Economics and Philosophy, **29** (2013) 199–212 © Cambridge University Press
doi:10.1017/S0266267113000199

MORAL CLEANSING AND MORAL LICENSES: EXPERIMENTAL EVIDENCE

PABLO BRAÑAS-GARZA
Middlesex University London, UK

MARISA BUCHELI
Universidad de la República, Uruguay

MARÍA PAZ ESPINOSA
BRiDGE: Universidad del País Vasco, UPV/EHU, Spain

TERESA GARCÍA-MUÑOZ
GLOBE: Universidad de Granada, Spain

Research on moral cleansing and moral self-licensing has introduced dynamic considerations in the theory of moral behaviour. Past bad actions trigger negative feelings that make people more likely to engage in future moral behaviour to offset them. Symmetrically, past good deeds favour a positive self-perception that creates licensing effects, leading people to engage in behaviour that is less likely to be moral. In short, a deviation from a 'normal state of being' is balanced with a subsequent action that compensates the prior behaviour. We model the decision of an individual trying to reach the optimal level of moral self-worth over time and show that under certain conditions the optimal sequence of actions follows a

We thank Pier Paolo Battigalli for his comments on a previous version of the paper. The editors and three anonymous referees made useful suggestions that allowed us to improve the paper. Financial aid from MICINN (ECO2009-09120, ECO2010-17049 and ECO2012-35820), Gobierno Vasco (DEUI, IT-313-07), Fundación Ramón Areces R+D 2011 and Junta de Andalucía-Excelencia (P07.SEJ.02547) is gratefully acknowledged. A previous version of this paper was presented in SEET-2011 Agadir, IMEBE-2011 Barcelona and Fundación Urrutial Elejalde Summer School-2011 San Sebastián.

regular pattern which combines good and bad actions. To explore this phenomenon we conduct an economic experiment where subjects play a sequence of giving decisions (dictator games). We find that donations in the previous period affect present decisions and the sign is negative: participants' behaviour in every round is negatively correlated to what they did in the past. Hence donations over time seem to be the result of a regular pattern of self-regulation: moral licensing (being selfish after altruistic) and cleansing (altruistic after selfish).

1. INTRODUCTION

How and *why* altruistic behaviour emerges is a critical question. Altruistic behaviour is not costless. Every single altruistic action generates a cost for the donor and thus good deeds need to come with a benefit to compensate the cost. There are a number of classical evolutionary arguments such as kin selection – Hamilton rule – or reciprocal altruism (Fehr and Fischbauer 2003) and a series of papers have dealt with more self-centred arguments such as identity, guilt-aversion or warm-glow, that describe the benefits of being moral (see Akerlof and Kranton 2000; Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2007; Aguiar *et al*. 2010). In this paper, we add the moral self-licensing and moral cleansing arguments that explore the relationship between past, present and future moral behaviour.

One motivation for good deeds is their positive effect on moral self-worth. When past actions make people feel confident about their moral behaviour, their moral self-regard could be high enough to allow them to engage in morally dubious behaviour in the present (Zhong and Liljenquist 2006; Merritt *et al.* 2010). This is the central argument of the moral self-licensing literature. In a review of the evidence, Merritt *et al.* (2010) present the two most frequent moral-licensing mechanisms used in the literature: the moral credits and the credentials models. The moral credits model uses a moral bank account metaphor: good deeds purchase 'moral credits' that diminish the discomfort of engaging in bad deeds in the future. In the credentials model, actions affect the meaning of future actions: the value of an ambiguous behaviour will be valued through the lens of past good deeds. As a consequence, a good action gives self-license to future transgressions. Note that according to the mechanism of the first model, the licensed person gets involved in what he considers a bad action but this is not the case in the second model. So the damage to self-value is different and we may expect self-license to lead to a lower number of transgressions under the moral credits than under the credentials mechanism.

