in the Treatment group. Not all of the “intensive” margin is caused by fewer subjects donating their entire amount to charity in Treatment.

In order to distinguish between various mechanisms that drive our findings we would have to examine how honest and dishonest subjects give in Treatment compared to Control. While we do not know for sure which subject misreported the dice rolls, we can safely assume that the subjects who never reported a successful roll were honest. These subjects have no guilt-related reasons to donate to charity, so we would like to see whether their giving is similar across treatments. The small and statistically insignificant coefficients on the interaction term in column (7) of Table 1 suggest that there are actually differences in giving between treatments among honest subjects.

Donations out of 10 dollars

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Const.	4.06	3.89	3.09	1.16	2.35	0.49	3.06	0.97	3.24	1.45	3.31	1.63
	(0.42)***	(0.81)***	(1.78)*	(3.41)	(1.51)	(2.90)	(1.81)*	(3.45)	(1.76)*	(3.34)	(1.77)*	(3.34)
Treat. (T)	−1.93	−3.70	−2.06	−3.91	−1.95	−3.73	−2.24	−4.02	−1.89	−3.61	−1.76	−3.25
	(0.60)***	(0.80)***	(0.59)***	(1.14)***	(0.60)***	(1.14)***	(0.79)***	(1.49)***	(0.59)***	(1.12)***	(0.63)***	(1.17)***
No Success (NS)					0.87	1.30	0.54	0.98				
					(0.61)	(1.13)	(0.84)	(1.54)				
NS x T							0.68	0.69				
							(1.19)	(2.22)				
Early Succ. (ES)									−1.77	−3.39	−1.05	−1.61
									(0.85)**	(1.65)**	(1.43)	(2.54)
ES x T											−1.11	−3.10
											(1.81)	(3.38)

Donations out of 25 dollars

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Const.	8.07	7.68	5.00	2.73	4.42	1.81	4.73	2.34	5.21	3.02	5.56	3.58
	(0.88)***	(1.18)***	(3.71)	(5.03)	(3.68)	(4.97)	(3.74)	(5.03)	(3.70)	(4.99)	(3.69)	(4.49)
Treat. (T)	−3.04	−4.44	−3.48	−5.02	−3.16	−4.57	−3.69	−5.51	−3.25	−4.72	−2.58	−3.65
	(1.23)**	(1.70)***	(1.23)***	(1.65)***	(1.23)**	(1.63)***	(1.63)**	(2.19)**	(1.23)***	(1.65)***	(1.31)	(1.73)
No Success (NS)					2.50	3.63	1.91	2.63				
					(1.25)**	(1.65)**	(1.73)	(2.26)				
NS x T							1.22	2.11				
							(2.46)	(3.24)				
Early Succ. (ES)									−2.43	−3.63	1.13	1.69
									(1.79)	(2.42)	(2.99)	(3.81)
ES x T											−5.59	−8.93
											(3.77)	(5.01)*
Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	130	130	130	130	130	130	130	130	130	130	130	130

Table 1: Ordinary least squares (odd columns) and Tobit censored from below and above (even columns). The dependent variable is donation out of 10 dollars (top part) and out of 25 dollars (bottom part) for all subjects in the experiment. No success takes a value of 1 if the subject did not report rolling 5-5-5 before the end of the first stage, 0 otherwise. Early success takes a value of 1 if the subject reported rolling 5-5-5 in the first 10 rounds of the first stage, 0 otherwise.Control variables are age, gender, Stanford University, business major and number of subjects per session. ***, **, and * represent significance at 1, 5, and 10 percent level.

Donations out of 10 dollars						
	All subjects		Unsuccessful subjects		Successful subjects	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3.09	1.16	5.41	3.87	5.62	4.55
	(1.78)*	(3.41)	(2.94)*	(5.83)	(2.82)*	(4.75)
Treatment	−2.06	−3.91	−1.76	−4.12	−1.99	−3.32
	(0.59)***	(1.14)***	(0.99)*	(2.06)*	(0.72)***	(1.25)***

Donations out of 25 dollars						
	All subjects		Unsuccessful subjects		Successful subjects	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.00	2.73	8.44	5.96	9.33	7.42
	(3.71)	(5.03)	(6.21)	(7.73)	(5.95)	(7.94)
Treatment	−3.48	−5.02	−2.52	−3.37	−3.26	−4.91
	(1.23)***	(1.65)***	(2.10)	(2.57)	(1.52)**	(2.09)**
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	130	130	57	57	73	73

Table 2: Ordinary least squares (odd columns) and Tobit censored from below and above (even columns). The dependent variable is donation out of 10 dollars (top part) and out of 25 dollars (bottom part) for all subjects in the experiment. Control variables are age, gender, Stanford University, business major and number of subjects per session. ***, **, and * represent significance at 1, 5, and 10 percent level.

