I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11592

Using Ethical Dilemmas to Predict Antisocial Choices with Real Payoff Consequences: An Experimental Study

David L. Dickinson
David Masclet

JUNE 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11592

Using Ethical Dilemmas to Predict Antisocial Choices with Real Payoff Consequences: An Experimental Study

David L. Dickinson
Appalachian State University and IZA

David Masclet
University of Rennes

JUNE 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Using Ethical Dilemmas to Predict Antisocial Choices with Real Payoff Consequences: An Experimental Study

Anti-social behaviours are costly to organizations, and the ability to identify predictors of such behaviours can be valuable. In this paper, we used a within-subjects laboratory design to study choices in the well-known (hypothetical) Trolley problem as well as in a real payoff money-burning experiment that can inform our understanding of moral preferences and antisocial behavior. Choices in both environments respond to incentives (i.e., the relative price of the ethical decision). Trolley problem decisions are consistent with previously known results – individuals prefer no action over action, and they prefer to avoid direct over indirect responsibility when negative consequences would be similar in either instance. In analyzing the determinants of anti-social money burning, our data identify money burning due to inequality aversion, but we also find evidence of pure nastiness (burning money of others to increase one's advantageous inequality). Importantly, we find that willingness to commit ethically dubious acts in the Trolley problem significantly predicts money burning and, more specifically, nastiness. We conclude that choices in hypothetical environments can predict consequential and inefficient antisocial behaviours. Also, utilitarian behaviour in the Trolley dilemma is *not* linked to antisocial money burning, which contrasts with conclusions in the literature.

JEL Classification: C90, C91, Z10

Keywords: experiments, money burning, ethical dilemmas, anti-social behavior, trolley problem

Corresponding author:
David L. Dickinson
Economics Department
Appalachian State University
Boone, NC 28608
USA

E-mail: dickinsondl@appstate.edu

1. Introduction

Economists have long challenged the assumption of *homo economicus* and recognize that people are not always own-payoff maximizing. Rather, they may be altruistic, fairness-minded, cooperative, or perhaps even anti-social. Using laboratory methods with real payoff, experimental economics has shown that participants in dictator games often share their endowment (Forsythe et al. 1994; Hoffman et al. 1994), they reciprocate in gift exchange or trust environments (Berg et al. 1995; Fehr et al. 1998), and they contribute positive amounts in public goods games (see surveys in Ledyard, 1995; Chaudhuri, 2011). Studies focusing on the darker side of human behaviour are surprisingly more limited in economics. This is all the more surprising given that unethical behavior within organizations is not rare and often results in high costs for the entire society. Anti-social behaviours can result in relational, workplace, or other costs to society that are nontrivial. Cyber-sabotage is now a growing concern, for example (see Line et al, 2014), and survey data from the U.S. and Europe document antisocial workplace behaviours that include mistreatment, verbal abuse, and sabotage, with estimates indicating that such antisocial behaviours may impact 10%-35% of people in the workplace (see Charness et al, 2013).[1] Field data examples often pose difficulties in our attempt to understand the core determinants of antisocial tendencies given that they may be confounded with self-interest, hidden from view, or contaminated by reputational concerns. Behavioural metrics that help identify the likelihood that someone may engage in antisocial behaviours can therefore be a useful way to prevent antisocial behavioural costs and improve overall welfare.

Moral philosophers have examined choices in ethical dilemmas, such as the famous Trolley problem (Foot, 1967), while in economics there has been some focus on the antisocial dimension of behaviour through experimental research on money burning (Zizzo and Oswald, 2001; Zizzo, 2004) and nasty behavior (Abbink and Sadrieh, 2009, Abbink and Herrman, 2009). An investigation into the components of moral preferences that predict antisocial behaviour, to our knowledge, does not yet exist. Our objective is to construct moral identifiers derived from the Trolley problem to predict antisocial (and consequential) behaviour in a money burning game.

The Trolley dilemma has captivated moral philosophers for decades (Foot, 1967; Thomson, 1985; Spranca et al, 1991; Petrinovich et al, 1993). The dilemma describes a runaway trolley that, unless an action is taken, will run over one or more persons on a train track unable to escape. Various versions of the problem exist (see Shallow et al, 2011), but we focus on perhaps two of the most classic scenarios. The first assumes a runaway trolley is bound to kill several individuals on a main set of tracks unless one pulls a lever to divert the trolley onto a side track where it will kill anyone who may be on the side track. Such a decision scenario is considered an "indirect" (or impersonal) moral choice in the sense that pulling the lever to save lives indirectly but intentionally results in the death of those on the side track. A

[1] Data from the U.S. includes research from the Workplace Bullying Institute (http://www.workplacebullying.org/) and the Bureau of Labor Statistics (www.bls.gov , considering that at least some of the workplace stoppage data represents an exercise of incurring some cost in order to impose even larger costs on a counterpart), and data from the French Ministry of Employment are from the SUMER medical monitoring survey of workplace risks (surveying over 50,000 workers in the 2010 wave, see https://www.eurofound.europa.eu/observatories/eurwork/articles/working-conditions/france-working-conditions-and-occupational-risks-sumer-2010).

second version is considered a "direct" (or personal) moral choice scenario where instead of pulling a lever one must push an individual onto the main track (and that person will die) in order to save the one or more individuals on the main track (Thomson (1985).

The Trolley dilemma has come under fire for its lack of realism and external validity (Rai and Holyoak, 2010; Bauman, et al 2014; Kahane, 2015). Nevertheless, others have found it useful for studying various components of moral reasoning (e.g., Cushman et al, 2006; Greene et al, 2001; Greene et al, 2011), such as the identification of behavioural norms or highlighting that certain moral dilemmas preferentially engage emotional centers in a way that may be important in predicting choice (e.g., Greene et al., 2001). Still others have noted how the Trolley dilemma can highlight the difference between acts of omission versus commission (Spranca et al, 1991; Cox et al, 2017), which is a relevant distinction in courts of law. And, while past criticism of the Trolley dilemma may have seemed justified due to the unrealistic nature of the decision it presents, the relevance of the Trolley dilemma is at a higher level than perhaps ever before with the recent rise in ethical concerns surrounding self-driving vehicles. Bonnefon et al (2016) highlight how the moral dilemma relates to the social dilemma of Autonomous Vehicle (AV) adoption, whereby most survey respondents agreed an AV should be programmed to sacrifice its passenger(s) if more pedestrians are saved as a result, but these same individuals thought it much less appropriate to program the AV as such if one's own life were at stake.

In economics, studies focusing on the antisocial dimension of behaviour include the seminal studies by Zizzo and Oswald (2001) and Zizzo (2004), whose results show that many subjects are willing to incur a real cost in order to reduce other's payoffs—"money burning". Money burning may be explained by inequality aversion (Zizzo and Oswald, 2001), but it may also result from a pure pleasure of being nasty (Abbink and Sadrieh, 2009).[2]

Our goal is to contribute to the literature in the following ways. First, we attempt to investigate the determinants of antisocial behaviours using measures derived from both a real-payoff money burning experiment and a hypothetical Trolley dilemma. Are money burning decisions driven by inequality aversion concerns, pure nastiness, pre-emptive retaliation? In conjunction with our exploration of the determinants of money burning, we hope to further our understanding of some key determinants of (un)ethical choices. Indeed, specific Trolley dilemmas can identify more ethically dubious choices, thus allowing us to classify one's morality. We are also able to distinguish an immoral act of commission from an immoral act of omission, which yields a more rich set of morality variables to consider as predictors of money burning decisions.

Secondly, our paper will also contribute to the literature by investigating the extent to which costly money burning decisions and Trolley choices obey the law of demand. Responses to ethical dilemmas surrounding the adoption of autonomous vehicle technologies, which bear resemblance to the Trolley dilemma, were recently shown to be sensitive to the relative number of lives saved in the scenario (Bonnefon, 2016). Within the context of demand for costly

[2]Individual characteristics may be yet another factor that explains money burning decisions. For instance, some previous studies have shown that high basal testosterone is associated with an increased threshold for conflict (see Carney and Mason, 2010, and references therein).

punishment. Nikiforakis and Normann (2008) showed that voluntary contributions to provide a public good increase monotonically in punishment effectiveness, and Anderson and Putterman (2007) found that the price of punishment is a significant determinant of punishment demand. These previous studies suggest that even the moral domain of choice should obey the law of demand. To our knowledge, no previous study has attempted to investigate the role played by relative cost in the context of money burning decisions, and our set of Trolley dilemmas allows us to explore efficiency (utilitarian outcomes), which implies that the dominant concern should be minimizing the number of lives lost.