In turn, immoral behaviour has a negative effect on moral self-worth. After engaging in bad deeds, people follow a moral behaviour to recover the lost self-worth; this mechanism is the so-called moral cleansing behaviour (see Sachdeva *et al*. 2009). One well documented example is that in response to sins, many religious practices require bodily purification.

Taking into account the two types of behaviour, moral licensing and moral cleansing, Sachdeva *et al*. (2009) consider 'moral behavior as being embedded within a larger system that contains competing forces. Moral or immoral actions may emerge from an attempt to find balance among these forces'. The process is symmetric: every deviation from the normal behaviour is subsequently balanced with either a more moral action (moral cleansing) or less moral action (moral licensing). In their experiment, Sachdeva *et al*. (2009) show that affirming a moral identity (participants were asked to write a self-relevant story containing positive traits) leads people to donate less to charities (moral licensing); when moral identity is threatened (story containing negative traits), generosity in donations to charity is a means to regain some lost self-worth (moral cleansing).

Our paper provides further evidence on this phenomenon of moral self-regulation in a dynamic context where a sequence of moral decisions is made. We analyse data from an economic experiment where subjects play a sequence of 16 dictator games, each with a different randomly chosen recipient (anonymity conditions). All the games have the same structure and they are framed. Besides a blind (baseline) game, we use three types of frames regarding the information given about gender (male/female), income (poor/rich) and political preferences (right wing/left wing) of the dictator and the recipient, to generate 15 different environments. Each subject played the 16 games in a different random order to control for order effects. The recipient was different at each round so that the subject could not compensate a high donation to a recipient in one round with subsequent low donations to the same recipient; as a consequence, if we observe a negative correlation, we can be sure that it is moral cleansing and not simple compensation to the recipient.

This design tries to recreate the sequentiality of decisions, to test the hypothesis of moral self-regulation that would lead individuals to reverse previous moral or immoral behaviour. The alternative hypothesis is that subjects would always behave according to their moral standards and therefore we would observe no reversion.

The design of the experiment is well suited to the dynamics of behaviour based on self-worth. First, the order of treatments is randomized so that in each round the donors are facing different frames and to some extent situations with a different moral content. Second, subsequent decisions cannot compensate the recipients because only one decision is paid and the recipients are different subjects. Thus, we eliminate any compensation effect and isolate the effect of moral self-worth. Finally, the treatments' frames refer to socio-demographic characteristics so that the sequence does not suggest nor lead the donor to focus on a dynamic based on moral licensing or moral cleansing behaviour.

Our estimation technique takes into account the dynamics of these actions; we estimate how a donation by each individual (dt-1) affects the subsequent one (dt). We find that donations over time follow an auto-regressive process of order one (AR(1)) with a negative coefficient.[1] We draw two important conclusions from this analysis:

(i) the negative sign of the effect of the immediate past actions (d_{t-1}) on current choices (d_t) indicates that subjects reverse in every round what they did in the past;
(ii) the length of the auto-regressive process (AR(1)) indicates that only the previous period affects present behaviour. Hence, subjects tend to balance in period t what they did in period $t-1$.

Our result implies that self-regulation is not a long memory process, since only the previous period matters. This could be due to the fact that decisions in our experiment are not overly asymmetric so that only one period is sufficient to reverse what the subjects did in the past.

The rest of the paper is organized as follows. In Section 2, we set the theoretical framework. Section 3 describes the experimental design and procedures, Sections 4 and 5 contain the results and an analysis of robustness and in Section 6 we present some concluding remarks.

2. THEORETICAL FRAMEWORK

This section presents a dynamic model for the paradox of moral self-regulation (Sachdeva *et al*. 2009). In this theoretical framework, decisions with a moral content have to be made over time and subjects self-regulate to achieve their utility-maximizing level of moral self-worth.

Assume that up to period t, a subject i is at her utility-maximizing level of goodness G^*, that is, she has made decisions that have placed her in a situation where her moral self-worth is at the chosen level. This level of goodness G^* is obtained taking into account the costs and benefits of moral self-worth (for example, G^* could maximize the net benefit, B(G)-C(G), the benefits minus the costs).

