Conscience Accounting: Emotion Dynamics and Social Behavior

Uri Gneezy*

Alex Imas†

Kristóf Madarász‡

October 19, 2013

Abstract

The paper reports the results of two experiments in which people who first made an immoral choice were then more likely to donate to charity than those who did not. In addition, those who knew that a donation opportunity would follow the potential moral transgression were more likely to behave immorally than those who were not told of the donation option. We interpret this increase in charitable behavior as being driven by a temporal increase in guilt induced by past immoral actions. We term such behavior conscience accounting, and discuss its importance in charitable giving.

JEL classification: D03, D06, D09

Keywords: Emotional Dynamics, Temporal Brackets, Deception, Prosocial Behavior

*Rady School of Management, University of California, San Diego, La Jolla, CA 92093

†Dept. of Economics, University of California, San Diego, La Jolla, CA 92093

‡London School of Economics, London WC2A 2AE, United Kingdom.

1 Introduction

This paper argues that dynamic fluctuations of guilt shape people's prosocial motivations. We present evidence that people who first made an unethical choice were then more likely to donate to charity than those who did not. We interpret these results in the context of our theory, which offers a new explanation for why people donate to charity: to offset a feeling of guilt associated with recent "bad" actions. We term this result "conscience accounting."

Throughout history, institutions have been built to take advantage of the effects of guilt on charitable behavior and to enable individuals to account for their conscience. For example, around the time of the Second Temple—500 B.C. to 70 A.D.—Jewish leaders formalized the use of chatot (sins) and ashamot (guilt) offerings as atonement for transgressions. The medieval Catholic Church adopted a similar technique when it began to grant "indulgences" that absolved an individual of sins through a system of "tariff penances." Today, Mass in the Catholic Church typically involves congregants reciting a prayer called the Confiteor (Mea Culpa) in which they confess, and are in turn reminded of, their sins. Afterwards, the Church solicits alms by requiring congregants to pass around a collection plate.

These kinds of institutions appear to take advantage of individuals' self-imposed moral constraints and the need to account for past transgressions through compensatory behavior. However, by providing people with the explicit opportunity to relieve their conscience, such practices may actually increase unethical behavior by lowering its cost. In this paper we present evidence for both effects, showing that individuals are more likely to be charitable within a temporal bracket after an unethical choice and that the knowledge of subsequent prosocial opportunities increases unethical behavior ex-ante.

Our paper begins with the observation that an individual experiences aversive feelings of guilt after she violates an internalized norm or acts in a way that she

views as unethical.[1] After the initial increase, such guilt depreciates over time as the individual's emotional state gradually reverts back to the original "cold" one (Elster, 1998). The identification of our proposed hypotheses relies on the fact that such emotional fluctuations lead to a time-inconsistency in the individual's ex-post prosocial behavior. A moral transgression creates a temporal bracket where the immediate onset of guilt creates a greater propensity for prosocial behavior which diminishes over time. We term this emotional response conscience accounting, and this systematic change in social preferences also helps us identify norm violations in choice behavior.

Being at least partially aware of their emotional responses, our mechanism implies that individuals may adjust their ex-ante behavior in accordance to available future prosocial opportunities. Since donating after a norm violation will make the guilty person feel better, if the charitable opportunity is small or limited, then knowing that this opportunity exists may *encourage* unethical behavior in the present. However, if the charity option is large or unlimited, individuals may initially choose to avoid or delay charitable opportunities after moral transgressions because they fear that their guilty self will be overly generous.

We test our behavioral hypotheses using two experimental paradigms, finding support for the prediction of conscience accounting: individuals who achieved a given payoff by deceiving or stealing were more likely to donate to charity than those who achieved the same payoffs in a more ethical manner. In addition, our results show that this effect occurs within a temporal bracket, such that the increase in prosocial behavior is greatest directly after the unethical act and decreases with time. We also find that individuals anticipate these effects, lying more when they know that a limited donation opportunity will follow.

Our results have a direct application to charitable contributions and volunteering behavior, suggesting an additional explanation for why people donate their money and time. The willingness to give has puzzled economists for decades;

[1]Guilt is associated with moral transgressions (e.g. Baumeister, Stillwell and Heatherton, 1994), and the desire to avoid violating the expectations of others has been established as equilibrium behavior within a game theoretic framework (Battigalli and Dufwenberg, 2007, 2009; see Charness and Dufwenberg, 2006 and Dufwenberg and Gneezy, 2000 for experimental evidence of such guilt aversion). Guilt is also considered an aversive feeling that discourages norm violations (e.g. Akerlof and Kranton, 2000), and once experienced, facilitates social behavior (Amodio, Devine and Harmon-Jones, 2007; de Hooge, Zeelenberg and Breugelmans, 2007).

not only because it contradicts the assumption that people are fueled solely by self-interest, but also because it does not seem to be driven by one simple alternative (Andreoni, 1990; Becker, 1976; Cappelen, Konow, Sørensen and Tungodden, forthcoming; Meier, 2007; Ariely, Bracha and Meier, 2009; Small and Loewenstein, 2003; Vesterlund, 2003). The findings presented here suggest that giving can potentially be driven by guilt induced by prior unethical behavior, which speaks to recent research showing that non-standard motivations may impel prosocial behavior (e.g. Dana, Weber and Kuang, 2007; Sachdeva, Iliev and Medin, 2009).

Besides charity, the results also highlight the potential importance of guilt-based pricing for businesses. For example, travelers flying out of some airports receive the opportunity to offset the carbon footprint of their flight. Using "Climate Passport kiosks," people can calculate how many pounds of carbon dioxide their trip will produce and the cost of offsetting this footprint using donations to programs aimed at greenhouse gas reduction. Several online travel retailers have begun to offer a similar option—giving customers the choice of offsetting their carbon footprint directly after ticket purchase. This kind of business is in line with our hypotheses: people clear their bad feelings by donating. According to the conscience accounting hypothesis, programs that ask for donations close to the time of unethical purchase should be more successful than alternatives that ask people to donate at a more remote time or before the unethical purchase is made.

In addition, the dynamic effects of guilt on prosocial behavior adds to the growing literature on the "demand" side of philanthropy (Della Vigna, List and Malmendier, 2012). Although the emotional response discussed here is temporary, it may be used strategically to increase prosocial acts by individuals and organizations wishing to maximize donations. Furthermore, reminders of past unethical actions might lead to similar emotional dynamics as outlined in this paper. People may want to avoid guilty feelings, but will still act more prosocially if reminded about the ethical dimensions of past or current actions. Similarly, reminders of past immoral choices—such as directing a person's attention to her broken promises or deceptions—may induce feelings of guilt that can facilitate greater donation revenue for charitable institutions and to increased loyalty within organizations.

Our proposed framework is also linked to the economic literature of incorporating procedural norms into economic behavior (e.g., Kahneman, Knetsch and Thaler, 1986, Akerlof and Kranton 2000, Cappelen, Hole, Sørensen and Tungodden, 2007 and Kircher, Ludwig and Sandroni, 2010). Particularly, in our framework, individuals would prefer to adhere to procedural norms when attaining a given consumption vector. Upon violating a norm, however, they exhibit a temporal altruistic preference reversal toward others. In this manner it is possible to identify norm violations in observable behavior.

The rest of the paper is organized as follows. In Section 2 we introduce and develop our behavioral hypotheses. In Section 3 we present evidence from a deception game experiment in which we provide support for our predictions. Section 4 lays out the results of an "over-paying" experiment that provides further support for our theory, and discusses the related moral licensing hypothesis (Monin and Miller, 2001) in more depth. In Section 5, we discuss several examples of how conscience accounting can be utilized by firms to maximize revenue, and posit how our framework can be generalized to other emotions.