Finally, we investigate the relationship between decisions in the money burning experiment and choices in the Trolley experiment. Specifically, by exploiting our within-subjects strategy method design, we examine the ability of moral identifiers derived from Trolley decisions to predict consequential choices in the money burning game. The predictive validity of ethical dilemma responses has been of interest in the recent literature although there is no clear cut evidence regarding the relationship. Some authors have recently argued that hypothetical ethical dilemmas are *not* useful for predicting behaviour in real dilemmas (Bostyn et al, 2018).[3] Other studies suggest that antisocial personality types may be more willing to make a difficult moral choice that produces a utilitarian outcome (Koenigs et al, 2007; Bartels and Pizarro, 2011; Bracht and Zylbersztejn, 2017).[4] In this paper we attempt to contribute to this existing literature that attempts to explore the connection between some version of anti-social behavior and choice in hypothetical moral dilemma. The main originality of our work is that in the typical ethical dilemma studied it is always utilitarian to sacrifice the life because even more are saved. Thus, existing studies cannot separate utilitarian behaviour from less savoury preferences that may perversely desire to cause a death through action. Our design can separate utilitarian from immoral preferences by including Trolley dilemmas not typically considered, which then allows us to construct moral identifiers.[5] These moral identifiers are shown to have power predicting antisocial behaviour in the consequential (i.e., non-

[3] Bostyn et al (2018) examine whether Trolley dilemma responses predict propensity to deliver electric shocks to mice in dilemmas with similar but nonfatal scenarios.

[4] Using self-report measure of antisocial personality tendencies (Bartels and Pizarro, 2011) or patients with brain damage in regions important to emotion generation (Koenigs et al, 2007), these studies find tendencies towards increased utilitarianism in individuals with antisocial personality. Another recent study (Bracht and Zylbersztejn, (2017) is quite related to ours in that it also examines ethical choice in hypothetical dilemmas as well as in a consequential money transfer game. The differences in our study are notable, however. First, we do not pool data across *indirect* versus *direct* moral choices as they do, which is important given we identify a highly significant ($p < .01$) impact of this factor on one's willingness to take action (we also show that one of their key results is qualified in our findings by conditioning on the direct versus indirect nature of the dilemma). Secondly, both our hypothetical and consequential choice experiments vary the relative efficiency or cost of one's action, thus allowing a more thorough examination of ethical and antisocial choice. Finally, we use a morality measure derived from the Trolley dilemma to predict behavior in the consequential money burning game, while Bracht and Zylbersztejn (2017) use the outcome in their consequential game as a regressor in the ethical dilemma choice. While of potential interest, we find the causation of hypothetical-to-consequential choice more valuable in terms of implications and use as a potential screening or identification mechanism (e.g., job application/interview screener).

[5] It is therefore important to note that many moral dilemmas confound the utilitarian choice from the choice one might make for non-utilitarian reasons. For example in the typical Trolley dilemma, it is utilitarian to pull the switch or push the individual, and yet one may be willing to act not because more lives are saved than lost, but rather because one prefers or perversely enjoys being responsible for someone's death.

hypothetical) money burning game. If choices in hypothetical moral thought experiments can help identify those likely to make antisocial choices, it may be possible to improve overall welfare.

To preview our main findings, we observe that choices in both environments respond to incentives (i.e., the relative price of the ethical decision. In analyzing the determinants of anti-social money burning, our data identify money burning due to inequality aversion, but we also find evidence of pure nastiness (burning money of others to increase one's advantageous inequality). Trolley problem decisions are consistent with previously known results—individuals prefer no action over action, and prefer to avoid direct over indirect responsibility when negative consequences would be similar in either instance. Importantly, we find that willingness to commit ethically dubious acts in the Trolley problem significantly predicts money burning and, more specifically, nastiness. For example, those willing to commit an immoral act of omission in the Trolley dilemma are significantly more likely to display nasty money burning behavior (i.e., destruction of another's payoff *even when* one is already at a payoff advantage). We conclude that ethically questionable choices in hypothetical environments can predict consequential and inefficient antisocial behaviours. Also, utilitarian behaviour in the Trolley dilemma is not linked to antisocial money burning, which contrasts with conclusions in the literature.

2. Experimental design

2.1. Overview

Both experiments were administered in strategy method format, where decisions were elicited on multiple decisions prior to a randomized draw of one (in the incentivized money burning task) for real payoff. The Trolley dilemma was always administered first to the subjects, and we allowed the option to opt out and *not* complete the Trolley dilemma if one preferred to avoid the dilemma altogether. Table 1 describes the menu of dilemmas administered in our version of the Trolley dilemma. Importantly, we highlight that our choice menu allows us to examine how the likelihood of taking action responds to the number of people saved (X) relative to killed (Y). We are also able to examine preferences for inaction over action when the number of lives lost would be unaffected (i.e., X=Y dilemmas). And finally, we can examine how one's likelihood to take action differs if action is indirect (i.e., pull a lever to divert the runaway trolley) versus a more direct act (i.e., push an individual(s) onto the track to stop the runaway trolley), which we call *INDIRECT* versus *DIRECT* decision scenarios.

In what follows, we scored immorality as derived from the Trolley dilemma choices as follows: *Immoral Omission* is an indicator variable equal to one if a subject chose to *not* take action in the *DIRECT* and *INDIRECT* (X,Y)=(6,0) scenarios, where action would save 6 individuals without any lives being lost as a result. Another dichotomous variable, *Immoral Commission*, is equal to one if the subject chose action in both the (X,Y)=(6,6) and (1,1) scenarios of both the *INDIRECT* and *DIRECT* choice dilemmas. In the case of *Immoral Commission*, the subject prefers to be responsible (via action) for a given number of deaths rather than simply allow that same number of deaths to occur as a bystander. We created a final variable by taking a subject's average propensity to act in the remaining scenarios not used in

the construction of the *Immoral Omission* or *Immoral Commission* variables. Such a variable, *Action Propensity*, represents one's willingness to take action, though it also describes utilitarian preferences in our dilemmas.

For the money burning game, a key treatment variable is whether only one in the pair (the "decider") or both could burn money.[6] Specifically, in the *Bilateral Burn* treatment, both players could mutually and simultaneously destroy a portion of each other's payoffs. That is, each of the two subjects in a randomly matched pair made money burning decisions and two random decisions were selected such that each subject was both a decider and passive recipient (i.e., potential money burn victim) in a consequential money burning choice. In *Unilateral Burn*, subjects were randomly assigned as decider or passive recipient *before* decision making, and only the deciders made decisions. After decisions were made in all 9 money burning scenarios, deciders and recipients were randomly matched and one scenario was selected at random to determine the payoff of both players in the money burning game.[7] This process was common knowledge.

2.2. Experimental procedures

The experiment was computerized and administered using the Z-tree platform (Fischbacher, 1999). We recruited 150 subjects at the University of Rennes 1 (France), each subject participated in only session, and none of the subjects had participated in a similar economic experiment. A total of 9 sessions were conducted (with 14 to 18 subjects per session), where in each session subjects were administered the Trolley dilemma and the Money Burning game. Table 3 contains summary information about number of participants in each treatment of the Money Burning game, which is identified by the order of presentation of the scenarios (see footnote 4) and whether the *Unilateral* or *Bilateral Burn* treatment. Importantly, subjects were given the choice to opt out of the Trolley Dilemma for whatever reason. A total of 12 subjects (8%) chose to opt out of the Trolley dilemma task, and we use this "opt-out" in the analysis of money burning choices below.

A session lasted approximately one hour (this includes the time spent to read the instructions). At the end of the experiment, one task was randomly selected for each pair of randomly matched subjects (and random role assignments, in the case of *Unilateral Burn*

[6]We also varied the ordering of the Money Burning (x,y) pairs in the menu received (presenting the decision maker's endowment in *Increasing*, *Decreasing*, or *Random* order. Each subject saw only one ordering). We did not have a formal hypothesis regarding the ordering of the money burning scenarios, and later analysis documents that the varied ordering does not significantly impact outcomes in the task. We considered variation in the order more exploratory. Of course, there is no theoretical reason to believe that the ordering should matter, but this possibility has been investigated on the more well-known risky choice task (Holt and Laury, 2002) lottery choice menu (see Bruner, 2009).