We assume that preferences concerning the level of goodness are single-peaked and symmetric around G^* so that at each period *t*, subjects minimize the distance $|G_t\text{-}G^*|$, where G_t is the moral self-worth at t.

[1] AR(p) is an auto-regressive process of length p, p being the number of previous periods which affect actual values.

https://doi.org/10.1017/S0266267113000199 Published online by Cambridge University Press

In a dynamic context, this level G^* may be difficult to maintain since life requires difficult decisions with a moral content to be made over time. To represent this, assume that at period t the individual must make a decision that will put her at a level of moral self-worth either higher than G^* or lower. We assume for simplicity that a single decision has to be made each period, it cannot be avoided and decisions are not neutral, that is, decisions always affect moral self-esteem.

The decision at each period t is binary; the subject may either have good behaviour, which increases goodness by $g_t > 0$ or bad behaviour which decreases it by $b_t > 0$. Depending on the decision made, she will enter period t+1 having a level of moral self-worth $G_t = G^* + g_t$ or $G_t = G^* - b_t$.

The subject decides again in period t+1 (see Figure 1 in the Appendix). If her decision was *good* in period *t*, she should choose *bad* in period t+1 as long as $G_{t+1} = G^* + g_t - b_{t+1}$ is closer to the optimal value G^* than $G_{t+1} = G^* + g_t + g_{t+1}$.

Note that if the decision *good* or *bad* is always symmetric, that is, if $g_t = b_t = g = b$ for all t, then the subject should always choose the decision opposite to the previous one, to get as close as possible to G^*.

Assuming that $g_t = g$ and $b_t = b$ for all t, what happens if the decisions are not symmetric ($g \neq b$)? Take for example the case $b = 3g$, that is, the cost of a bad action is three times the benefit of a good one. Then starting from G^*, to minimize $|G_t - G^*|$ at each t, the subject's decisions should follow a regular pattern: (... gg b ggg b ggg b ...), three good actions are always followed by a bad one.

More generally, if $b = ng$, where n is an integer and an even number, starting from G^* the optimal sequence of actions follows a regular pattern: (n/2 actions g, one action b, n actions g, one action b, n actions g, one action b ...). If n is an odd number, the sequence is: ((n+1)/2 actions g, one action b, n actions g, one action b, n actions g, one action b ...). If 1/n is an integer, and even number: (n/2 actions b, one action g, n actions b, one action g, n actions b, one action g ...). If 1/n is odd: ((n+1)/2 actions b, one action g, n actions b, one action g, n actions b, one action g ...).[2]

This result implies that individuals self-regulate to achieve their optimal level of moral self-worth G^* and this self-regulation follows a regular pattern. Whenever decisions with a moral content cannot be avoided, individuals will alternate bad and good actions over time.

In our experiment, we test whether these regular patterns predicted by the theory appear when subjects have to make sequential decisions involving moral self-worth.

[2] See the appendix for proof of this result.

https://doi.org/10.1017/S0266267113000199 Published online by Cambridge University Press

3. THE EXPERIMENT

3.1. The dictator game

In the dictator game (Forsythe *et al.* 1994), the first player (dictator), determines an allocation (split) of some endowment (such as a cash prize). The second player, the recipient, simply receives the share of the endowment left by the dictator. The recipient's role is entirely passive.

Formally, given an endowment of size D, the dictator must decide any value of $d_i \in [0,D]$ to pass to the recipient. Therefore the final distribution of benefits is a pair:

$$(D - d_i,\ d_i)$$

where $D - d_i$ is the dictator's benefit and d_i is the donation. Since the Nash equilibrium is giving zero to the recipient, any strictly positive donation, $d_i > 0$, is interpreted as pure altruism.[3]

3.2. Participants

One hundred and seventy-six subjects distributed in four sessions participated in the experiment (dictators and recipients). We will focus only on the sample of 88 dictators (32% of women) since recipients do not play any active role in our analysis. The participants were undergraduate students at the *Universidad de la República* (Uruguay), recruited via posters and verbal information in regular class time.