2 Behavioral Hypotheses

In this Section, we outline three behavioral predictions based on a simple formal model of emotional decision making in the presence of moral constraints.[2] Consider an individual who faces a temporal sequence of allocation decisions between herself and others leading to a final consumption vector. In specifying the decision maker's preferences, we extend the basic model of altruism, e.g., Becker (1976). In particular, we assume that the decision maker internalizes a set of moral constraints or norms and experiences feelings of *guilt* upon violating these constraints. Similar to Akerlof and Kranton (2000), norms here describe what the individual should not do; rather than prohibiting specific payoffs, these constraints prohibit procedures and actions by which these payoffs are attained. Examples of such procedural norms include attaining the same payoff allocation by either lying or

[2]To better focus on the key mechanisms, we present a very simple setting. For a more broadly applicable model and demonstration that the effects described here extend to general allocation problems, see Gneezy, Imas and Madarász (2012).

telling the truth, by stealing from business partners or receiving a gift, with the former procedure constituting a moral transgression in each case. Conditional on attaining the same payoff, individuals in our framework have a clear preference for not violating a moral constraint because doing so induces negative feelings of guilt.

To formalize this idea, we assume that the individual makes a sequence of allocation decisions in periods $t = 1, 2, 3$ the sum of which produces a final allocation vector be consumed in period $T = 4$.[3] To incorporate emotions further, we take that in each decision period t, the individual derives anticipatory utility based on her expectation of the final consumption vector and her current feelings of guilt (Loewenstein, 1987). Specifically, the decision maker's anticipatory utility at time t depends on her emotional state experienced in that period, $g_t \in \mathbb{R}^+$, which is interpreted as the intensity of her guilt. Since consumption takes place in period $T = 4$, her anticipatory utility in a periods $t < T$ can be expressed as

$$u(x, y, g_t)$$

where x is her own final consumption in period T, y is the total consumption given to others in period T. In line with standard models of altruism, both x and y are taken to be normal goods and u to be twice differentiable.

In a context where the individual experiences multiple periods of anticipatory emotions, her preferences at time t are defined by maximizing the sum of current and future anticipatory utilities and final consumption utility under rational expectations. Thus the individual's preferences in period t are given by:

$$U_t = E_t \sum\nolimits_{s=t}^{s=4} u(x, y, g_s)$$

Note that if the level of guilt remains the same over time, then preferences over the final allocations are stable and do not change. In contrast, when the level of guilt changes from one period to another—which will always be the case after a moral transgression—her preferences over the final allocations may also change over time. Thus emotional fluctuations due to moral transgressions lead to time-

[3]The model directly extends to any finite number of periods T.

inconsistency in behavior. We will exploit such time-inconsistency in identifying the implications of the model.[4]

2.1 Psychological Framework

We now turn to the framework characterizing how guilt affects preferences, first describing a simple dynamic of how guilt evolves over time, then discussing the relationship between guilt and preferences.

As is typical of many emotions, a class of events triggers a rapid and relatively large subsequent change in an individual's emotional state. With the passage of time, the individual "cools off" and the emotional state reverts back to baseline (Loewenstein, 1996; Elster, 1998; Van Boven, White and Huber, 2009). Describing the dynamics of guilt, we assume that the violation of a moral constraint is followed by a *subsequent* increase in feelings of guilt, and that this guilt gradually subsides over time. Formally,

$$g_{t+1} = \gamma g_t + z \text{ where}$$
$$z = 1 \text{ if choice at } t \text{ violates a moral constraint, } z = 0 \text{ otherwise.}$$

where $\gamma \in (0, 1)$ expresses the speed at which guilt decays.

In our setting, guilt affects utility in two ways: it is not only an aversive feeling, but one that also facilitates prosocial interactions (Baumeister, Stillwell and Heatherton, 1994; de Hooge, 2012). Guilt is costly and one would want to avoid experiencing it conditional on being able to attain the same payoff allocation. Additionally, guilt is accompanied by the subsequent desire to treat others well and behave more prosocially (Kelter and Lerner, 2010). In turn, guilt can be characterized as a relative substitute of a person's own consumption: feelings of guilt increase the importance of improving the consumption of others relative to one's own. We assume that people prefer to avoid violating norms because of guilt, and conditional on violating a norm, behave more prosocially. Formally,

$$u_g < 0 \text{ and } u_{g,x} < 0 \leq u_{g,y}$$

[4]As standard, given time-inconsistency and rational expectations, the solution of the decision maker's maximization problem is obtained via backward induction.

2.2 Hypotheses

Below we derive three predictions of the model outlined above. Since emotions are inherently temporal, the identification of our proposed mechanism relies critically on the link between the timing of choice and the resulting behavior.

To test the main implications of the above model, it will suffice to consider two simple choice sets arranged over three periods. Let $M = \{(x_1, y_1), (x_2, y_2)\}$ be a choice set with two allocation options, where x refers to the decision-maker's own material payoff, y to the material payoff of others and assume that $x_1 > x_2$. Let $D = \{(-d, d), (0, 0)\}$ be a donation choice set between a positive transfer from the decision-maker to others (a donation) and an allocation $(0, 0)$ with no transfer (no donation). Finally, let the empty choice set $\emptyset$ simply describe a period where there is no allocation decision to be made. To render the problem non-trivial, unless mentioned otherwise, we assume that choosing the more selfish allocation from M requires a norm violation / moral transgression

We can now consider two distinct choice problems. Let $S_{\text{hot}} = \{M, D, \emptyset\}$ describe the choice problem where the donation option directly follows the potential norm violation, and let $S_{\text{cold}} = \{M, \emptyset, D\}$ describe the alternative choice problem where the donation option is presented after some time has passed. Given these choice problems we first describe a hypothesis that directly encapsulates conscience accounting.

In our framework, guilt dynamics imply that after a moral transgression people exhibit a greater willingness to act prosocially. Hence, in a temporal bracket following a potential norm violation we expect to see more prosocial behavior. As guilt decays with time, this increased tendency for prosocial behavior diminishes as well.

Claim 1 (Conscience Accounting Hypothesis) *The decision maker is more likely to donate after a norm violation if the donation option is presented earlier than if it is presented later, i.e. in S_{hot} versus in S_{cold}. If the donation option is known ex-ante, the decision maker is also less likely to jointly violate a norm and not donate in S_{hot} than in S_{cold}.*

Note that the conscience accounting hypothesis holds both if the donation

option is a surprise as well as when it is known ex-ante. Furthermore, if the donation is a surprise, the propensity to donate after an immoral choice is higher in S_{hot} than in S_{cold}, while it is the same after a moral choice. We test these predictions directly in the next Section.[5]

To account for the fact that violating a moral constraint may lead to a higher earning for the decision maker (e.g. $x_1 > x_2$), we will also test whether the same increased tendency for prosocial behavior is observed when the decision maker's payoffs are kept constant but there is no possibility of a norm violation. Our mechanism suggests that if the decision maker attains the same payoff *without* violating a moral constraint then we should not observe a temporal increase in prosocial behavior. This prediction is discussed further when addressing potential income effects in Sections 3 and 4.

In the absence of knowing what acts violate norms and what acts do not, the above result also allows us to identify moral constraints in dynamic choice behavior. Holding the material payoffs constant, our first hypothesis links a specific temporal choice pattern to the presence of a moral violation. If an act violates a constraint, the level of subsequent altruism follows a predictable pattern—greatest after the act and returning back to baseline after some time has passed.