[7] Our design made use of the strategy method, as opposed to direct elicitation method, in order to generate multiple observations from each subject in each decision experiment (other than the passive recipients in the money burning game, which were randomly selected *prior* to decision making in that game). Brandts and Charness (2011) survey experimental results comparing strategy method versus direct elicitation and conclude that the strategy method for response elicitation, in general, provides a conservative estimate of what choice would be using direct response elicitation—in our case, money burning choices may therefore be a conservative estimate of outcomes one would find using direct elicitation of just a single response in a single scenario.

treatments). Payments were made anonymously at the end of the session and the average earnings were 25.52 Euros per subject.

3. Behavioural Predictions

3.1. Trolley Problem

The literature identifies two clear predictions we can make regarding outcomes in the Trolley dilemma. First, a widely reported result is that individuals are more willing to take action and save lives in the *INDIRECT* frame where a level is pulled, as compared to the *DIRECT* frame where an individual is pushed onto the track, holding constant the relative number of lives saved. This is related to the distinction between personal and impersonal moral dilemmas (Greene et al, 2001). Thus, our first hypothesis stems from this "contact principle" (Cushman et al, 2006). This hypothesis implies that for each pair (X,Y) of lives (saved, lost), we predict an individual is more likely to take action in the Indirect frame.

> **H1 (Trolley):** For each (X,Y) dilemma, action is more likely in the *INDIRECT* Frame

Secondly, both individuals and courts of law consider an act of omission to be a lesser "sin" than an act of commission that results in similar consequences (see Cox et al, 2017). This principle with respect to the Trolley dilemma has been labelled the "action principle". In our set of Trolley dilemma choices we can then focus on (X,Y) pairs (6,6) and (1,1), where an equal number of individuals would perish whether or not action is taken. In these cases, we hypothesize a lesser likelihood to act given a preference to not be responsible (via action) for the deaths. To take action in such cases could be considered an immoral act of *commission.* Also of interest would be the (0,6) Trolley dilemma, where action costs no lives. In such a dilemma, to *not* take action would be consider an immoral act of *omission*.

> **H2 (Trolley):** When lives lost are unaffected (X=Y dilemmas), inaction is preferred over action (moral omission). Also, action is preferred over inaction when action is costless (any (X,0) dilemma) (moral commission).

Rejection of H2 implies acts of *Immoral Omission* or *Commission*.

This final Trolley Problem H3 hypothesis posits that, independent of one's moral tendencies, a decreased relative cost of action will increase action likelihood. In other words, we predict a downward sloping demand curve for lives saved in this moral dilemma.

> **H3 (Trolley):** The likelihood of action will increase in the relative number of lives saved.

3.2. Money Burning

First consider the theoretical predictions in the money burning game. Purely selfish decision makers are completely indifferent towards the other person's well-being and would not incur costs to affect others' payoffs, and Utilitarian preferences would never choose to reduce total welfare. Thus, under the assumption of either pure selfishness or Utilitarianism, there should be no money burning. Behavioral considerations, however, may be invoked to generate alternative hypotheses of interest. For example, individuals may be inequality averse such that utility depends not only on one's own payoff but also on the equality of the income distribution (see Fehr and Schmidt, 1999).[8] In our design, advantageous inequality (i.e., $x \geq y$) cannot be reduced by burning money, and so we are only able to identify disadvantageous inequality aversion if money is burned when $x < y$). Alternatively, individuals may have so-called *nasty* preferences, which we define as a desire to lower others' payoff to gain or increase one's advantageous inequality (i.e., money burning scenarios where $x \geq y$). Because decisions in the *Bilateral Burn* treatment are impacted by any (unmeasured) expectations of others' money burning choices, the pure effect unconfounded by expectations is measured from the *Unilateral Burn* treatments. Thus, we have our first money burning hypothesis:

> **H4 (Money Burning):** Money will not be burned in *Unilateral burn* treatments (*homo economicus* or Utilitarian preferences assumed).

Rejection of H4 when $x < y$ (tasks A1-A4 in Table 2) is evidence of disadvantageous inequality aversion, and when $x \geq y$ is evidence of nasty preferences.

If individuals burn others' money and it is common knowledge that money burning is *Bilateral*, then observed money burning may be due to inequality aversion (for $x < y$) or nastiness (for $x \geq y$) or anticipatory negative reciprocity. This type of "pre-emptive retaliation" may be exist even in the presence of disadvantageous inequality if both players in a pair make a simultaneous choice to burn money. In other words, in the simultaneous choice *Bilateral*.treatments one may burn the counterpart's money on the expectation that the counterpart may burn some of one's payoff (see Abbink and Sadrieh, 2009).

> **H5 (Money Burning)***:* Money burning will be similar in *Unilateral* and *Bilateral Burn* treatments.

Rejection of H5 is evidence that people believe others will burn money.

[8]Indeed, a very appealing hypothesis about distributional preference is inequality aversion (see Loewenstein et al. 1989; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). These approaches assume that individual utility depends not only on one's own payoff but also on the equality of the income distribution.

Just as in the case of the Trolley dilemma, we expect that antisocial tendencies to burn resources of others will nevertheless respond to the price of doing so. Previous studies have shown that punishment decisions in a VCM context obey the law of demand. (Nikiforakis and Normann, 2008; Anderson and Putterman, 2006). Based on these papers' findings one may reasonably conjecture that money burning decisions also obey the law of demand and, though the cost of burning is fixed in our design, the amount burnt varies. This implies that the cost of burning money *relative* to one's payoff in the chosen payoff distribution varies and we can expect an increase in money burning when the relative cost of burning money is low. This leads to H6.

H6 (Money Burning): Burning money will be negatively related to its relative cost.

Finally, a unique contribution we offer in the paper is to consider moral descriptors of one's choices in the Trolley dilemma as an explanatory variable regarding one's choice to burn money. Immoral acts of commission and omission are defined in H2 based on a subset of our Trolley dilemmas. Someone who takes action in the (X,Y) Trolley dilemmas not implicated in H2 has a higher *Action Propensity* (or is more Utilitarian). The morality of those with higher *Action Propensity* is difficult to assess given that one may be willing to sacrifice one or more lives to save others for more than one reason. Such reasons may include both ethically dubious reasons (i.e., I prefer to push someone to save others) as well as Utilitarian reasons (e.g., I will do whatever leads to the most lives saved (least lives lost)). For this reason, clean moral descriptors of immorality for our final hypothesis focus on metrics derived from a subset of the Trolley dilemmas:

H7 (Money Burning): Moral descriptors derived from the Trolley dilemmas—(X=Y) and (X,0) dilemmas—will predict increased money burning.

4. Results

4.1 Trolley Results

We start by showing summary data from the subjects who made Trolley dilemma choices in Figures 1 and 2 (9 Trolley dilemma choices per subject). Of the 150 participants in our experiment, n=12 subjects opted out of the Trolley dilemma, leaving us with n=138 Trolley subject decision makers (we code these "Trolley opt-out" subjects for later use as a regressor in the money burning estimations in the next section).[9] Figure 1 shows the proportion of choices in each treatment (*Direct and Indirect* dilemmas) for the subset of dilemmas that hold constant the number of individuals saved. Left to right on the horizontal axis shows dilemmas that increase the number of individuals sacrificed for a constant X=6 individuals saved. Two things stand out in Figure 1: the proportion of individuals who take action decreases as the relative

[9] We conducted a probit estimation of the determinants of the decision to opt out of the Trolley dilemma. Though few subjects opted out, we found one variable, "happiness" (self-reported current level of happiness in life) was a marginally significant determinant of the opt-out choice ($p < .10$). Specifically, those self-reporting higher levels of life happiness were marginally more likely to opt out of the Trolley dilemma.

cost, Y/X, increases; more surprisingly, greater than 20% of subjects did *not* choose to take action in the (6,0) dilemma where 6 individuals could be saved at zero cost, and some chose (indirect) action in the (6,6) dilemma where the same number of individuals would perish even if nothing were done. Both represent instances of what we call "Trolley immorality". Figure 2 organized the remaining subset of Trolley choices to hold constant the number of individuals who perish at Y=1. Left to right in Figure 2 organizes the Trolley dilemmas by a decreasing number of lives saved while holding constant the number who perish.