3.3. Procedures and materials

The subjects were given verbal and printed information: they had to take 16 decisions and each one was explained on one page of a printed booklet. They were not allowed to speak to one another and they were seated in such a way that they could not see the written responses of the other subjects.

The baseline treatment consisted of a standard dictator game in which each participant was a dictator or a recipient (the participants knew that no one would play both roles). The dictator had to allocate 10 bills of 20 Uruguay pesos (around US $10) between herself and a randomly chosen student placed in a different room. Following List (2007) instructions, the task was explained on one sheet of paper inside a printed booklet and the possible payoffs were presented on a line in which the subject had to mark her decision with a circle. The amount of money ranked from

[3] See Brañas-Garza and Espinosa *et al.* (2006); Brañas-Garza *et al.* (2009, 2012).

0 pesos (left-end) to 200 pesos (right-end) and the donations were restricted to multiples of 20 including zero, that is 0, 20, 40, . . . 200.

The rest of the treatments were identical to the baseline (blind) with the exception of the framing. In order to frame the task, we used information that participants gave at the moment they registered for the experiment: sex, income category and ideological category. This information was used to label the participants as women/men, rich/poor and right-wing/left-wing.[4] This label appeared on the first sheet of the booklet. Thus, the participant knew it at the beginning of the experiment.

The sheets for each task had a similar appearance. In three treatments, the donor was told that the recipient would know the donor's sex, income category or ideological category, respectively. In these sheets, the six possible donor's characteristics were printed in a column under the title 'information about you': man, woman, poor, rich, left, right. One of them was circled: the circle indicated the donor's characteristic that the recipient would know. In six treatments, the donor knew one characteristic of the recipient (sex, or income category, or ideological category). In these sheets, the six possible recipient's characteristics were printed in a column under the title 'information about your partner'. One of them was circled, indicating that recipient was a subject with that specific characteristic. We used one of these sheets with no circled characteristic for the baseline treatment. In another six treatments, the donor knew his own characteristic, a characteristic of the recipient and besides, he was informed that the recipient would know the frame (for example, the recipient would know that the donation was done from a woman to a man). The characteristics of the donor and the recipient were always within the same category (sex, income, political preferences).

The entire booklet consisted of 16 tasks that were presented in a different random order for each subject. For the randomization we used the following procedure: we associated to each task a random number based on a uniform distribution ranged from 0 to 1. We ordered the tasks based on the order of those random numbers; we assigned this ordered booklet to the first participant. We repeated the procedure for all the participants. This is an important characteristic of the design: as in each round the donors are facing different frames, even if all participants had the same preferences, we would not necessarily observe an equalizing pattern common to all subjects.

[4] We asked the participants to fill out a questionnaire where they revealed their personal ideological position and the socio-economic status of their household in a 10-step scale where 1 was extremely poor/left wing and 10 was extremely rich/right wing. In order to build binary labels (poor/rich, left-wing/right-wing), the threshold was the median value of the reported distributions.

We paid only one decision (randomly chosen) to each dictator, which avoids the effect of accumulation of earnings in the course of the session. Besides, the use of different recipients and frames at each decision helped to maintain subjects' interest. Notice that once a decision is taken, subsequent decisions by the same subject cannot actually hurt or help the same recipient. Thus, if the donor makes what he thinks is a selfish (generous) decision, the subsequent action will not compensate the prior recipient since the recipients are different individuals; with this design any compensation effect affects exclusively moral self-worth.