We report in Table 2 the coefficients of ordinary least squares and censored Tobit regressions for all the subjects pooled together and separately for the subjects who did not report a success and those who did. The top coefficients of Columns (3) and (4) in Table 2 show that an honest subject donates a smaller amount to charity out of 10 dollars if she learned about the opportunity to give before the first stage. These coefficients are significantly different from zero at the usual significance levels. The bottom coefficients in the same columns show similar magnitudes and suggest the same behavior, but they are estimated with less precision. We can therefore report the following result.

Result 5: Subjects who did not cheat give less in Treatment than in Control.

Column (5) in Table 2 shows that subjects who reported rolling a 5-5-5 donate a smaller amount to charity out of both earnings in Treatment compared to Control. Not all subjects who report a success are dishonest, so we cannot conclude immediately that dishonest subjects are stingier in Treatment than in Control. We use the estimates in columns (3) and (5) of Table 2, as well as the observed proportions of subjects who fail to report a success, to compute the difference in giving across treatments among dishonest subjects. We then use Monte Carlo simulations to derive its precision.[6] We obtain a statistically significant estimate of -1.92 for the effect of Treatment on donations out of 10 dollars among dishonest subjects. Like it is the case with donations out of 25 dollars of honest subjects, the coefficient on Treatment among dishonest subjects for donations out of 25 dollars is economically significant (-2.95) but it is estimated with slightly less precision.

Result 6: The subjects who reported a success give less in Treatment than in Control. This is indicative of subjects

[6]The details of these calculations are presented in Appendix 2.

who cheated giving less in Treatment than in Control.

Differences in realized giving between subjects in Control and Treatment

	Donation if 10	Donation if 25	Won 10	Won 25	Actual donation
Control	4.06	8.07	39	25	4.82
Treatment	2.13	5.03	34	32	3.48
Difference	1.92	3.04	5	-7	1.34
p-value	0.001	0.003	0.14		0.007

Table 3: Mann-Whitney U and Pearson' equality of proportion test results

Finally, Table 3 shows the difference in realized giving between subjects who knew about the possibility to give to charity and subjects who did not. While subjects in Treatment give less both out of 10 and out of 25 dollars than subjects in Control, they are more likely to earn the larger amount. Do subjects in Treatment end up giving more to charity simply because they earn the larger amount more often? It turns out that subjects in Treatment end up giving less than subjects in Control. Subjects in Control end up giving \$1.34 more than subjects in Treatment; the difference is statistically significant in a Mann-Whitney test at 1-percent significance level. Controlling for other variables, the coefficient on Treatment is -1.45, still statistically significant at 10-percent significance level. This establishes the last result of the paper:

Result 7: Although they earn a larger amount more often, subjects in Treatment give less than subjects in Control.

5.1 Discussion

Our results on charitable giving reject all predictions derived from the transgression-compliance hypothesis. We find (Result 3) that on average charitable giving decreases in Treatment compared to Control, which goes against the idea of subjects using charitable donations to soothe their larger guilt. This contradicts Prediction 1 of transgression-compliance hypothesis. We also find (Result 6) that dishonest subjects give less in Treatment than in Control, which contradicts Prediction 2 of transgression-compliance hypothesis. Result 3 and 6 are precisely the predictions of our alternative hypothesis, that an initial transgression numbs one's conscience, resulting in more transgressions.

Although we do not have a formal statistical test, a few results strongly indicate that Prediction 3 of transgression-compliance hypothesis does not hold either. In column (5) of Table 1, the coefficient on subjects who did not report a success is positive, although not statistically significant. So subjects who did not report a success, and who are certainly honest, give more on average to charity than the subjects who reported a success, some of whom are honest and some dishonest. This can only happen if the dishonest subjects give less to charity than the honest ones. Column (9) of the same table reports the coefficient on subjects who reported a success very early. These are most likely cheaters, and they give significantly less than the rest of the subjects. Again, this result can only happen if the dishonest subjects give less than the honest ones.