While the first hypothesis centered on the retrospective effects of guilt on behavior, the next two hypotheses focus on the prospective effects of guilt, i.e., how people respond to emotional fluctuations ex ante. The dynamic pattern of emotions implies that as guilty feelings change, so do preferences over allocations. If a person knew that she would have an opportunity to donate after an unethical act, then being at least partially aware of the conscience accounting response can lead to a dynamic impulse control problem: the moral violation produces an altruistic urge that the individual would potentially want to control prior to the violation. If the individual believes that the subsequent guilt will induce her to be more charitable than she finds optimal ex ante, she would prefer to limit her ability to donate, e.g. by delaying the opportunity until she feels less guilty, limiting ex

[5]In the case where the donation option is known ex-ante, the timing of the donation option may effect the initial decision to violate a norm. We discuss this below.

post donation options, or not violating the norm at all.[6]

At the same time, our assumption that guilt is a relative substitute of own consumption implies that there is always a general complementary relationship between moral violations and prosocial actions—moral violations increase one's desire to be prosocial. In particular, if a desirable allocation can only be attained by violating a norm, then this allocation is more likely to be chosen if a limited donation option is available shortly after the unethical act. "Paying for her sins" by donating after unethical behavior will make a guilt-prone individual feel better: since guilt is a relative substitute for own-consumption over the consumption of others, moving towards a more altruistic allocation will help reduce the overall utility loss from immoral behavior.

To demonstrate the latter effect, take the case where an individual always prefers the selfish allocation when it does not require a moral transgression, but prefers the less selfish one when it does —finding the utility cost of guilt associated with deception to outweigh the payoff difference. Since guilt is a bad in our model, the required moral transgression leads the individual to choose the selfish allocation less often than when immoral actions are not required to attain it.[7] Here, adding the option to donate after the potential norm violation will allow the decision-maker to 'lower' the overall utility cost of transgressing, making her more likely to transgress and subsequently donate. In turn, informing the decision-maker of a limited prosocial opportunity will increase the likelihood of norm violation, and if she does violate a norm, the decision-maker will also choose to subsequently donate.

To state the second hypothesis, consider a choice problem where there is no donation option: $S = \{M, \emptyset, \emptyset\}$.

Claim 2 (Paying for Sins Hypothesis) *Suppose that the decision maker does not violate the norm in S. Informing the decision maker about a future donation*

[6]Della Vigna, List and Malmendier (2012) provide evidence that the ex-ante ability to avoid later donation solicitations reduces donations by 28%-44%. Andreoni, Rao and Trachtman (2011) provide further evidence for such avoidance techniques.

[7]Such preferences are consistent with Gneezy (2005), who demonstrates that individuals are significantly more averse to obtaining a selfish allocation through unethical behavior such as lying than choosing it from a set of allocations, e.g. in a Dictator game. For example, 90% of individuals chose $(15, 5)$ over $(5, 15)$ in a Dictator game while only 52% were willing to obtain the $(15, 5)$ allocation when it required a lie ($p < .01$).

option ex-ante, i.e., that she is facing S_{hot}, will increase the likelihood of a norm violation, and if she now transgresses, she will also donate in S_{hot}.

Given an ex-ante choice to be presented with a binary donation option either soon after a moral transgression or after some time has passed, one can demonstrate the interplay between the demand for altruism and the worry of paying too much for one's sins. Since guilt diminishes over time, a decision to delay the donation option can serve as a commitment to give less. Similarly, a decision to be presented with the donation option early can serve as a commitment to give more. People who morally transgress may exhibit both preferences. It is always true, however, that those who find it optimal to "pay for their sins" would strictly prefer the early opportunity—because they fear that their later and colder self will not be generous enough—and donate. On the other hand, those who prefer not to donate and are worried about the above discussed impulse control problem will strictly prefer the late donation opportunity, and then will not donate. This link between donation behavior and preference for timing leads to our third hypothesis:

Claim 3 (Choice Hypothesis) *When presented with the ex-ante choice of either facing an early or a late donation option, S_{hot} or S_{cold}, those who strictly prefer an early opportunity and violate the norm will donate. Those who strictly prefer a late opportunity and violate the norm will not donate.*

Above, we described how varying the temporal distance between choices or the information about this distance could be used to identify guilt dynamics. In addition, the proposed mechanism speaks to how behavior might change when varying the *order* of decisions. Particularly, consider the comparison between choice sets where the donation option precedes the potential moral violation, $S_{\text{pre}} = \{D, M, \emptyset\}$, and the case where it follows, S_{hot}.

In this context, one can state an *alternate hypothesis* to conscience accounting where people simply alternate between being prosocial or not, but are not subject to guilt dynamics. Under the alternate hypothesis, the ordering of choices should not matter with regards to the overall pattern of behavior. Specifically, the proportion of individuals who choose to donate and then lie and those who do not donate and tell the truth should be similar whether the donation option follows

the decision to lie or precedes it. If lying positively predicts donations in S_{hot}, then under this alternative hypothesis, donating should predict lying to a similar extent in S_{pre}.

In contrast, the presence of guilt dynamics under the conscience accounting hypothesis suggests a different prediction. Since guilt increases prosocial motivation, there should be more donations by those who violate a norm when the donation option follows the potentially immoral choice than when the ordering is reversed. This prediction follows from our model because guilt increases only *after* a moral transgression. Hence, there should be less individuals who jointly lie and do not donate in S_{hot}, which suggests a significant correlation between the immoral choice and donations in S_{hot} but not in S_{pre}. The predicted difference in correlations rests on the assumption that not donating is considered to be less of an unethical act than the moral transgression in M, e.g., lying or stealing. We present empirical support for this assumption in the next section.

3 A Deception Game

3.1 Procedure

To study conscience accounting empirically, we conducted a two-stage experiment. First, participants could lie to increase their profits at the expense of another participant. Second, after choosing whether to lie, we gave participants the option to donate to a charity.

We used a setup similar to Gneezy (2005). In this two-player deception game, one player, the Sender, has private information and the other, the Receiver, makes a choice based on a message conveyed by the Sender. The payoffs for both players depend on the choice the Receiver makes. We constructed payoffs such that lying (sending a "wrong" message) resulted in a higher payoff for the Sender.

In the instructions (see Online Appendix)[8], we told participants that the experiment had two possible payment outcomes. Although the Receiver's choice would determine the outcome, only the Sender knew about the monetary outcomes of each option—the Receiver had no information regarding the alignment

[8]https://sites.google.com/site/alexoimas/ca

of incentives. As was shown using this game (e.g. Gneezy, 2005; Sutter, 2009; Dreber and Johannesson, 2008), most Receivers choose to follow the message sent by the Sender, and most Senders expect this.

After choosing whether to lie, Senders were given the option to donate to a charitable foundation. Depending on the treatment, the donation option was either presented directly after the decision to lie or with some delay, and Senders were either aware of this option when choosing whether to lie or not.

We recruited 528 undergraduate students at the University of California, San Diego for a classroom experiment. The rules of the experiment were both read aloud and presented in written form to the participants.

We informed them that neither the Sender nor Receiver would ever know the identity of the player with whom they were matched. Participants in both roles knew that 1 of 10 students assigned the role of Sender would be randomly chosen to be paid, and be matched with Receivers in a different class.

Senders could choose from one of 10 possible messages to send the Receiver. Each message was in the form of "Choosing _ will earn you more money than any other number," with the blank corresponding to a number from 0 to 9. We told the Sender that if the Receiver chose a number that corresponded to the last digit of the Sender's Personal Identification number (PID), both players would be paid according to payment Option Y, and if the Receiver chose any other number, both players would be paid according to Option X. We informed Senders of the monetary consequences of both Option X and Y, and that the Receivers were not informed of this. We constructed the payments such that Option Y earned the Receiver more money than the Sender, and Option X earned the Sender more money than the Receiver. Hence, if the Sender expected the Receiver to follow her message, she had a monetary incentive to send one that did not correspond to the last digit of her PID—to lie—so the Receiver would choose the wrong number. Gneezy (2005) showed that Senders in the deception game expected the Receivers to follow their message (82%), and that Receivers indeed largely followed the message sent (78%). Receivers in our experiment largely followed the Senders' messages as well, with 75% of participants assigned the role of Receiver choosing the number the Sender wrote would earn them the most money.