We again see that action in the Trolley dilemma is responsive to the relative cost (or effectiveness) of the action--subjects are less likely to take action when the relative cost, Y/X, increases (or, as the relative benefit X/Y decreases). In Figure 2, we also see that a nonzero number of subjects choose an immoral act of commission in Trolley dilemma (1,1) where action was preferred even though an individual would perish even with inaction.

As noted in Section 3 (Experimental Design), we elicit choices in the strategy method to maximize data generated per subject. Due to multiple decisions per subject, all models in Table 4 include standard errors clustering at the individual subject level. The model structure is a Probit estimation where the dependent variable is equal to one if that subject chooses to take "action" (i.e., pull the lever or push the individual(s) in that particular dilemma scenario). The different columns of Table 4 show estimations using different sets of independent variables. The first two columns use a dummy variable for each (X,Y) pair of lives saved (X) and sacrificed or killed (Y) compared to the omitted baseline scenario of (X,Y)=(6,0). Columns 3-5 replace the dummy variables with continuous variables measuring the number of lives sacrificed and saved.

The dummy variable identifying the *DIRECT* version of each Trolley dilemma has a consistently negative and significant coefficient estimate across all models, which supports Hypothesis H1. Individuals are significantly less likely to take action when it is a more personal moral dilemma (action would be direct) compared to impersonal (action would be indirect). Interestingly, this effect is somewhat muted for male subjects as seen by the significant and positive coefficient on *Male*DIRECT* in model 5.[10] Because many of the dilemmas confound morality of choice with utilitarian actions, we next examine hypothesis H2 using only the subset of Trolley dilemmas (X,Y)=(6,6), (1,1), and (6,0). The comparison of coefficients in our Table 4 estimations are not a transparent way to assess whether a statistically significant number of subjects chose *action* in the (6,6) and (1,1) dilemmas, or *inaction* in the (6,0) dilemma. Rather, we can test the null hypothesis that the proportion of subjects choosing *action* in the (X=Y) dilemmas is equal to zero against the alternative that it is greater than zero. For the test of immoral action, we test the null hypothesis that the proportion of subjects choosing *action* in the (6,0) dilemma is equal to 100% against the alternative hypothesis that it is less than 100%. For the case of n=138 observations, the observed proportions in both the case of *DIRECT* and *INDIRECT* framing of the Trolley dilemmas lie outside of the 95% confidence interval. This

[10] This result is somewhat related to the gender result found in Bracht and Zylbersztejn (2017), who find males more likely to take action in a set of moral dilemmas. The study includes a variety of dilemmas in addition to a limited number of Trolley dilemmas, but they do not distinguish dilemmas in their set that involve a *direct* versus *indirect* action in the moral choice. As such, our result is an important qualification of what they report given our evidence suggests the gender effect may not be as general as they suggest.

evidence implies rejection of H2 in favor of the existence of immoral acts of omission and commission being greater than zero.[11]

Finally, we show support for Hypothesis H3 by using estimates in models (4) and (5) of Table 4. Here the marginal effect on *# Lives Sacrificed* (*Y*) holds constant the *# Lives Saved* (*X*), and vice versa. Thus, the negative and positive, respectively, effects of these variables on the likelihood of taking action confirm that Trolley choices respond to the relative number of lives saved to lost, which supports Hypothesis 3. In short, action in the Trolley dilemma responds to incentives and displays a downward sloping demand curve for lives saved. Nevertheless, a nontrivial number of subjects make choices that can be classified as immoral acts of commission or omission in the Trolley dilemma choices designed to highlight such dubious moral viewpoints. Having established the results from our Trolley Dilemma, shown them consistent with the extant literature, and also documented our morality metrics as revealing, we next turn to the results from the Money Burning game.

4.2 Money Burning Results

Summary results from the Money Burning game are shown in Figures 3 and 4, and in Table 5. Figures 3 and 4 summarize the frequency of money burning choices for categories of (*x*,*y*) pairs. Figure 3 shows money burning choices in each possible scenario, for both instances of *Unilateral Burn* and *Bilateral Burn*. Figure 4 highlights the apparent downward trend in money burning as the cost of burning money is larger relative to the recipient's budget—money burning also obeys the law of demand (Hypothesis 6). Table 5 shows the total number of instances (out of 9 scenarios) in which the subject burned money, on average (top row), along with summary information on the proportion of money burning choices for the different possible types of money burners. Depending on the relationship between the decider's payoff, *x*, and the passive recipient's payoff, *y*, one can consider decisions to burn money are reflective of disadvantageous inequality aversion (burning money when $x<y$) or nasty preferences (burning money when $x > y$). Others may never burn money (Homo Economicus or Utilitarian preferences), and some burn money in all 9 scenarios and so reflect an unconditional desire to behave antisocially (i.e., destroy resources and reduce total welfare of the pair). Though we have only limited data from deciders in the *Unilateral Burn* treatments, the *Bilateral Burn* data in Table 5 reflect similar proportions of burn choices in both *Bilateral* and *Unilateral Burn* treatments. This most likely indicates that *Bilateral Burn* choices are *not* driven primarily by expectations that others will burn money.

Tables 6 and 7 show results from Probit estimations of the probability that someone makes the dichotomous choice to burn money and select the End Distribution over the Start Distribution in the Table 2 scenarios. Errors in both tables are clustered at the level of the individual subject, and we report marginal effects in the tables. The set of independent variables

[11]One sample proportions tests, the Z statistic cannot be calculated for the boundary hypothesized proportions of 0% and 100%, and so we rather calculate our tests using null hypothesis proportions of 1% and 99%, respectively. Our conclusions remain intact even if allowing for a 5% « error » in decision making (at the p <.10 level for the (6,6) *DIRECT* and (6,0) *INDIRECT* dilemmas, but at the p < .01 level in all other cases). That is, if assuming that a small percentage of subject may make mistaken choices in ours ample, our conclusions are H2 result is largely unchanged.

in Table 6 includes: controls for the presentation order of the (*x*,*y*) distributions in the Money Burning menu set (*Random, Increasing, or Decreasing*); an indicator variable for the scenarios where burning was *Unilateral* (*Bilateral* is the reference group); indicator variables capturing payoff equality/inequality in the different (*x*,*y*) payoff distributions; a variable measuring the relative cost of money burning compared to one's own payoff; a variables measuring the simple preferences in the Trolley dilemma to take action); a set of subject-specific controls. Importantly, model (3) in Table 6 and the models in Table 7 include indicator variables to identify whether the subject committed *Immoral Commission* or *Immoral Omission* in the Trolley problem (6,6), (1,1), and (6,0) dilemmas. So, these two indicator variables capture a sense of the moral preferences of the subject as derived from the Trolley choices, and the test of significance on their coefficients is a test of whether such measures from hypothetical decision scenarios may yet hold power to predict decision in consequential decision tasks than contain at least some type of moral element. Table 7 focuses on estimates separating the subsamples of the data for the (*x*,*y*) distributions where $x < y$ (disadvantageous payoff inequality) versus $x \geq$ y (advantageous inequality).