Our Result 5, that the honest subjects give less in Treatment than in Control, is surprising. Both hypotheses would predict that the charitable giving of honest subjects is not affected by subjects knowing about the opportunity to donate from the beginning. This strong result is obtained in the model from a strong assumption that all honest subjects experience the guilt from stinginess in the same way. Even after allowing heterogeneity in this guilt function,

neither the transgression-compliance hypothesis nor conscience numbing would predict Result 5 for a simple reason: since the honest subjects did not cheat, their guilt would not be systematically enhanced or numbed.

We offer a possible rationalization for our puzzling finding based on the idea of "moral licensing". Monin and Miller (2001) found that subjects regard past ethical decisions as a license to engage in less moral behavior. Closer to the behaviors examined here, Sachdeva et al. (2009) find that subjects whose moral identity was threatened express a stronger desire to donate to charity than subjects whose moral identity was enhanced. We conjecture that in our data moral licensing only operates when subjects make multiple moral decisions at the same time. This could be justified if subjects try to maintain their self-regard whenever they make moral decisions, once in Treatment but twice in Control. In Treatment, subjects decide whether to be honest and how much to give at the same time. A subject who just decided to be honest might feel less compelled to give in order to maintain her self-regard. On the other hand, the benefits to self-regard a subject in Control may have derived from her initial decision to be honest are "sunk" when she decides how much to give. In order to maintain her self-regard, she may feel the need to give more compared to honest subjects in Treatment. If social norms change in this fashion, subjects could donate less in Treatment than in Control even when their stinginess remained roughly the same across treatments.

6 Conclusions

We conduct a two-stage experiment in the laboratory to examine how the opportunity to donate to a charity in the second stage affects the likelihood of dishonest behavior in the first stage and the subjects' subsequent charitable donations. Our results show that fewer subjects behave honestly when they are aware of a future opportunity to donate to a charity. We also find that when subjects are aware of the opportunity to donate to charity they give significantly less. Both the honest subjects, and the dishonest subjects give less when they know about the opportunity to donate in the second stage, even though the latter were slightly more dishonest in the first stage than the cheaters who did not know about charitable giving.

Our findings do not support the transgression-compliance hypothesis. Subjects do not appear to give more in order to alleviate the guilt they feel from cheating. They also do not seem to always regard past honest behavior as a license to give less to charity. Our results are consistent with the assumptions we propose about moral behavior. First, initial transgressions appear to lower, rather than increase, the marginal guilt of additional violations of social norms. Second, moral licensing only seems to apply when people make multiple moral decisions at the same time. This has important consequences: initial transgressions do not result in future compliance to social norms but rather in even more transgressions.

The results of our paper emphasize that ethical and unethical behavior should not be studied in isolation. In certain situations there could be significant interactions between these behaviors. Policies designed to promote the former and discourage the latter based only on studies that look at these behaviors in isolation could lead to adverse outcomes.

References

Andreoni, J. (2006). Philantropy. In L. A. Gerard-Varet, S. C. Kolm, and J. M. Ythier (Eds.), *Handbook of Giving, Reciprocity and Altruism*, pp. 1201–69. North-Holland: Elsevier.

Benabou, R. and J. Tirole (2011). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Carpenter, J., C. Connolly, and C. Myers (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics 11*(3), 282–298.

Cialdini, R. B., B. L. Darby, and J. E. Vincent (1973). Transgression and altruism: A case for hedonism. *Journal of Experimental Social Psychology 9*(6), 502516.

Gill, D., V. Prowse, and M. Vlassopoulos (2012). Cheating in the workplace: An experimental study of the impact of bonuses and productivity.

Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review 95*, 384–394.

Gneezy, U., A. Imas, and K. Madarasz (2012). Conscience Accounting: Emotional Dynamics and Social Behavior.

Hurkens, S. and N. Kartik (2009). Would I Lie to You? On Social Preferences and Lying Aversion. *Experimental Economics 12*, 180–192.

Kotchen, M. (2006). Green Markets and Private Provision of Public Goods. *Journal of Political Economy 114*(4), 816–834.

Lundquist, T., T. Ellingsen, E. Gribbe, and M. Johannesson (2002). The Aversion to Lying. *Journal of Economic Behavior and Organization* (70), 81–92.

Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* (45), 633–644.

Mazar, N. and D. Ariely (2006). Dishonesty in Everyday Life and Its Policy Implications. *Journal of Public Policy and Marketing* (25), 117–126.

Monin, B. and D. T. Miller (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology 81*(1), 33–43.

Sachdeva, S., R. Iliev, and D. L. Medin (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science 20*(4), 523–528.

Appendix 1: Instructions

Please read the following instructions carefully. You should not communicate with other participants in the room at any time during the experiment. If you have questions raise your hand and the administrator will assist you.