All treatments (other than baseline) offered Senders the option to donate either $1 or $2 to the Make-A-Wish foundation. These numbers were used so that the donation amount would always be relatively small in comparison to the amount that could be gained through deception. In the Incentive and No Incentive treatments, we presented the donation option directly after Senders made their message choices. In the Incentive Delay treatment, we presented the donation option with some delay: after their message choice, Senders received anagrams to solve for 10 minutes before we presented them with the option to donate. Importantly, in these three treatments Senders were not aware of the subsequent donation option when choosing what message to send, but were informed of it only after they made their initial choice. The Incentive Reverse treatment was similar to the Incentive treatment, except that the donation option was presented before the message choice. Senders were not aware of the deception game when making the donation decision.

In the Informed Incentive and Incentive Choice treatments, however, Senders knew in advance they would have the opportunity to donate. Other than Senders being informed that this opportunity would directly follow their message choice (and the different payoffs), the Informed Incentive treatment was similar to the Incentive treatment. In the Incentive Choice treatment, we asked Senders to choose whether they wanted to make the decision to donate sooner (directly after their message choice) or later (at the end of the experiment), at the same time as deciding what message to send. Senders made the actual donation decision according to this choice. Ten minutes of anagrams once again served as the delay.

The last treatment was a baseline containing the same payoffs as the Informed Incentive treatment but excluding the donation option.

Table 1 presents the payoffs we used in the experiment. We designed the Incentive, Incentive Delay, Incentive Reverse and Incentive Choice treatments such that if the Receiver chose the wrong number, the Sender stood to earn $10 more and the Receiver $10 less than if the Receiver chose the correct number. Senders had a smaller incentive to lie in the Informed Incentive and Baseline treatments, earning $5 more and the Receiver $5 less if the Receiver chose the wrong number.

In the No Incentive treatment the Sender had no monetary incentive to lie: both the Sender and Receiver stood to potentially earn $10 less if the Receiver

chose the wrong number. Note that a Sender in the No Incentive treatment could obtain the same payoff by telling the truth as she could obtain by lying in the Incentive treatment.

Since Receivers were not informed about the Senders' payoffs, we did not expect nor did we observe a difference in the Receivers' behavior.

We established subject identification through the PID numbers students provided as part of the experiment. We used the PID numbers to pay the participants according to the outcome of the experiment and to determine whether the Sender had lied in her message. Donations were either \$1 or \$2 in each available case, and we deducted the amount from the Senders' payments if they chose to donate. We then made the donations on behalf of the Senders directly through the Make-A-Wish website. Using these treatments we test our hypotheses.

Conscience Accounting Hypothesis: The Incentive and Incentive Delay treatments allow us to test the Conscience Accounting hypothesis. Conscience accounting predicts that those who lied in the Incentive treatment should donate to a greater extent than those who lied in the Incentive Delay treatment. Furthermore, we should also observe a greater correlation between the decision to lie and the decision to donate in the Incentive treatment than in the Incentive Delay treatment.

In addition, comparing the Incentive and Incentive Reverse treatments, conscience accounting implies that there should be a greater correlation between lying and donating when the decisions follow in that order (Incentive treatment), than when the donation option precedes the decision to lie (Incentive Reverse treatment). Particularly, the decision to lie or not should predict the decision to donate in the former treatment but not the latter.

We use the No Incentive treatment to test whether the moral character of a choice has an impact on prosocial behavior after controlling for earnings. The prediction on donation rates in the Incentive and No Incentive treatments follows from our model, and allows us to rule out that conscience accounting is driven by differences in the Senders' income. Those who lie in the Incentive treatment stand to earn the same amount as those who tell the truth in the No Incentive treatment (\$20). Our framework predicts that those who tell the truth to earn \$20 in the No Incentive treatment should feel less guilty, and in turn, be less likely to donate

than those who lie to earn $20 in the Incentive treatment.

Paying for Sins Hypothesis: The Informed Incentive and Baseline treatments were designed to test the Paying for Sins hypothesis. Given the small size of the donation, the hypothesis would predict that those in the Informed Incentive treatment should be more likely to lie than those in the Baseline treatment, where Senders were not informed of a subsequent donation opportunity. This prediction holds under the assumption that more individuals would prefer the selfish allocation in the Baseline treatment if they did not have to lie to attain it (for evidence that this assumption holds, see Gneezy, 2005).

Choice Hypothesis: The Incentive Choice treatment allowed us to test the Choice hypothesis which predicts that those who lied and chose the early donation opportunity should donate more than those who lied and chose the late opportunity.

Is Lying a Moral Violation? To provide external evidence in support of the moral category assumptions in the experiment, we adopted the incentive-compatible elicitation method of Krupka and Weber (2013) to assess people's moral attitudes towards the various choice options in our paradigm. We presented a separate group of participants with a description of the deception game and the possible choices available to the Sender, and used the method to elicit whether they viewed each choice as morally appropriate or inappropriate. Participants were instructed to view a morally inappropriate choice as one that, if chosen, would make the acting individual feel guilty. To incentivize truthful reporting, participants were paid an additional fee if their response to the moral appropriateness of a choice matched the response selected by most other people in the experiment. In turn, each was prompted to coordinate on the belief held by the social group as to whether a choice violates a moral constraint or not.

Participants (N=43) were recruited using the Amazon Mechanical Turk platform. Each read the Sender's instructions for the deception game and was asked to judge the extent to which each of the Sender's possible message choices was "morally appropriate" and "consistent with moral or proper social behavior" or "morally inappropriate" and "inconsistent with moral or proper behavior." Participants were told that at the end of the experiment, we would randomly select one of their choices, and if their moral appropriateness rating matched the modal

response of others in the experiment, they would be paid an additional $3 on top of a base fee of $0.50. If their rating did not match the modal response, no extra payment would be awarded.[9]

To test the assumption that donating nothing is not viewed as a moral violation, we separately elicited the moral appropriateness ratings for the choice to donate $2 from experimental earnings to charity as well as the choice *not* to donate.

[9]See Online Appendix for instructions

Table 1: Results by Treatment

Treatment	**Option**	**Sender($)**	**Receiver($)**	***N***	**Lying**	**Delay**	**Informed**
Incentive	X	20	10	57	65%	No	No
	Y	10	20				
Incentive Delay	X	20	10	36	75%	Yes	No
	Y	10	20				
Incentive Choice	X	20	10	38	61%	Yes	Yes
	Y	10	20				
Informed Incentive	X	20	15	62	63%	No	Yes
	Y	15	20				
Baseline	X	20	15	63	48%	-	-
	Y	15	20				
No Incentive	X	10	10	54	24%	No	No
	Y	20	20				
Incentive Reverse	X	20	10	48	60%	No	No
	Y	10	20				

3.2 Results

3.2.1 Moral Attitudes

Following Krupka and Weber (2013), we converted participants' responses into numerical scores.[10] See Online Appendix for summary ratings of all evaluated choices as well as robustness checks of the analysis.

Looking at message choices that did not match the Sender's personal code, the average rating given was -.40. Pairwise t-test comparisons revealed no significant differences between messages that we classify as lies—each was judged to be a moral violation if chosen by the Sender (all p-values $> .5$). On the other hand, the message matching the Sender's personal code had an average rating of .71. In pairwise comparisons the message that we classify as telling the truth was rated significantly more morally appropriate than any of the messages classified as lies (all p-values $< .001$).

Looking at donation decisions, the decision to donate had an average rating of .66: significantly more morally appropriate than any of the decisions to lie (all p-values $< .001$). Importantly, the decision *not* to donate, with an average rating of .09, was also judged to be significantly less of a moral violation than any of the choices to lie (all p-values $< .01$). In turn, we find support for our assumption that the decision not to donate does not constitute a moral violation while the decision to lie does. In light of our theory, this classification is consistent with the behavior revealed in the experiments described below.