We first focus on the results in Table 6, and note that there is no evidence that the ordering of the (*x*,*y*) options matters. Our data support for hypothesis H5—money burning is equally likely in *Bilateral* and *Unilateral Burn*, which suggests that beliefs that others will burn money do not impact money burning decisions in our data. Statistically significant positive coefficients on *Income<other* in all three models support rejection of our hypothesis H4 in favor of the alterative hypothesis of disadvantageous inequality aversion (Fehr and Schmidt, 1999) when $x > y$ in the Start Distribution (Table 2). The marginally significant ($p < .10$) coefficient on the *Relative Cost* of burning money indicates that money burning is responsive to how much of one's payoff the burning choice will cost—a lower relative budget impact of burning marginally increases the likelihood that one burns money, which supports hypothesis H6. Model (2) includes an indicator variable for those who opted out of the Trolley dilemma, and we find a marginally significant impact of *Opt-Out* on the probability that one will burn money. This variable is absent is model (3) where we include the Trolley immorality measures as regressors, which necessarily implies we focus on the money burning data from those who also completed the Trolley dilemma choices. Importantly, in model (3) of Table 6, we find evidence that making a morally dubious choice(s) in the Trolley dilemma predicts a significantly increased likelihood of money burning.[12] This is support for hypothesis H7. Thus, we offer first evidence in the literature, to our knowledge, that moral indicators from a hypothetical dilemma can predict significant increases in anti-social money burning choices with real payoff consequences.[13]

Table 7 shows results of related estimations where the subsample of $x < y$ versus $x \geq y$ are used as a way to identify general money burning from "nastiness", which would be defined as a willingness to burn money for $x \geq y$ payoff distributions (i.e., a willingness to pay to burn money even when my payoff is at least as higher my counterpart's). The results from Table 7

[12]The difference between the impact of *Immoral Commission* versus *Immoral Omission* is not statistically significant ($p > .10$ for the Wald test of coefficient equality)

[13] Model (3) also indicates a marginally significant impact of higher self-reported life happiness predicting a lower probability that one burns money.

show that the *Diff Income* (= $|y - x|$) represented in the Start Distribution only predicts a significantly higher probability of money burning when a subject's payoff is lower than the counterpart's, which again implies rejection of H4 in favor of disadvantageous inequality aversion that is sensitive to the size of inequality. Looking at the *advantageous inequality* subset of data in model (2), we see that the relative cost of burning money matters in terms of anti-social "nasty" choices. The higher the advantageous payoff inequality is in our design, the lower the relative cost to make the money burning choice. For this reason, we see the predicted marginally higher nastiness in those scenarios where the decider is at the largest payoff advantage (see also right-half of Figure 3). This offers some evidence of nasty preferences as the alternative upon rejection of the Utilitarian or Homo Economicus H4 hypothesis.

Interestingly, the immorality measures from the Trolley dilemma are significant predictors of the probability one burns money. Model (1) shows that both immoral acts of commission and omission in the Trolley dilemma predict a 36%-38% increase in the likelihood one burns money ($p < .01$ in both cases. The difference between these two effects is statistically insignificant, $p > .10$). We identify predictors of nastiness in model (2) of Table 7 and a key result is that we find that the immoral act of omission in Trolley dilemma #12 (i.e., *not* acting when 6 lives could be saved at the expense of zero lost lives) predicts a 30% increased likelihood of making a "nasty" money burning choice ($p < .01$). In a sense, our strongest way to judge morality from the Trolley dilemma is whether someone chose the immoral act of omission. In sum, we find strong support for hypothesis H7 and conclude that Trolley morality, though hypothetical, can be a significant predictor of consequential antisocial decisions.

5. Discussion.

In this paper we attempt to explain some key determinants of costly antisocial behaviours using measures derived from both a money burning game and a moral thought experiment. While ethical dilemmas and thought experiments have been of significant interest to moral philosophers for decades, we believe our study to be unique. Our particular innovation has been to use responses in the iconic Trolley dilemma to generate immorality indicators that have predictive power regarding one's decision making in consequential environments. The consequential environment we explore allows for costly antisocial choice and may be considered a type of behavioural marker for the likelihood of costly behaviours in field settings.

Our results highlight the importance of the relative cost of the ethical behaviour across the domains of both the hypothetical Trolley dilemma and the consequential Money Burning game. Subjects are more likely to make an ethically dubious choice if the costs of doing so are lower. Aside from identifying typical response patterns in the Trolley dilemma, we identified choices made from our set of Trolley dilemmas that would constitute morally questionable acts of omission or commission. We then estimated a significant increase in the likelihood of burning money for those subjects identified as willing to commit an immoral act of omission or commission in the Trolley dilemma. Upon further investigation, we found that immoral Trolley respondents' increased willingness to burn money was linked more strongly to disadvantageous inequality aversion than to nastiness. Nevertheless, we identified that choice in one Trolley scenario (not typically considered in the existing literature) is a highly significant

predictor of the probability of nasty money burning. These results call into question some recent conclusions in the literature regarding increased Utilitarianism among those with anti-social personality traits (Bartels and Pizarro, 2011). Specifically, our research connects immoral preferences and actual anti-social behaviour in a stylized game.

As always, there are limitations to our study. First, it is likely the case that reputational concerns may be important if one is aware that selection based on Trolley dilemma responses may be at stake. And of course, the validity of a hypothetical ethical dilemma may always be a point of concern. For this reason, one of our main purposes is to highlight that response patterns in such hypothetical dilemmas may be instructive towards an understanding of consequential behavioural tendencies. At some level, the criticism of selection bias would apply to any number of hypothetical or self-response instruments used to screen individuals or assess situational risk. We believe the key is that we first understand the link between hypothetical responses and consequential behaviours, because researchers often have no alternative approach to study high stakes moral dilemmas.

Our hope is that this research will stimulate further investigations into the value of hypothetical choices towards predicting outcomes in other non-hypothetical but related decision domains. These findings may have interesting implications for how hypothetical scenario instruments could be used to screen individuals for antisocial tendencies that could be costly to an organization. Because the type of anti-social decision making we studied involves resource destruction when outcome inequality is present, it is intriguing to consider that the markers for such behavioural tendencies may already exist in well-known hypothetical thought scenarios. Imagine that an employer could use responses to the Trolley dilemma as a way to identify workers who may be more willing to engage in antisocial resource destruction. While this may seem like the type of worker one may wish to avoid (i.e., avoid such individuals in designing self-driving auto accident avoidance algorithms), those willing to destroy resources may have value to the employer in certain specialized roles (e.g., lead negotiator who must credibly be willing to walk away from a contractual arrangement or wage negotiations). Of course, such implications of our findings are themselves only a thought experiment, but we hope them to be useful at motivating why this may be a fruitful area for research extensions. If choices in hypothetical dilemmas can serve as behavioural markers that predict real world ethical choice, then we feel this is a useful step forward in an important area of behavioural research.

Current technological developments, such as the increased experimentation and utilization of drones and self-driving vehicles, render hypothetical moral dilemmas like the Trolley dilemma increasingly relevant to policy-makers as society attempts to understand barriers to technology adoption and implementation (e.g., Crockett, 2016). Our results have

implications for how choices in hypothetical moral dilemmas may be used to understand or even possibly forecast certain types of behaviour in real world environments.

REFERENCES

Abbink K, & Herrmann B. (2011). The moral costs of nastiness. *Economic Inquiry*, *49*(2): 631-633.

Abbink K, & Sadrieh A. (2009): The pleasure of being nasty. *Economics Letters*, *105*: 306-308.

Anderson, C. M., & Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, *54*(1), 1-24.

Bauman CW, McGraw AP, Bartels DM, & Warren C (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9): 536-554.

Bartels DM, & Pizarro DA. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*: 154-161.

Berg J, Dickhaut J., & McCabe K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1): 122-142.

Bonnefon JF, Shariff A, & Rahwan I (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293): 1573-1576.

Bostyn, DH, Sevenhant S, & Roets A (2018). Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas. *Psychological Science*, [first published online May 9, 2018]: DOI: 10.1177/0956797617752640

Bracht J, & Zylbersztejn A. (2017). Moral judgments, gender, and antisocial preferences: An experimental study. *Working Paper #1630, GATE.*

Brandts J, & Charness G (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, *14*(3): 375-398.

Bruner DM. 2009. Changing the probability versus changing the reward. *Experimental Economics*, *12*(4): 367-385.

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: when the ends justify the means. *Journal of Experimental Social Psychology*, *46*(4), 668-671.

Charness G, Masclet D, & Villeval MC (2013). The dark side of competition for status. *Management Science*, *60*(1): 38-55.

Chaudhuri A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1): 47-83.

Cox JC, Servátka M, & Vadovič R (2017). Status quo effects in fairness games: reciprocal responses to acts of commission versus acts of omission. *Experimental Economics*, *20*(1): 1-18.

Crockett M (2016). The trolley problem: Would you kill one person to save many others? *The Guardian*, Dec. 12, 2016.

Cushman F, Young L, & Hauser M (2006). The role of reasoning in moral judgments: Testing three principles of harm. *Psychological Science, 17*: 1082–1089.