You will make individual decisions and your payment will depend entirely on your decisions; what other participants do will have no effect on you. In addition to the money you will make in the experiment we will pay you $5 as a show up fee.

All decisions you make are recorded only by the anonymous subject number you received and will only be used for research purposes. Your decisions will remain completely anonymous.

Before describing what the experiment consists of, you are asked to read the information on the following page. After you are done reading it, please move on to the rest of the instructions.

The American Red Cross is empowering people to perform extraordinary acts in the face of emergencies. A hot meal delivered to victims after a disaster, blood when it is needed most, shelter when there is nowhere else to turn, an emergency message delivered to a member of the Armed Forces from their family – these are just some of the ways in which the American Red Cross helps. The American Red Cross is a volunteer-led, humanitarian organization. Each year, it responds immediately to more than 70,000 disasters, including fires, hurricanes, floods, earthquakes, tornadoes. When a disaster threatens or strikes, the Red Cross provides shelter, food, and health services to address basic human needs. In addition, the Red Cross provides assistance to individuals and families affected by disaster to enable them to resume their normal daily activities independently. Here are some examples of how the American Red Cross uses its funds to help disaster victims:

- $25 provides five blankets at an emergency shelter.
- $75 can cover a doctor's visit for an individual injured in a disaster.
- $350 provides emergency food and shelter for 25 disaster victims for one day.
- $2,500 deploys one Emergency Response Vehicle and drivers (including housing and meals for drivers) to a disaster relief operation.

There are two stages of the experiment. The stages are independent and your decisions in one stage will not affect your payoff in the second stage.

For the first stage you will use the three dice and the box you find on your desk. You are asked to throw the three dice simultaneously in the box. Your objective is to get three 5s (5, 5, 5). After each throw you are asked to input in the computer the three numbers you obtained, click "Continue", and wait until you are instructed to roll the dice again. If you input three 5s, you will be prompted to the second stage of the experiment. If you did not input three 5s, the computer will ask you to input the results of the new throw. You can throw the three dice as many as 80 times. Some subjects in the room are asked to throw the three dice 80 times regardless if they obtain three 5s or not.

Your payment could be influenced by how quickly you get the three 5s. You will be rewarded with either $25 or $10. The amount will be randomly determined by the computer and the sooner you obtain three 5s the more likely it is you will be rewarded with $25. The table below describes how your reward will be determined:

If you get three 5s in	Your chance of getting $25 is:	Your chance of getting $10 is:
the first throw	90%	10%
the second throw	89%	11%
the third throw	88%	12%
...	...	...
the 32nd throw	59%	41%
...	...	...
the 79th throw	11%	89%
the 80th throw	10%	90%

If you do not get three 5s after 80 throws the computer will reward you the same way as if you would have got the three 5s in the 80th throw (you will have a 10% chance of getting $25 and a 90% chance of getting $10).

After you either input three 5s or you reach the limit of 80 throws, you will be asked to follow the instructions on the screen corresponding to the second stage.

For Subjects in Control:

If you finish the experiment before other subjects, you are asked to remain at your desk and to not disrupt the other participants. After the program terminates you will be able to use the computer in any way you like, for example to browse the internet.

For Subjects in Treatment:

In the second stage of the experiment we give you the opportunity to donate a part of your earnings in the first stage of the experiment to the American Red Cross. You can choose to donate none or all of your earnings, as well as any integer amount in between.

The computer will display your total compensation for the experiment: your earnings in the first stage + $5 show up fee your charitable donation to Red Cross. When everybody finishes the experiment you will be paid this amount by the experimenter. If you finish before other subjects, you are asked to remain at your desk and to not disrupt the other participants. After the program terminates you will be able to use the computer in any way you like, for example to browse the internet.

After the conclusion of the experiment, everybody's donations will be totaled and a check will be mailed to American Red Cross. One of you will be randomly selected to walk with the experimenter to the nearest mail box to witness the mailing of the check. For this task, the monitor will be paid an additional 5 dollars.

Appendix 2: Point Estimates and Confidence Intervals for The Effect of Treatment on Cheaters

A subject who never reports rolling a 5-5-5 is certainly honest. A subject who reports rolling three fives is either honest and lucky or dishonest. We can therefore identify a subset of honest subjects, but we cannot identify which of the successful subjects are dishonest. Instead, we perform some calculations and Monte Carlo simulations to estimate how giving in Treatment compares to giving in Control among dishonest subjects.