The money donated to recipients was delivered to them in a different session. Taking all the games into consideration, the average donation in the 16 games was 60.8 Uruguay pesos (around US $3) and the average recipient's earnings were 57.5 Uruguay pesos (around US $3).[5]

4. RESULTS

According to the theoretical framework described in Section II, we would expect a negative correlation between the donation at t and that at period t+1. We test this hypothesis in a dynamic panel data model where we estimate the donation at period t (d_t) as a function of past donation (d_{t-1}):

$$d_{it} = \alpha_i + \gamma d_{i,t-1} + x_{it}^{'}\beta + v_{it},\ i = 1, \ldots, 88\ \text{individuals}, t = 1, \ldots, 16\text{rounds}$$

where α_i denotes the unobserved individual-specific time-invariant fixed effect;[6] x_{it} is the it-th observation of the explanatory variables, in our case, treatment dummies and temporal trend; the disturbance terms v_{it} have zero mean, constant variance and are uncorrelated across time and individuals.

We use two-step GMM[7] estimators with the Windmeijer correction using lagged levels (t−2, t−3 and t−4) of the dependent variable as instruments (Arellano and Bond 1991; Windmeijer 2005).

Table 1 shows the results of three regressions. In the first column, the only covariate is the previous donation (d_{t-1}); in regression (2) we also include the treatment dummies and in regression (3) we add a temporal trend.

The treatments dummies group the treatments in four categories. One of them is the baseline that is the omitted treatment. The included dummies capture the effect on the donor's decision of three types of information: (a) the fact that one of the characteristics of the donor will be

[5] Only one round was paid and the average donation in that round was 57.5 pesos.
[6] By construction, $d_{i,t-1}$ is correlated with α_i. It then makes no sense to use a random effects estimation method since one regressor is correlated with the individual effects.
[7] Generalized Method of Moments.

https://doi.org/10.1017/S0266267113000199 Published online by Cambridge University Press

	(1)	(2)	(3)
Round (t)	–	–	0.195
			(0.430)
d_{t-1}	**−0.085**	**−0.088**	**−0.075**
	(0.035)	**(0.036)**	**(0.031)**
Constant	61.125	48.081	45.157
	(0.000)	(0.000)	(0.000)
Treatment controls	Not	Yes	Yes
Arellano-Bond serial correlation test	−0.635	−0.808	−0.695
	(0.525)	(0.419)	(0.487)
Instruments	*40*	*43*	*44*
Sample Size	*1220*	*1220*	*1220*

p-values in parentheses.

TABLE 1. Moral cleansing and licensing

known; (b) a characteristic of the recipient; and (c) the fact that both donor and recipient would share the information about their characteristics. Note that these variables do not capture the effect of the characteristic (for example the effect that the recipient would be a man or a woman) but the type of information (for example, the effect of the knowledge of the recipient's characteristic).

In the three estimations, the coefficient of past donation (d_{t-1}) is negative, significant at 5% and less than one in absolute value. Besides, the trend is not significant. In the bottom part of Table 1 we show Arellano–Bond tests.[8]

The important result here is that donations follow a stationary AR(1) process with negative coefficient. Hence, subjects tend to balance a donation above the mean in a round with a donation below in the following round. We have also computed the autocorrelation coefficient of order 1 for the donations of each individual during the 16 rounds and 67.95% of the individuals have negative coefficient (Figure 2 in the appendix shows donations for each individual).

This result does not support the alternative hypothesis that subjects would always donate according to their moral standards and show consistent preferences for a given level of donation. On the contrary, the pattern of donations over time shows a self-regulation behaviour and emerges as the result of a systematic process of dynamic equalization:

[8] Arellano and Bond (1991) proposed a test to detect serial correlation in the disturbances. Validation of the instrumental variables is obtained given that the null hypothesis of this test (no serial correlation) is not rejected.

	(1)	(2)	(3)
Round (t)	–	–	−0.021
			(0.938)
d_{t-1}	**−0.128**	**−0.135**	**−0.115**
	(0.073)	**(0.048)**	**(0.045)**
d_{t-2}	−0.055	−0.064	−0.052
	(0.257)	(0.165)	(0.197)
Constant	67.010	55.564	53.991
	(0.000)	(0.000)	(0.000)
Treatment controls	Not	Yes	Yes
Arellano-Bond serial correlation test	0.177	0.159	0.140
	(0.860)	(0.873)	(0.888)
Instruments	*39*	*42*	*43*
Sample Size	*1130*	*1130*	*1130*

p-values in parentheses.