Fehr E, Kirchsteiger G, & Riedl A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, *42*(1): 1-34.

Fehr E, & Schmidt KM (1999): A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*: 817-868.

Foot P (1967). The problem of abortion and the doctrine of the double effect, *Oxford Review*, *5* : 5-15.

Forsythe R, Horowitz JL, Savin NE, & Sefton M (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3): 347-369.

Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, & Cohen JD (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*: 364–371.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, & Cohen JD (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science, 293*(5537): 2105–2108.

Hoffman E, McCabe K., Shachat K., & Smith V (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*(3): 346-380.

Holt CA, & Laury SK. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5): 1644-1657.

Kahane G (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience, 10*(5): 551-560.

Koenigs M., Young L, Adolphs R, Tranel D, Cushman F, Hauser M, & Damasio A (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*(7138): 908-911.

Ledyard J (1995) Public goods: A survey of experimental research. In J. Kagel, A. Roth (Eds.), The Handbook of Experimental Economics, Princeton Univ. Press, Princeton: 111-194.

Line MB, Zand A, Stringhini G, & Kemmerer R (2014). Targeted attacks against industrial control systems: Is the power industry prepared? In *Proceedings of the 2nd Workshop on Smart Energy Grid Security,* November(2014): 13-22. ACM.

Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, *11*(4), 358-369.

Petrinovich L, O'Neill P, & Jorgensen M (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*: 467–478.

Rai TS, & Holyoak KJ (2010). Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science*, *34* : 311–321.

Shallow C, Iliev R., & Medin D (2011). Trolley problems in context. *Judgment and Decision Making*, *6*(7) : 593-601.

Spranca M, Minsk E, & Baron J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*, 76–105.

Thomson J. (1985). The Trolley Problem, *Yale Law Journal*, *94*: 1395–1415.

Zizzo D (2004): Inequality and Procedural Fairness in a Money Burning and Stealing Experiment. In F.A. Cowell (ed.), Research on Economic Inequality, vol. 11. Elsevier.

Zizzo D, & Oswald AJ (2001): Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Economie et de Statistique 63-64*: 39-62.

Table 1. Trolley Dilemmas.

	INDIRECT				**DIRECT**			
	Are you willing to pull a lever to divert the trolley to a different track to save X people, where Y on that side track will die.				Are you willing to kill Y people by pushing them onto the track to save X people?			
Trolley Dilemma #	Total people killed	Total people saved			Total people killed	Total people saved		
1	Y=6	X=6	Yes	No	Y=6	X=6	Yes	No
2	Y=5	X=6	Yes	No	Y=5	X=6	Yes	No
3	Y=4	X=6	Yes	No	Y=4	X=6	Yes	No
4	Y=3	X=6	Yes	No	Y=3	X=6	Yes	No
5	Y=2	X=6	Yes	No	Y=2	X=6	Yes	No
6	Y=1	X=6	Yes	No	Y=1	X=6	Yes	No
7	Y=1	X=5	Yes	No	Y=1	X=5	Yes	No
8	Y=1	X=4	Yes	No	Y=1	X=4	Yes	No
9	Y=1	X=3	Yes	No	Y=1	X=3	Yes	No
10	Y=1	X=2	Yes	No	Y=1	X=2	Yes	No
11	Y=1	X=1	Yes	No	Y=1	X=1	Yes	No
12	Y=0	X=6	Yes	No	Y=0	X=6	Yes	No

Note: Trolley dilemmas numbered here for discussion in the text (dilemmas were not numbered for subjects)

Table 2: Money Burning choice tasks (*Increasing* treatment)
Subjects chose the Start or End Distribution for each of the 9 tasks

Task #	**Start Distribution**	**Damage**	**Burning costs**	**End Distribution**
A1	(50 , 250)	50	10	(40 , 200)
A2	(50 , 200)	50	10	(40 , 150)
A3	(50 , 150)	50	10	(40 , 100)
A4	(50 , 100)	50	10	(40 , 50)
A5	(50 , 50)	50	10	(40 , 0)
A6	(100 , 50)	50	10	(90 , 0)
A7	(150 , 50)	50	10	(140 , 0)
A8	(200 , 50)	50	10	(190 , 0)
A9	(250 , 50)	50	10	(240 , 0)

Table 3. *Summary of the Money Burning Treatments*

Session	Participants	Treatment Description
1	18	*Increasing*—Unilateral burn
2	16	*Increasing*—Bilateral burn
3	18	*Increasing*—Bilateral burn
4	18	*Decreasing*—Unilateral burn
5	16	*Decreasing*—Bilateral burn
6	14	*Decreasing*—Bilateral burn
7	18	*Random*—Unilateral burn
8	16	*Random*—Bilateral burn
9	16	*Random*—Bilateral burn
Total	150	(n=96 Bilateral Burn, n=54 Unilateral burn)

Table 4. Probability of Action (Pull level or Push person) in Trolley Dilemma

Marginal Effects Reported (robust st errors in parenthesis)

Independent variable	(1)	(2)	(3)	(4)	(5)
DIRECT Action	-0.2061***	-0.2083***	-0.1917***	-0.1997***	-0.2974***
	(0.0314)	(0.0317)	(0.0302)	(0.0305)	(0.0502)
*Male*DIRECT*	---	---	---	---	0.1863***
					(0.0570)
(X,Y)=(6,0)	Reference	Reference	---	---	---
(X,Y)=(6,6)	-0.6207***	-0.6243***	---	---	---
	(0.0263)	(0.0268)			
(X,Y)=(6,5)	-0.3996***	-0.4026***	---	---	---
	(0.0422)	(0.0427)			
(X,Y)=(6,4)	-0.3765**	-0.3794***	---	---	---
	(0.0448)	(0.0454)			
(X,Y)=(6,3)	-0.3298***	-0.3324***	---	---	---
	(0.0488)	(0.0495)			
(X,Y)=(6,2)	-0.3095***	-0.3120***	---	---	---
	(0.0495)	(0.0503)			
(X,Y)=(6,1)	-0.2391***	-0.2403***	---	---	---
	(0.0529)	(0.0542)			
(X,Y)=(5,1)	-0.2547***	-0.2570***	---	---	---
	(0.0508)	(0.0517)			
(X,Y)=(4,1)	-0.2851***	-0.2877***	---	---	---
	(0.0468)	(0.0476)			
(X,Y)=(3,1)	-0.2991***	-0.3019***	---	---	---
	(0.0472)	(0.0478)			
(X,Y)=(2,1)	-0.3329***	-0.3359***	---	---	---
	(0.0443)	(0.0449)			
(X,Y)=(1,1)	-0.5839***	-0.5881***	---	---	---
	(0.0266)	(0.0268)			
# Lives Sacrificed (Y)	---	---	-0.1102***	-0.1117***	-0.1126***
			(0.0073)	(0.0072)	(0.0073)
# Lives Saved (X)	---	---	0.0873***	0.0885***	0.0892***
			(0.0062)	(0.0061)	(0.0062)
Religion ∈ [1 ,10]	---	-0.0026	---	-0.0022	-0.0023
(10=very important)		(0.0118)		(0.0113)	(0.0114)
Happiness ∈ [1 ,10]	---	0.0205	---	0.0196	0.0198
(10=highest current life happiness)		(0.0204)		(0.0196)	(0.0197)
Age	---	0.0289	---	0.0271	0.0270
		(0.0187)		(0.0178)	(0.0178)
Male (=1)		0.1020*	---	0.0979*	-0.0013
		(0.0599)		(0.0576)	(0.0657)
Observations	3312	3312	3312	3312	3312
#Clusters/Participants^	138	138	138	138	138
Log likelihood	-1921.8550	-1889.6102	-1998.7432	-1968.3391	-1954.1628

Notes: *.10, **.05, ***.001 for the 2-tailed test. Standard errors clustered at the individual subject level. Total observations reflect n=138 subjects who opted to complete the Trolley dilemma task. Each of the 138 made 12 *Direct* and 12 *Indirect* Trolley dilemma choices.