We decompose the difference in giving among successful subjects observed in the data $\Delta G_S = E[G|T = 1, S = 1] - E[G|T = 0, S = 1]$ as follows:

$$\begin{aligned}\Delta G_S = {} & P(c=1|T=1,S=1) \cdot \{E[G|T=1,c=1] - E[G|T=0,c=1]\} \\ & + P(c=0|T=1,S=1) \cdot \{E[G|T=1,c=0] - E[G|T=0,c=0]\} \\ & + \{P(c=1|T=1,S=1) - P(c=1|T=0,S=1)\} \cdot \{E[G|T=0,c=1] - E[G|T=0,c=0]\},\end{aligned}$$

where $P(c = 1|T = 1, S = 1)$ is the probability a subject is a cheater conditional on being in Treatment and reporting a success, and $P(c = 1|T = 0, S = 1)$ is the probability a subject is a cheater conditional on being successful in the control group.

We also observe the difference in giving between successful and unsuccessful subjects in Control. This can be written as: $E[G|T = 0, S = 1] - E[G|T = 0, S = 0] = P(c = 1|T = 0, S = 1) \cdot \{E[G|T = 0, c = 1] - E[G|T = 0, c = 0]\}$. Plugging the resulting expression for $\{E[G|T = 0, c = 1] - E[G|T = 0, c = 0]\}$ back in the decomposition for ΔG_S above gives:

$$\begin{aligned}\Delta G_S = {} & P(c=1|T=1,S=1) \cdot \{E[G|T=1,c=1] - E[G|T=0,c=1]\} \\ & + P(c=0|T=1,S=1) \cdot \{E[G|T=1,c=0] - E[G|T=0,c=0]\} \\ & + \frac{P(c=1|T=1,S=1) - P(c=1|T=0,S=1)}{P(c=1|T=0,S=1)} \cdot \{E[G|T=0,S=1] - E[G|T=0,S=0]\}.\end{aligned}$$

We then solve for $E[G|T = 1, c = 1] - E[G|T = 0, c = 1]$ using point estimates for all the other differences from Table 2 and a few additional ordinary least squares regressions. For example, for donations out of 10 dollars, we observe: $\Delta G_S = -1.99$, $E[G|T = 1, c = 0] - E[G|T = 0, c = 0] = -1.76$, and $E[G|T = 0, S = 1] - E[G|T = 0, S = 0] = -0.455$. Point estimate for conditional probabilities could be obtained from information in Table **??** in the following way. 32 out of 64 subjects did not report a success in Control and are therefore certainly honest. The proportion of honest subjects who would fail to report a success after 80 dice rolls is about 73.2%. Thus the expected number of honest subjects in Control is $\frac{32}{0.732} = 43.71$. Of these, $43.71 - 32 = 11.71$ were successful, so $P(c = 0|T = 0, S = 1) = \frac{11.71}{32}$. Similarly, 25 subjects fail to report a success in Treatment, so there must be around $\frac{25}{0.732} = 34.15$ honest subjects in Treatment. Thus, out of the 41 subjects who reported a success in Treatment, only $34.15 - 25 = 9.15$, are honest. Thus, $P(c = 0|T = 1, S = 1)$ is approximately $\frac{9.15}{41}$.

With these values, the equation above for donations out of 10 dollars becomes:

$$-1.99 = \frac{31.85}{41} \cdot \{E[G|T=1,c=1] - E[G|T=0,c=1]\} + \frac{9.15}{41} \cdot (-1.76) + \frac{\frac{31.85}{41} - \frac{20.29}{32}}{\frac{20.29}{32}} \cdot (-0.455),$$

which yields the point estimate for $E[G|T = 1, c = 1] - E[G|T = 0, c = 1]$ of -1.92. Similarly, the point estimate for $E[G|T = 1, c = 1] - E[G|T = 0, c = 1]$ out of 25 dollars is -2.95.

To obtain confidence intervals for these point estimates we run Monte Carlo simulations. We draw a value for each difference in giving we used above from a normal distribution centered at the point estimate and with the standard deviation estimated from regressions. For conditional probabilities we draw a value from the simulated posterior distribution of the number of cheaters in each group, conditional on observing 32 and 41 subjects reporting a success, respectively. We then solve for the estimate for $E[G|T = 1, c = 1] - E[G|T = 0, c = 1]$ according to the equation above. With 50,000 draws, $E[G|T = 1, c = 1] - E[G|T = 0, c = 1]$ was above 0 around 9% of the time for donations out of 10 dollars, and around 13% of time for donations out of 25 dollars.