TABLE 2. Moral cleansing and licensing, with 2 lags

moral licensing (being selfish after altruistic) or cleansing (altruistic after selfish).

We also check if donations follow an AR(2) process. We find that the coefficient of d_{t-2} is not significant, whereas the coefficient of d_{t-1} is still negative and significant (see Table 2).

5. ROBUSTNESS

As a simple robustness test, we check whether our results change in the final rounds of the experiment. Table 3 shows the same regressions as Table 1 but using only the last 12 periods ($t = 5, 6, \ldots, 16$) and the last 8 periods ($t = 9, 10, \ldots, 16$), respectively. Given that every individual played the 16 games in a different random order, we lose different treatments' observations for each individual.

There are no remarkable differences when we compare results from Table 1 and Table 3. Hence, there are no differences in behaviour at the end of the game. Using all or only the final rounds of the experiment does not make any difference.

Lastly, as an additional robustness test, we check whether our key finding holds for different subsamples. We estimate the AR(1) model – with controls – for a sample of 68 subjects randomly selected, that is, we drop 20 subjects. We repeat the exercise removing another 20 different subjects and finally we repeat the process a third time. Table A.1, in the appendix, shows the estimated AR(1) coefficients for the three

	Rounds 5 to 16			Rounds 9 to 16		
	(4)	(5)	(6)	(7)	(8)	(9)
Round (t)	–	–	0.021	–	–	0.240
			(0.931)			(0.625)
d_{t-1}	**−0.098**	**−0.098**	**−0.101**	**−0.119**	**−0.132**	**−0.137**
	(0.061)	**(0.066)**	**(0.036)**	**(0.078)**	**(0.042)**	**(0.010)**
Constant	63.335	51.237	50.744	65.612	50.090	47.245
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Treatment controls	Not	Yes	Yes	Not	Yes	Yes
Arellano-Bond serial correlation test	−0.592	−0.677	−0.724	−0.369	−0.536	−0.584
	(0.554)	(0.498)	(0.469)	(0.712)	(0.592)	(0.559)
Instruments	37	40	41	25	28	29
Sample Size	1046	1046	1046	695	695	695

p-values in parentheses.

TABLE 3. Robustness checks

sub-samples (elimination #1, #2 and #3). The results are similar to those presented in the text.

Two additional robustness checks are shown at the bottom of Table A.1. We estimate the AR(1) coefficients when observations from the baseline are not included; results are even stronger (p-value = 0.01). We also run a model removing subjects who donate the same quantity in all rounds, and the results were similar.

Our experimental results indicate that the coefficient of the participant's previous donation is significant and negative, which is consistent with our hypothesis that over time individuals self-regulate to attain the optimal level of self-worth.

6. CONCLUSIONS

This research contributes to the literature that focuses on the role of moral cleansing and moral self-licensing on behaviour. Our results show that donations show stability across time as the result of equalization. In the estimations, the past donation (d_{t-1}) coefficient is always negative, significant and its absolute value is less than one – indicating that subjects who behaved nicely yesterday are selfish today and vice versa. In short, a systematic moral self-licensing and moral cleansing pattern emerges.

Our findings are related to the current theories of identity (Akerlof and Kranton 2000). When decisions are not morally neutral, each decision affects the sense of identity and implies a deviation from the optimal level of moral self-worth, which requires a compensating subsequent decision.

We have identified this self-regulation behaviour empirically and our results are consistent with moral licensing and moral cleansing.

Our results suggest that self-image maintenance is the factor responsible for the negative autocorrelation in behaviour. Further experimental research could be directed to examine this relationship more closely.

REFERENCES

Aguiar, F., P. Brañas-Garza, M. P. Espinosa and L. Miller. 2010. Personal identity. A theoretical and experimental analysis. *Journal of Economic Methodology* 17: 261–275.