Table 5: Descriptive statistics of money burning decisions

	All	**Unilateral Burn**	**Bilateral Burn**
# Money Burning choices (out of 9)			
Mean	***1.33***	***1.22***	***1.35***
[standard deviation]	[2.11]	[1.82]	[2.19]
Never burn	**82**	**18**	**64**
Homo Economicus or *Utilitarian*	(66.67%)	(66.67%)	(66.67%)
Burn only when income < other's	**6**	**2**	**4**
(*Disadvantageous inequality aversion*)	(4.87%)	(7.40%)	(4.17%)
Burn only when income ≥ other's	**19**	**4**	**15**
(*Pure nastiness*)	(15.45%)	(14.81%)	(15.62%)
Always burn	**2**	**0**	**2**
(*Unconditionally anti-social*)	(1.63%)	(0%)	(2.08%)
Other	**14**	**3**	**11**
	(11.38%)	(11.12%)	(11.46%)
Total # Subjects	***123***	***27***	***96***

Notes: # subjects in **bold**, % subjects in parenthesis ( )

Table 6. Probability of burning money

Marginal Effect (st. error) displayed **Independent Variable**	(1) Marg. Effect (st. error)	(2) Marg. Effect (st. error)	(3) Marg. Effect (st. error)
Increasing (*x*,*y*) order (=1)	-0.0077 (0.0502)	-0.0047 (0.0498)	-0.0196 (0.0510)
Decreasing (*x*,*y*) order (=1)	-0.0186 (0.0467)	-0.0173 (0.0460)	-0.0147 (0.0493)
Unilateral Burn (=1)	-0.0146 (0.0456)	-0.0110 (0.0454)	-0.0156 (0.0440)
|Income < other (*x* < *y*)|	0.0005*** (0.0001)	0.0005*** (0.0003)	0.0005*** (.0002)
Income > other (*x* > *y*)	-0.0002 (0.0003)	-0.0002 (0.0003)	-0.0002 (0.0003)
Income = other (= 1)	0.0525 (0.0524)	0.0513 (0.0519)	0.0597 (0.0583)
Relative cost	-0.9551* (0.5447)	-0.9398* (0.5370)	-1.0741* (0.5787)
Trolley Opt-Out (=1)	---	-0.1171* (0.0409)	---
Action Propensity (Trolley dilemmas 2-10)	---	---	0.0675 (0.0546)
Immoral Commission (=1) (action in Trolley 1&11)	---	---	0.2655*** (0.1154)
Immoral Omission (=1) (inaction in Trolley 12)	---	---	0.3428*** (0.1282)
Male (=1)	---	---	-0.0157 (0.0433)
Happiness ∈ [1 ,10] (10=highest current life happiness)	---	---	-0.0278* (0.0161)
Religion ∈ [1 ,10] (10=very important)	---	---	0.0053 (0.0078)
Age	---	---	-0.0135 (0.0133)
Observations	1107	1107	1026
# Participants^	123	123	114^
Log likelihood	-458.2919	-452.741	-406.183

Notes: *.10, **.05, ***.001 for the 2-tailed test. Standard errors clustered at the individual subject level. *Increasing, Decreasing, Random* (reference group) control for the differing order of the money burning allocation scenarios. *Relative Cost* takes on the value of the 10 experimental monetary units (EMU) cost divided by the payoff in EMU if choosing not to burn money. *Trolley Opt-Out* is an indicator for those subjects who chose *not* to complete the Trolley dilemma task.

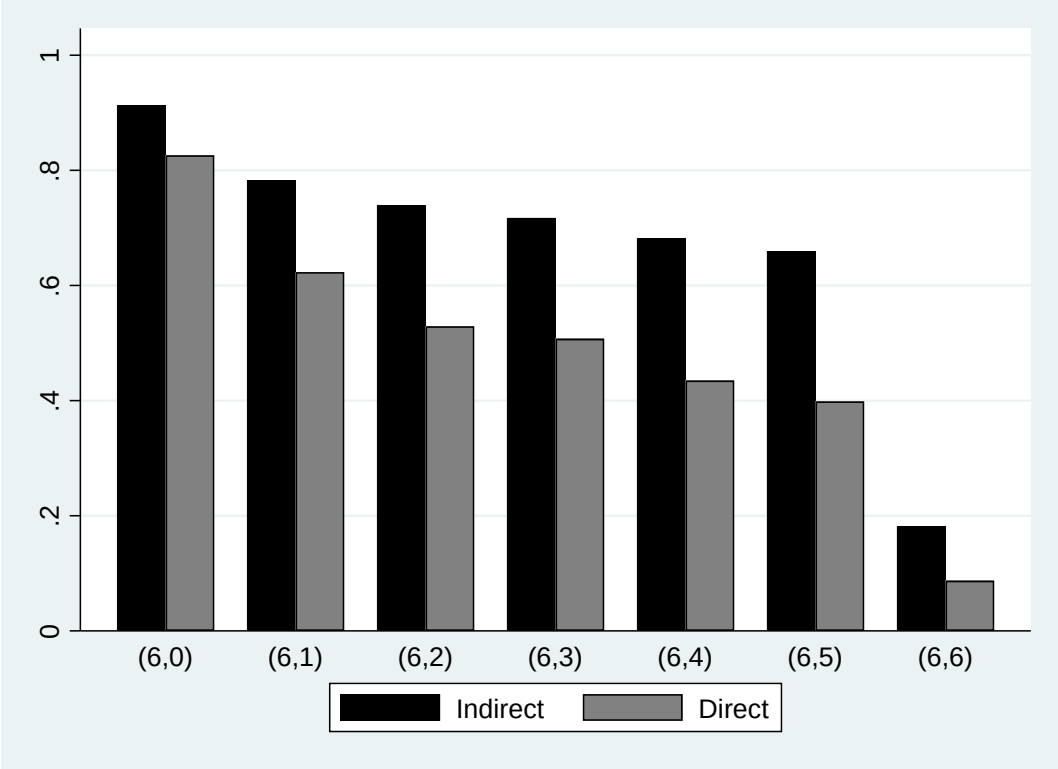
^reduced as a result of those opting out of the Trolley dilemma choice, which is used to score morality variables.

Table 7. Probability of burning money

Marginal Effect (st. error) displayed		
	(1)	(2)
Independent Variable	Income ≤ other's	Income ≥ other's
Increasing (*x*,*y*) order (=1)	0.0546	-0.0665
	(0.0705)	(0.0574)
Decreasing (*x*,*y*) order (=1)	0.0312	-0.0510
	(0.0649)	(0.0597)
Unilateral Burn (=1)	-0.0326	-0.0123
	(0.0527)	(0.0569)
|Diff Income|	0.0004***	0.0007
	(0.0001)	(0.0005)
Equal Income (*x* = *y*)	0.0513	0.2075
	(0.0509)	(0.1820)
Relative cost	---	-2.3937*
		(1.3978)
Action Propensity	0.0379	0.1121
(Trolley dilemmas 2-10)	(0.0572)	(0.0717)
Immoral Commission (=1)	0.3669***	0.1326
(action in Trolley 1&11)	(0.1797)	(0.1146)
Immoral Omission (=1)	0.3823***	0.3043***
(inaction in Trolley 12)	(0.1774)	(0.1387)
Male (=1)	0.0084	-0.0441
	(0.0527)	(0.0541)
Happiness ∈ [1 ,10]	-0.0203	-0.0357*
(10=highest current life happiness)	(0.0172)	(0.0195)
Religion ∈ [1 ,10]	0.0085	-0.0071
(10=very important)	(0.0092)	(0.0102)
Age	-0.0211	-0.0105
	(0.0151)	(0.0176)
Observations	570	570
# Participants^	114	114
Log likelihood	-196.646	-239.281