3.2.2 Behavioral Hypotheses

Lying rates by treatment are presented in Table 1. The differences in lying rates between the Incentive and Incentive Delay treatments ($Z = 1.02,\ p = .15$), Incentive and Incentive Reverse ($Z = .47,\ p = .32$) the Incentive and Incentive Choice treatments ($Z = .43,\ p = .33$) were not statistically significant.[11] However, this difference between the Incentive and No Incentive treatments ($Z = 4.32,\ p < .001$)

[10] Participants chose from amongst five categories and numerical scores were assigned such that "very morally inappropriate"=-1, "somewhat morally inappropriate"=-1/3, "somewhat morally appropriate"=1/3, "morally appropriate"=2/3 and "very morally appropriate"=1.

[11] p-values were calculated from a one-tailed test of the equality of proportions using a normal approximation to the binomial distribution.

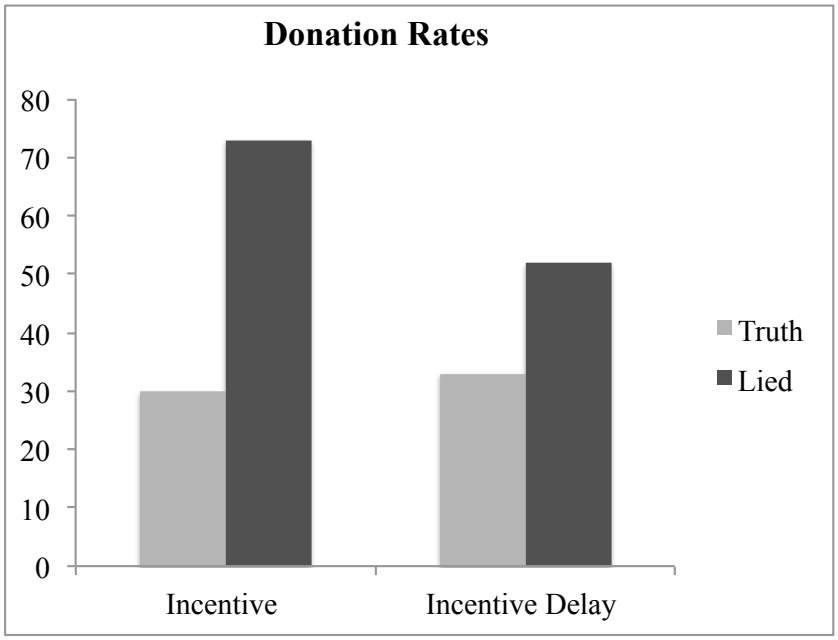
was statistically significant.

To test our first hypothesis, we rely on identification through time by examining behavior in the Incentive and Incentive Delay treatments. Our first key finding is that in the Incentive treatment, when the donation option came as a surprise directly after the message choice, 30% (6) of the participants who told the truth chose to donate, compared to 73% (27) of those who lied ($Z = 3.14,\ p < .001$): the participants who chose to lie—and potentially earn \$10 from lying—were significantly more likely to donate to charity than those who chose to tell the truth. Note that this finding demonstrates that classifying individuals into simple "types," where some always behave in a "moral" way (not lying and donating) and others never do (lying and not donating), is problematic. In our experiment, those who had previously lied were *also* more likely to donate to charity.

In the Incentive Delay treatment, where the option to donate was presented some time after the message choice, 33% (3) of the participants who sent a true message chose to donate compared to 52% (14) of those who lied ($Z = .96,\ p = .17$). Looking at the conditional propensities to donate in the Incentive and Incentive Delay treatments, those who lied and had the opportunity to donate directly after their message choice did so significantly more often than those who lied and faced a delay between the two choices ($Z = 1.74,\ p = .04$). On the other hand, the delay had *no effect* on the donation rates of truth tellers ($Z = .18,\ p = .43$). These results are summarized in Figure 1.

Since the only difference between the Incentive and Incentive Delay treatments was the time between the message choice and donation option, the observed difference in donation rates of liars also rules out a number of potential alternate explanations. Particularly, if the higher donation rate of liars compared to truth tellers in the Incentive treatment was due to Senders having heterogeneous preferences over allocations between Receivers, themselves, and the charity, then Senders who lied should have donated at the same rate in both treatments. An explanation based on some form of self-image preservation—of trying to cancel a bad deed with a good one—would also predict that donation rates should not change between treatments. In addition, since those who lied in both Incentive and Incentive Delay treatments were expecting to earn the same higher payoff of \$20,

Figure 1: Fraction of Senders Who Donated by Message Type



the greater donation rates of liars in the Incentive treatment cannot be explained by differences in earnings.

Looking at the relationship between lying and donating, an OLS regression revealed that the decision to lie had a significant influence on the decision to donate in the Incentive treatment ($\beta = .401,\ p = .001$). However, in line with our prediction, there was no relationship between the decision to lie and donating in the Incentive Reverse treatment where the order of the decisions was reversed ($\beta = .02,\ p = .87$). Regressing the choice to donate on the decision to lie, a treatment dummy, and the interaction of the decision to lie and the treatment dummy revealed a significant interaction effect ($\beta = -.404,\ p = .041$). The positive relationship between lying and donating was significantly greater when the donation option followed the decision to lie than if the choices were reversed. Individuals chose to lie and donate less in the Incentive Reverse treatment than in the Incentive treatment.

The difference in the relationship between lying and donating in the Incentive and Incentive Reverse treatments provides additional support for our conscience accounting hypothesis. If the increased donation rate of liars in the Incentive treatment was due to individuals broadly bracketing moral choices, offsetting a lie with a donation or vise versa, then the ordering of choices should not effect

observed behavior. However, guilt dynamics imply that the positive relationship between lying and donating should be larger when the donation option *follows* the decision to lie and not when the order of choices is reversed. Our results are consistent with the latter prediction.

To provide further support that differences in donation rates between liars and truth tellers in the Incentive treatment are not due to differences in the Senders' material payoffs – an income effect – we compare the donation rates of Senders whose material payoffs are held constant but who differ in whether these payoffs were attained by lying or telling the truth. If the higher donation rates of liars were caused by greater material payoffs in the deception game, then the same higher donation rates should be observed when Senders were truthful in the No Incentive treatment. In the No Incentive treatment, Senders did not have a monetary incentive to lie. Specifically, their expected payoffs from truth telling in the No Incentive treatment was the same as from lying in the Incentive treatment: \$20. Hence, the only difference between earning \$20 in the Incentive and No Incentive treatments is that a moral violation is required in the former but not the latter.

Consistent with our hypothesis, we found that in the No Incentive treatment, of those who told the truth, 51% (21) chose to donate compared to 73% (27) of those who lied in the Incentive treatment. Those who lied in the Incentive treatment were still significantly more likely to donate than those who had told the truth in the No Incentive treatment ($Z = 1.97,\ p = .02$), despite the fact that the expected own payoffs were the same.

Combined, these results provide direct support for the Conscience Accounting hypothesis which predicts that when the subsequent donation option comes as a surprise, individuals who violated a norm will be more likely to donate than those who did not within a temporal bracket.

We now turn to our predictions for contexts where Senders are informed of the donation option in advance. To test our Paying for Sins hypothesis, we compare the lying rates of Senders in the Informed Incentive treatment to those in Baseline. In line with our predictions, 63% (39) of Senders lied in the Informed Incentive treatment compared to 48% (30) of those in the Baseline treatment($Z = 1.72,\ p =$

.04).[12] Of those who lied, 82% donated. Given the relatively low rate of deception in the Baseline treatment, these results provide support for the Paying for Sins hypothesis: knowing that they would be presented with the opportunity to "pay for their sins" in the future led to Senders more willing to lie.

Although a direct comparison to prior research is not appropriate, we note that Gneezy (2005) observed 66% of individuals chose the selfish allocation in a Dictator game for a gain of $1, while we observed that only 48% of Senders were willing to lie to obtain a larger gain of $5 in the Baseline treatment. This suggests that Senders would be more willing to choose ($20, $15) over ($15, $20) if it did not require a lie.