Akerlof, G. and R. E. Kranton. 2000. Economics and identity. *Quarterly Journal of Economics* 115: 715–753.

Arellano, M. and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Battigalli, P. and M. Dufwenberg. 2007. Guilt in games. *American Economic Review* 97: 170–176.

Brañas-Garza, P. and M. P. Espinosa. 2006. Altruism with social roots: an emerging literature. *Desarrollo y Sociedad* 58: 245–260.

Brañas-Garza, P., M. A. Durán and M. P. Espinosa. 2009. The role of personal involvement and responsibility in unfair outcomes: a classroom investigation. *Rationality and Society* 21: 225–248.

Brañas-Garza, P., M. A. Durán and M. P. Espinosa. 2012. Favouring friends. *Bulletin of Economic Research* 64: 172–178.

Charness, G. and M. Dufwenberg. 2006. Promises & partnership. *Econometrica* 74: 1579–1601.

Fehr, E. and U. Fischbauer. 2003. The nature of human altruism. *Nature* 425: 1579–1601.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6: 347–369.

List, J. A. 2007. On the interpretation of giving in dictator games. *Journal of Political Economy* 115: 482–493.

Merritt, A. C., D. Effron and B. Monin. 2010. Moral self-licensing: when being good frees us to be bad. *Social and Personality Psychology Compass* 4/5: 344–357.

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

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51.

Zhong, C. and K. Liljenquist. 2006. Washing away your sins: threatened morality and physical cleansing. *Science* 313(5792): 1451–1452.

APPENDIX 1. THE OPTIMAL SEQUENCE OF ACTIONS

We assume that whenever the subject is indifferent between a good and a bad action, she chooses the one with the lowest payoff: b if $b < g$ and g if $g \leq b$.

Assume $b = ng$, n a positive integer and an even number. Starting from G^*, the subject has the choice between G^*-b or G^*+g, and she should choose G^*+g since it is closer to G^*. The same is true in the following periods up to period $n/2$. After $n/2$ periods, the subject is at $G^*+(n/2)g$. She is then indifferent between $G^*+(n/2)g$ and $G^*+(n/2)g–b = G^*+(n/2)g –ng = G^*-(n/2)g$, so that the next decision should be g since $G^*+(n/2)g–b$ is closer to G^* than $G^*+(n/2)g+g$.

https://doi.org/10.1017/S0266267113000199 Published online by Cambridge University Press

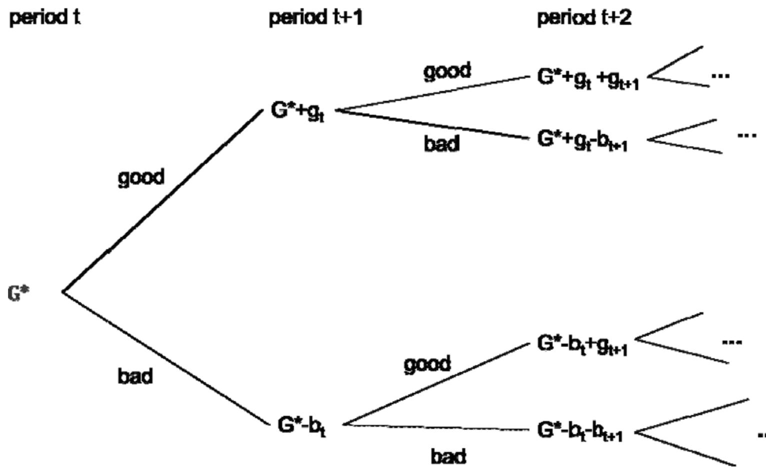

FIGURE 1. Decision tree

Example. Assume $G^* = 100$, $b = 10$ and $g = 2$. Then $n = 5$. The subject would follow the sequence (g,g,g,b,g,g,g,g,g,b, . . .): 100, 102, 104, 106, 96, 98,100,102, 104, ...