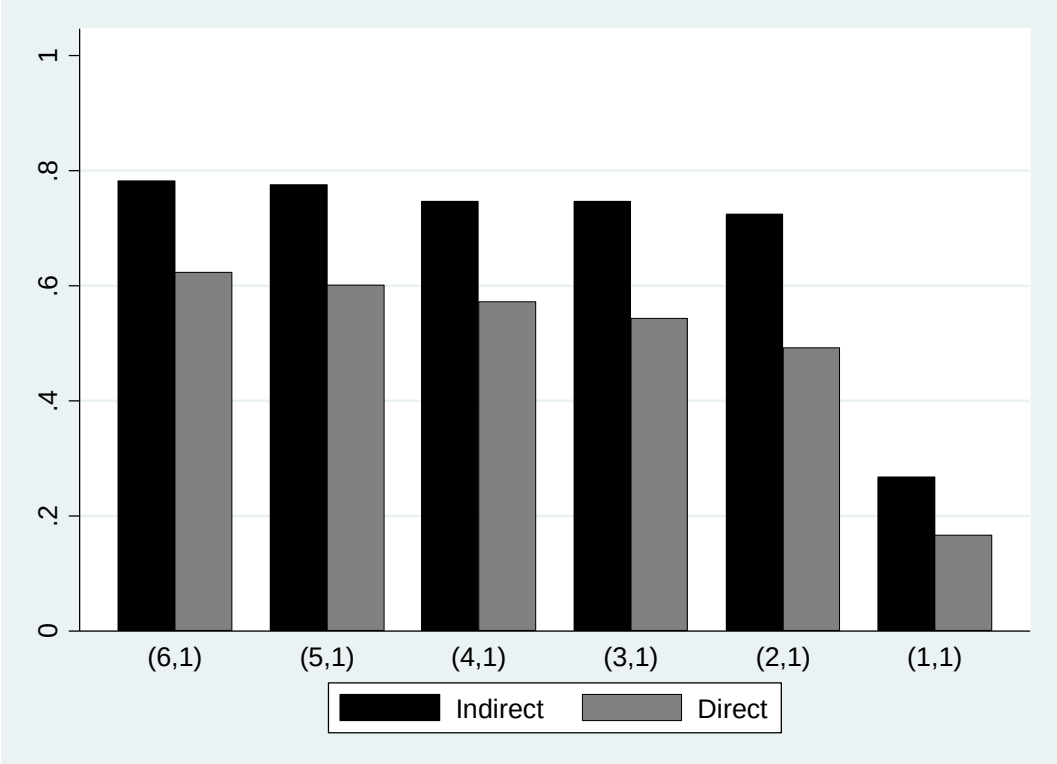
Notes: *.10, **.05, ***.001 for the 2-tailed test. Standard errors clustered at the individual subject level.
^reduced as a result of those opting out of the Trolley dilemma choice, which is used to score morality variables.

Figure 1. Frequencies of taking action in the Trolley dilemma by treatment (number saved unchanged)



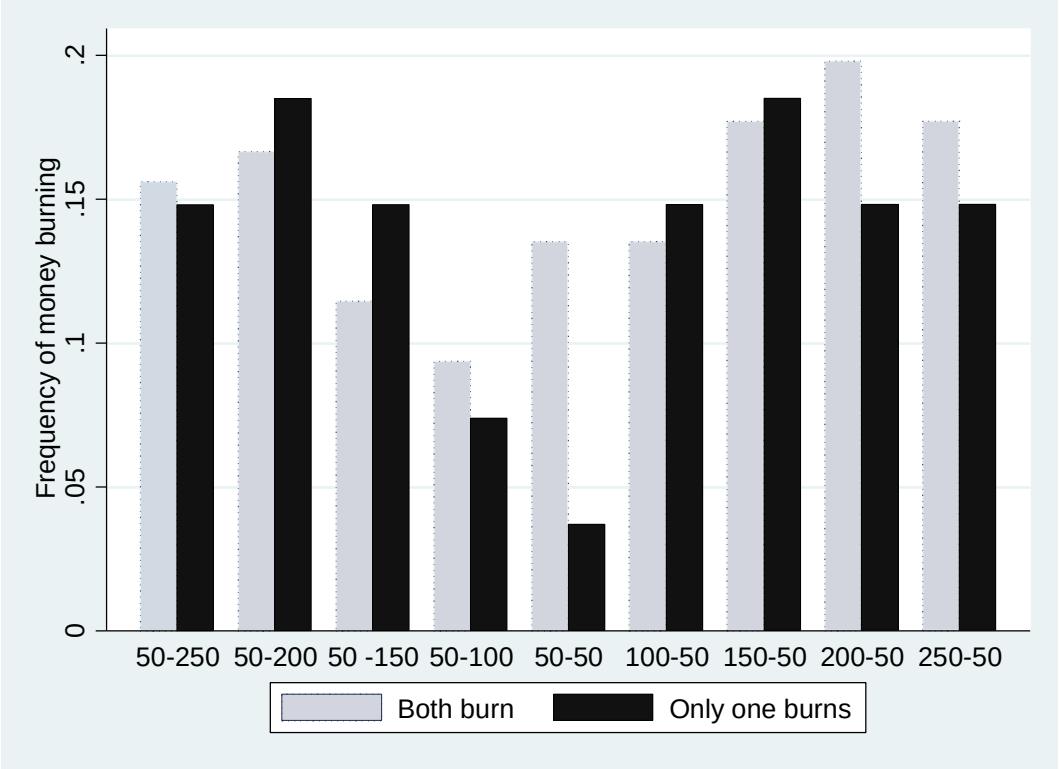
Notes: (X,Y) dilemmas represent number saved (X) and number sacrificed (Y)

Figure 2. Frequencies of taking action in the Trolley dilemma by treatment (number killed unchanged)



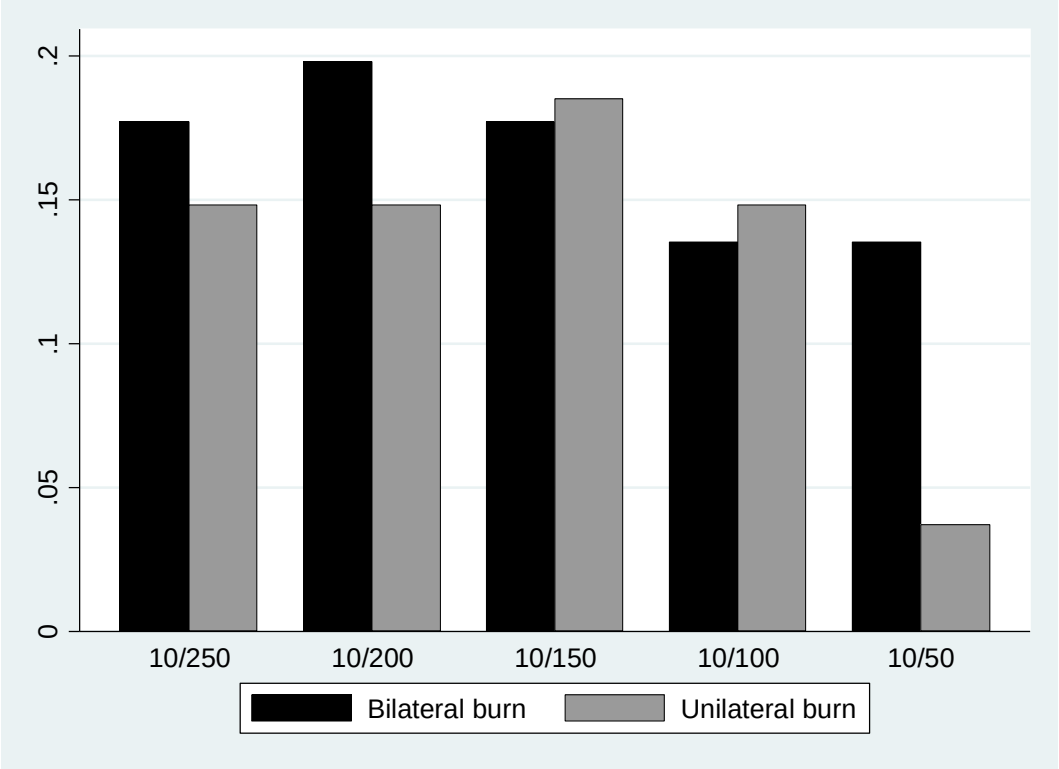
Notes: (X,Y) dilemmas represent number saved (X) and number sacrificed (Y)

Figure 3. Frequencies of money burning decision by treatment



Notes: Allocation *x-y* describes own payoff-recipient payoff

Figure 4. Money burning per relative cost



Notes: ratio along horizontal axis reflect the size of the burning cost, 10, relative to the decider's payoff level prior to burning the recipient's payoff. Left to right along the axis reflects an increasing cost of burning the recipient's payoff, relative to one's own payoff in the allocation.

Appendix A: Experiment Instructions