In the Incentive Choice treatment, where Senders could choose when to be presented with the donation option, of those who lied 43% (10) chose to make their donation decisions early and 57% (13) chose to make their donation decisions late. Of those who chose to make their donation decision early and lied, 90% (9) actually donated, compared to 31% (4) of those who chose to decide later and lied ($Z = 2.84,\ p < .001$). These results provide support for the Choice Hypothesis: when given the choice of when to be presented with the donation option, those who chose to be presented with it early donated much more than those who chose to be presented with it late.[13]

We did not find any significant gender differences in behavior.

4 An Over-paying Experiment

4.1 Procedure

In the deception game experiment, participants knew we were observing whether they had lied. We designed the second experiment such that participants were

[12] To test the robustness of these results, we ran these treatments again with a different group of subjects using the same instructions ($Z = 1.74,\ p = .04$). Combining results across both iterations yielded similarly significant results ($Z = 2.02,\ p = .02$).

[13] It should be noted that there was no increase in deception in the Incentive Choice treatment when compared to the Incentive and Incentive Delay treatments where individuals were not informed of the subsequent donation option ($Z = .91,\ p = .18$). We believe that this was due to the fact that in addition to being informed of the subsequent donation option, individuals in the Incentive Choice treatment were also asked to make a choice on the timing of this option, which may have interacted with their decision to lie.

unaware we were studying their moral choices. This unawareness should reduce behavior based on the experimenter demand effect and/or experimenter scrutiny.

We paid groups of subjects for their participation in an unrelated experiment. Two groups received payment according to how much we promised them. A third group received more than they were promised by "mistake" and had the opportunity to either return or keep the extra money.[14] We also gave all three groups the option to donate $1, and we recorded donation rates across the groups. According to our hypothesis, we expected conscience accounting to manifest itself in the third group, predicting participants who decided to keep the extra money for themselves would be more likely to donate than those who had returned it.

We recruited 160 undergraduate students at the University of California, San Diego to participate in a coordination game experiment with an advertised expected payoff of $10 (see Blume and Gneezy, 2010). We invited subjects to the lab in pairs and seated them far apart for the duration of the game, which took approximately 15 minutes. We guaranteed all participants a $5 show-up fee, and those who did not succeed in coordinating did not get any extra money.

In addition, participants received $10 or $14, depending on the treatment, if they were able to coordinate with the individuals whom they were matched with. We randomly assigned those who had succeeded in coordinating to one of three treatments. In the Low treatment, subjects learned they would receive an additional $10 if they had succeeded in coordinating with their partners. In the High treatment, participants learned the additional payment would be $14. In the Mistake treatment, we told participants they would get $10 if they had succeeded, but gave them $10 and an extra $4 by "mistake"–nine $1 bills and one $5 bill interspersed amongst them. Table 2 summarizes payments for all three treatments. After receiving their pay at the end of the experiment, participants in all three treatments were given a description of a child with cancer and asked if they wanted to donate $1 from their final payment to the child.

When they received their pay participants were told, "Here is your _. Please count it and sign this form," with the blank corresponding to the promised pay-

[14]The study of individuals who do not know they are participating in an experiment is a common practice in field experiments, see gift exchange experiments e.g. Gneezy and List (2006), and is used in part to minimize experimenter demand effects that may be present in the lab.

Table 2: Payoffs Used by Treatment

Treatment	**Payment Promised**	**Money Given by Mistake**	**Donation($)**	***N***
Low	10	-	1	40
High	14	-	1	40
Mistake	10	4	1	80

ment ($10 in the Low and Mistake treatments, $14 dollars in the High treatment). Then the experimenter left the room. All payments were made in $1 bills except for the extra $5 bill in the Mistake treatment. Participants in all three treatments then decided whether to donate.

This framework allowed us to test the Conscience Accounting hypothesis in a different context. Particularly, we expected participants in the Mistake treatment who did not return the extra money to be more likely to donate than those who had returned it. In addition, since attaining $14 in the Mistake treatment required a moral transgression while earning $14 in the High treatment did not, we predicted that donation rates in the former treatment would be significantly greater than in the latter.

To support that our moral category assumptions were met, we again used the method of Krupka and Weber (2013) to elicit moral judgements about the choices participants faced in this paradigm. Particularly, we gave the same group of participants as in the first study an exact description of the over-paying experiment. We presented them with the choices available to an individual in the Mistake treatment, and asked each to judge the extent to which keeping the extra money and returning the extra money was morally appropriate.

4.2 Results

4.2.1 Moral Attitudes

As in the first study, we converted participants' responses into numerical scores. Looking at the choice of not returning the $4 given by mistake, the average rating given was -.67. On the other hand, returning the $4 given by mistake was given an average rating of .86 ($t(84) = 23.82,\ p < .001$). In line with our assumptions,

keeping the extra money was viewed as a moral violation while returning the extra money was viewed as morally appropriate.

4.2.2 Behavioral Hypothesis

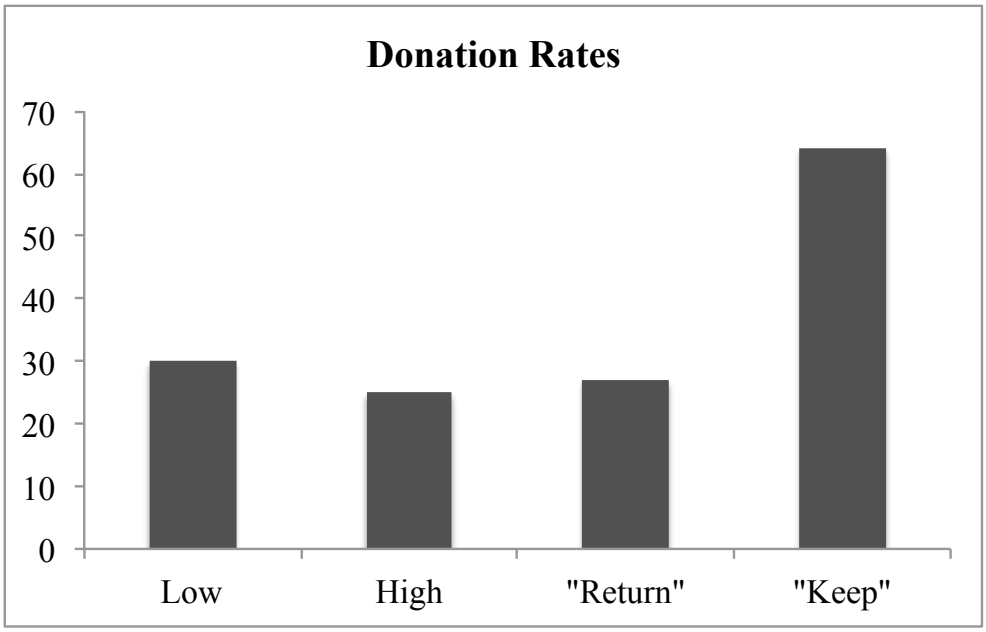
In the Mistake treatment, 41% (33) of participants returned the extra money they had received by "mistake." Donation rates by treatment are presented in Figure 2. Overall donation rates of participants were 30% (12) in the Low treatment, 25% (10) in the High treatment, and 49% (39) in the Mistake treatment. The overall donation rate in the Mistake treatment was significantly higher than in both the Low ($Z = 1.96,\ p = .03$) and the High ($Z = 2.50,\ p = .01$) treatments. Consistent with the Conscience Accounting hypothesis, of those who returned the extra money in the Mistake treatment, 27% (9) made a donation, whereas 64% (30) of those who did not return the extra money made a donation ($Z = 3.22,\ p < .001$).