When n is not an integer, the optimal sequence of actions takes a slightly more complicated form. For example, if $n = 3.5$, the optimal sequence is (. . . 3 gs, b, 4 gs, b, 3 gs, b, 4 gs, b,. . ...).

The case $b = g/n$ follows by symmetry.

Appendix 2

	AR(1) Coefficient	p-value	Sample Size
Removal of 20 participants			
elimination #1	−0.115	0.026	944
elimination #2	−0.105	0.057	941
elimination #3	−0.089	0.068	940
without "Blind"	−0.133	0.014	989
without "Constant"	−0.089	0.034	1094

TABLE A.1. Additional robustness checks

Appendix 3

Figure 2 shows individual donations. Subjects with constant donations in all rounds are not shown (11 participants).

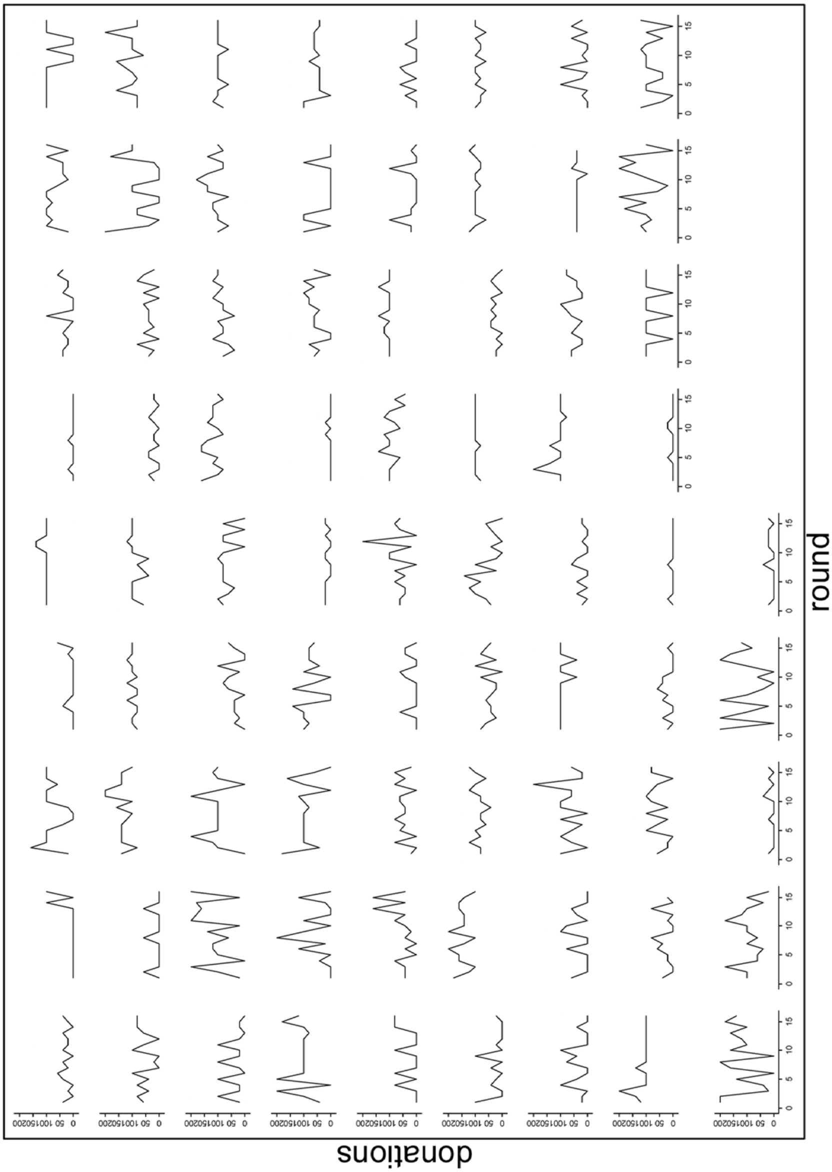

FIGURE 2. Individual donations

https://doi.org/10.1017/S0266267113000199 Published online by Cambridge University Press