An income effect of earning \$14 rather than \$10 does not explain the discrepancy in donation rates. Subjects in the High treatment, who were told they would earn \$14,[15] donated at about the same rate as those who returned the extra money but significantly less than those who kept it. Namely, although the donation rate for participants who returned the extra money is similar to those in the Low ($Z = .17,\ p = .43$) and High ($Z = .22,\ p = .41$) treatments, the donation rate for those who kept the money is significantly higher ($Z = 3.15,\ p < .001$ and $Z = 3.62,\ p < .001$, respectively). The difference in behavior in the Mistake treatment also suggests many participants, including those who did not return the money, did notice the mistake.

Our results also speak to the "moral licensing" hypothesis proposed by Monin and Miller (2001), where past moral actions can justify less moral choices down the road (see also Khan and Dhar, 2006; Mazar, Amir and Ariely, 2008; Zhong, Ku, Lount and Murnighan, 2009). For example, Monin and Miller (2001) showed

[15]Note that all participants were recruited to participate in the study for \$10. Hence in this case, as well as in the case of the mistake, the extra \$4 could be treated as a windfall because participants were initially expecting only \$10. Particularly, since the reference point coming into the study was the same in all treatments – participants had the same payoff expectations of earning \$10 – we believe that a windfall argument is not sufficient to explain the differences in donation rates.

Figure 2: Fraction of Participants Who Donated by Treatment



that participants allowed to establish themselves as not being prejudiced were more likely to later make remarks deemed socially offensive. One way to interpret moral licensing in the context of the over-paying experiment is to say that people who behaved morally and returned the extra money should be less likely to subsequently choose to donate than those achieving the same payoff without a moral act because they had earned the "license" not to. The results presented in Figure 2 reject this prediction: people who returned the extra money, and hence did not violate a norm, donated at the same rate as those who had no option to make such a moral choice.

In addition, looking at the deception game, we can compare the lying rates of Senders in the Incentive Reverse and Incentive treatments. Participants in the Incentive Reverse treatment were given an opportunity to donate before making their decision lie or tell the truth, while those in the Incentive treatment were not. Since not donating was not viewed as an immoral act, moral licensing would predict that giving individuals the opportunity to make a donation should increase the amount of subsequent deception (conditional on at least some individuals donating). We did not observe this effect: 65% of Senders lied in the Incentive treatment while 60% of Senders lied in the Incentive Reverse. Although the difference was not significant, directionally individuals appeared to lie less when they

had a prior opportunity to donate.

It should be noted that an important feature of studies demonstrating licensing is that the initial prosocial act was costless to the subject. For example, the subjects in the Monin and Miller (2001) study had the opportunity to establish themselves as unprejudiced at no cost to themselves. In another example, Khan and Dhar (2006) demonstrated licensing by having a group of individuals engage in one of two hypothetical volunteer assignments; they were then more likely than controls to choose a luxury item over a necessary item. However, a recent study by Gneezy, Imas, Brown, Nelson and Norton (2012) found that cost is a critical factor in licensing, showing that when the initial prosocial act came at a cost to the subject (as in our experiment), the licensing effect disappeared.

5 Discussion and Conclusion

We examine emotional dynamics in the context of social behavior. We posit and test several behavioral hypotheses where individuals care about the procedural aspects of their choices and, upon violating a norm, exhibit a specific time-inconsistency in their attitude towards others, preferring a more prosocial allocation after a norm violation than in its absence. This suggests an additional explanation for prosocial behavior in the presence of moral constraints: people donate to account for their conscience after making a morally bad choice. The fact that people who lie are more likely to donate to charity than people who tell the (costly) truth may seem counter intuitive. One goal of this paper is to reshape this intuition.

Throughout the paper we have focused on the specific emotion of guilt. However, other negative emotions such as anger may fit a very similar retrospective temporal pattern in the context of social behavior. While guilt changes preferences to be more prosocial, events that provoke anger affect preferences so that hurting the other party becomes more desirable within a temporal bracket (Card and Dahl, 2011). Angry individuals may lash out at others even at a cost to themselves if such an opportunity arises soon after a trigger, but may prefer to control this impulse ex ante. In this manner, anger functions as a temporal shock to choices

directed against the payoff of others. Such effects of anger on decision making are greater immediately after the incitement than after some delay—consistent with the folk wisdom of anger management: "count to 10 before reacting."

Incorporating emotional dynamics into models of charitable giving and prosocial behavior would provide further insight for theory that aims to better understand both the incidence of altruism and norm violations. For example, in moral choices such as the decision to lie, research has shown that people have a cost associated with breaking internal moral constraints that manifests itself as a conditional aversion to lying (Dreber and Johannesson, 2008; Gneezy, 2005; Sutter, 2009). Our findings suggest that feelings of guilt may play a role in those costs.

Our paper also sheds light on how businesses may optimally bundle products when pricing goods whose purchase violates internalized moral constraints. Climate Passport kiosks—as mentioned earlier—or check-out donations in liquor stores might speak to the existence of these practices. By offering such bundles in close temporal proximity (in a bracketing sense), businesses may not only increase the propensity to spend on the prosocial activity, but simultaneously increase the likelihood that individuals choose the products which may violate their moral constraints.

Additionally, the general relationship between emotional brackets and decision making outlined in our predictions provides an important avenue for future research, both on how emotions affect identifiable economic choices and the ways in which these effects are used strategically by individuals and organizations.

References

[1] Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, (2000), *115*(3), 715-753.

[2] Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A Dynamic Model of Guilt: Implications For Motivation and Self-Regulation in the Context of Prejudice. *Psychological Science*, *18*(6), 524-530.

[3] Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *Economic Journal*, *100*, 464-477.

[4] Andreoni, J., Rao, J. M., & Trachtman, H. (2011). Avoiding The Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. *mimeo.*

[5] Ariely, D., Bracha, A., & Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, *99*(1), 544-555.

[6] Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games. *American Economic Review. P&P*, *97*(2), 170-176.

[7] Battigalli, P., & Dufwenberg, M. (2009). Dynamic Psychological Games. *Journal of Economic Theory*, *144*(1), 1-35.

[8] Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin*, *115*(2), 243-267.

[9] Becker, G. S. (1976). A Theory of Social Interactions. *Journal of Political Economy*, *82*(6), 1063-1093.

[10] Blume, A., & Gneezy, U. (2010). Cognitive Forward Induction and Coordination without Common Knowledge: An Experimental Study. *Games and Economic Behavior*, *68*, 488-511.

[11] Cappelen, A. W., Konow, J., Sørensen, E. Ø., & Tungodden, B. (*forthcoming*). Just Luck: An Experimental Study of Risk Taking and Fairness. *American Economic Review.*

[12] Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The Pluralism of Fairness Ideals: An Experimental Approach. *American Economic Review*, *97*(3), 818-827.

[13] Card, D., & Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics*, *126*(1), 103-143.

[14] Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, *74*(6), 1579-1601.

[15] Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory*, *33*(1), 67-80.

[16] Della Vigna, S., List, J. A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*(1), 1-56.

[17] Dreber, A., & Johannesson, M. (2008). Gender Differences in Deception. *Economics Letters*, *99*(1), 197-199.

[18] Dufwenberg, M., & Gneezy, U. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior*, *30*(2), 163-182.

[19] Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature*, *36*(1), 47-74.

[20] Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to Be Nice: Consistency and Costly Prosocial Behavior. *Management Science*, *58*(1), 179-187.

[21] Gneezy, U., Imas, A., and Madarasz, K. (2012). Conscience Accounting: Emotional Dynamics and Social Behavior. *TE Series 563, London School of Economics*.

[22] Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review*, *95*(1), 384-394.

[23] Gneezy, U., & List, J. A. (2006). Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. *Econometrica*, *74*(5), 1365-1384.

[24] de Hooge, I. E. (2012). The Exemplary Social Emotion Guilt: Not So Relationship-Oriented When Another Person Repairs for You. *Cognition & Emotion*, *7*(26), 1189-1207.

[25] de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral Sentiments and Cooperation: Differential Influences of Shame and Guilt. *Cognition & Emotion*, *21*(5), 1025-1042.

[26] Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *The Journal of Business*, *59*, S285-300.

[27] Khan, U., & Dhar, R. (2006). Licensing Effect in Consumer Choice. *Journal of Marketing Research*, *43*(2), 259-266.

[28] Kircher, P., Ludwig, S., & Sandroni, A. (2010). Procedural Fairness: A Note on the Difference Between Social and Private Goods. *mimeo*, London School of Economics.

[29] Krupka, E. & Weber, R. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator-Game Sharing Vary? *Journal of the European Economic Association*, forthcoming.

[30] Keltner, D. & Lerner, J. (2010). Emotion. In D. T. Gilbert, S. Fiske, & G. Lindsay (Eds.), *The Handbook of Social Psychology* (pp. 312-347). New York, NY: McGraw Hill.

[31] Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *Economic Journal*, *97*(387), 666-684.

[32] Loewenstein, G. (1996). Out of Control?: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272-292.

[33] 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.

[34] Meier, S. (2007). Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment. *Journal of the European Economic Association*, *5*(6), 1203-1222.

[35] Monin, B., & Miller, D. T. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology*, *81*, 33-43.

[36] Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning Saints and Saintly Sinners: The Paradox of Moral Self-Regulation. *Psychological Science*, *20*(4), 523-528.

[37] Small, D. A., & Loewenstein, G. (2003). Helping a Victim or Helping the Victim?: Altruism and Identifiability. *The Journal of Risk and Uncertainty*, *26*(1), 5-16.

[38] Sutter, M. (2009). Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams. *The Economic Journal*, *119*(534), 47-60.

[39] Van Boven, L., White, K., & Huber, M. (2009). Immediacy Bias in Emotion Perception: Current Emotions Seem More Intense Than Previous Emotions. *Journal of Experimental Psychology: General*, *138*(3), 368-82.

[40] Vesterlund, L. (2003). The Informational Value of Sequential Fundraising. *Journal of Public Economics*, *87*(34), 627-657.

[41] Zhong, C. B., Ku, G., Lount, R. B., & Murnighan, J. K. (2009). Compensatory Ethics. *Journal of Business Ethics*, *92*(3), 323-339.

6 Appendix

In this Section, we present the three hypothesis formally. To see that the results hold more generally for convex D and M sets—as well as for further results—please consult Gneezy, Imas, Madarász (2012).

Proposition 1 (Hypothesis 1) *Suppose the donation option is a surprise. Given a norm violation, the DM is more likely to donate in S_{hot} than in S_{cold}. Suppose the donation option is known ex-ante. The DM is less likely to jointly act immoral and also not donate in S_{hot} than in S_{cold}.*

Proof. Consider first the problem where the donation option is a surprise. Since the availability of a donation option is not known at $t = 1$, the perceived problems in $t = 1$ are identical. Hence the initial choices in S_{hot} and in S_{cold} are the same.

Suppose that there is no norm violation at $t = 1$, then there is no increase in guilt—$g_2 = g_1 = 0$—and hence continuation behaviors in the two choice problems S_{hot} and S_{cold} are also identical.

Suppose now that there is a norm violation at $t = 1$. This implies that $g_2 > 0$. To prove the above claim, we need to show that the DM will be more altruistic at $t = 3$ than at $t = 2$. To show this, we need to compare the marginal rate of substitution (MRS) between own consumption and the consumption of others at $t = 2$ versus at $t = 3$:

$$\begin{aligned} MRS_{t=2}(x,y) &= \frac{u_x(x,y,1)+u_x(x,y,\gamma)}{u_y(\pi,1)+u_y(\pi,\gamma)} \leq \frac{u_x(x,y,\gamma)+u_x(x,y,\gamma)}{u_y(x,y,1)+u_y(x,y,\gamma)} \\ &\leq \frac{u_x(x,y,\gamma)+u_x(x,y,\gamma)}{u_y(x,y,\gamma)+u_y(x,y,\gamma)} = MRS_{t=3}(x,y) \end{aligned}$$

where the inequality follows from the assumptions that $u_{x,g} < 0 \leq u_{y,g}$ and that $\gamma \leq 1$. Thus whenever the decision-maker prefers to donate at $t = 3$—in the cold problem S_{cold}—she will do so at $t = 2$—in the hot problem S_{hot}—as well.

Consider now the case when the DM is informed in-advance about the donation option. This can have two effects on the $t = 1$ choice, (i) encourage norm violations by the ability of donating later or (ii) deter norm violations by fear of donating later. Note however that effect (i) is always greater in S_{hot} than in S_{cold} given the above inequality, and effect (ii) is also always greater in S_{hot} than in S_{cold} for the

same reason. Hence it follows that those who transgress in S_{hot} but not in S_{cold} will always donate in S_{hot}. Similarly, those who transgress only in S_{cold} but not in S_{hot}, will not donate in S_{hot}. Hence it follows that for any given pool of subjects the likelihood of both a norm violation and no donation is always greater in S_{cold} than in S_{hot}, i.e., $\Pr((x_1, y_1), \text{"no donation"} \mid S_{\text{hot}}) < \Pr((x_1, y_1), \text{"no donation"} \mid S_{\text{cold}})$. ■

Proposition 2 (Hypothesis 2) *Suppose the DM chooses (x_2, y_2) in S. Suppose in S_{hot} the donation option is known ex-ante. She is then more likely to choose (x_1, y_1) in S_{hot} than in S, and if she does so, she will donate as well.*

Proof. Note first that since $u_{g,y} > u_{g,x}$, conditional on a norm violation, planning to donate at $t = 2$ can improve the decision maker's overall utility at $t = 1$. In other words, at $t = 1$, the DM can prefer to violate a norm and donate, to not violating the norm and not donating, while at the same time prefer the latter to simply violating the norm. Formally,

$$\sum_{t=1}^{4} u(x_1 - d, y_1 + d, g_t) \geq \sum_{t=1}^{4} u(x_2, y_2, 0) \geq \sum_{t=1}^{4} u(x_1, y_1, g_t)$$

Furthermore, since conditional on a norm violation, $g_2 > g_1$, if the DM at $t = 1$ would prefer to donate, given that $MRS_{t=2} < MRS_{t=1}$, she will have the same preference at $t = 2$. Hence if she she violates the norm in S_{hot}, but not in S, she must also donate in S_{hot}. ■

Proposition 3 (Hypothesis 3) *Suppose that the donation option is known ex-ante. If the DM ex-ante strictly prefers S_{hot} to S_{cold}, then she will choose to donate in S_{hot}. If the DM ex-ante strictly prefers S_{cold} to S_{hot}, then she will choose not to donate in S_{cold}.*

Proof. Note first that since the set of feasible options are the same, an ex-ante strict preference for one choice problem over another can arise only if the choice behavior in the two problems differ. It then follows that a strict preference can only arise when there is dynamic inconsistency in at least one of the choice problems, i.e., a norm violation occurs in M.

Given an initial norm violation from M, it follows from the proof of Hypothesis 1 that whenever "no donation" is preferred at stage $t = 2$, it must also be preferred at stage $t = 3$. For the same reason, whenever a "donation" is preferred at stage $t = 3$, it is also preferred at stage $t = 2$. Hence the difference in final allocation choices can only arise when "donation" is only implementable in S_{hot} but not in S_{cold} or when "no donation" is implementable in S_{cold} but not in S_{hot}. As a consequence, a strict preference for S_{hot} over S_{cold} implies a strict preference for a "donation" from D. Similarly, a strict preference for S_{cold} over S_{hot} implies a strict preference for "no donation" from D. This implies that if the decision maker strictly prefers S_{hot} over S_{cold} , she will donate, and if she strictly prefers S_{cold} over S_{hot}, she will not donate. ■