Altruism begets altruism*

Stephanie A. Heger†
School of Economics, The University of Sydney

Robert Slonim
School of Economics, The University of Sydney

July 11, 2020

Abstract: Guided by Bem's (1972) self-perception theory, we design an experiment to ask whether morally-motivated behaviour, e.g., charitable giving, is history-dependent. Using a popular policy nudge, the default option, we exogenously vary altruism "now" and show that the nudge-induced choice to give "now" causes a 41 percentage point (or 200%) increase in the probability of giving "later"; that is, altruism begets altruism. We further show that, consistent with self-perception theory, the *choice* to behave altruistically "now", rather than the nudge itself, is the crucial element in the causal relationship. These findings are consistent with positive path-dependence, which we interpret as moral consistency.

*This project benefitted from many helpful comments and conversations, in particular, we thank Christine Exley, Lata Gangadharan, Tom Wilkening, Nina Xue and seminar and conference participants at the University of Sydney, the University of Melbourne, the Early Career Behavioral Economics Conference (ECBE 2019), and the Science of Philanthropy Initiative (SPI). Slonim gratefully acknowledges funding from ARC DP 150101307.

†[Corresponding author] Email: saheger@gmail.com.

1 Introduction

Policy shapes society by encouraging socially desirable behavior. For example, the United States government successfully incentivizes charitable giving by allowing individuals to deduct donations from their pre-taxed income (Clotfelter, 1980; Meer and Priday, 2019). This policy can also have additional and often unaccounted for consequences: decreases in the after-tax price of giving increase charitable giving and are also associated with increases in other socially desirable behaviors, such as volunteerism (Feldman, 2010).

Charitable giving not only made up 2% of GDP in the USA in 2019 (see Giving USA, 2020), but it is also associated with increased happiness (Anik et al., 2009) and health (Yörük, 2014). Thus, it is not surprising that there is a large body of research that examines interventions aimed at increasing charitable giving, including the effects of price (Karlan and List, 2007), various fundraising schemes (Huck, Rasul, and Shephard, 2015), efficiency concerns (Gneezy, Keenan, and Gneezy, 2014; Exley, 2015b), social pressure (List and Lucking-Reiley, 2002; Frey and Meier, 2004; Shang and Croson, 2009), and identity (Kessler and Milkman, 2016). And in addition to examining the immediate effects of interventions, there is a growing literature looking at the intertemporal spill-overs of pro-social behaviours generated by interventions at charitable giving(Shang and Croson, 2009; Cairns and Slonim, 2011; Gneezy et al., 2012; Castillo, Petrie, and Samek, 2017). For example, Shang and Croson (2009) find that subjects who experienced social pressure to give to public radio in $t-1$ were more likely to give in $t-1$ but also more likely to renew their membership in time t and donate more in time t relative to those subjects who did not experience social pressure at time $t-1$. Gneezy et al. (2012) show that when subjects are randomly assigned to incur the cost of a charitable donation at $t-1$, they will be more likely to behave honestly in time t, relative to those randomly assigned to make a costless donation at time $t-1$. [1] In other words, the existing literature has examined how an intervention at time $t-1$ affects behavior in time $t-1$ and behaviour in time t.

In this paper, by contrast, we estimate how *choices* at $t-1$ affect *choices* at t. We hypothesize and show that altruism begets altruism;[2] a nudge induces people to give now,

[1] Cassar, d'Adda, and Grosjean (2014); Peysakhovich and Rand (2016); Engl, Riedl, and Weber (2018) also study positive spill-overs in prosociality, but not due to an increase in charitable giving. For example, Peysakhovich and Rand (2016) examine how random exposure to incentive structures for cooperation in infinitely repeated prisoner's dilemma games affect future cooperation in one-shot coordination games, while Cassar, d'Adda, and Grosjean (2014) study the effect of institutional strength in a market game on subjects' behavior in a trust game.

[2] Our hypothesis is supported by the review in Gee and Meer (2019), who conclude that while there is some evidence of donor fatigue (Meier, 2007; Damgaard and Gravert, 2018), "the preponderance of evidence finds that gifts today do not cannibalize gifts tomorrow."

and this increase in *choosing to give* now causes an increase in later giving.[3] We do this by instrumenting for choices at $t-1$ with the random assignment to a default option nudge that encourages people to give at time $t-1$ and then we also elicit giving decisions at time t. Our experimental and econometric approach is thus distinct from previous literature that shows a positive correlation in giving over time (Landry et al., 2010; Adena and Huck, 2019);[4] we provide causal evidence of the effect of altruistic choices *now* on altruism *later*.

Our hypothesis comes from self-perception theory (Bem, 1972), which provides a framework for modeling morally-motivated behaviors over time. Self-perception theory posits that individuals use past choices to make inferences about their own identity, which then inform future choices. Bénabou and Tirole (2011) formalize self-perception theory in economics; they model individuals with imperfect memories of their identity who use their past choices to make inferences about their identity, which serves as a guide for current choices. Further, self-perception theory emphasizes the role of *choice* in linking behaviors to identity (Zanna, 1972).[5] Specifically, if an individual is forced to donate then they cannot infer much about their identity as an altruistic person, whereas if they had made an active choice to donate then they can make an inference about their identity as an altruistic person from the past donation. Thus, self-perception theory predicts a path-dependency between *chosen* moral actions over time.[6]

Motivated by the history-dependence in chosen actions modelled by Bem (1972) and Bénabou and Tirole (2011), we model our decision-maker's utility at time t as dependent on his current choice of charitable giving and his $t-1$ choice of altruism using a habit formation model (Pollak, 1970). Further, his $t-1$ choice depends on his previous choices of altruism as well as on whether he is nudged towards altruism or selfishness at $t-1$. We thus model moral consistency in altruism as habit persistent charitable giving, meaning that charitable giving is not just positively correlated over time, but that an increase in past giving *causes* an increase in giving today. This is also consistent with Meer (2013), who finds evidence consistent with habit formation preferences for giving—using the performance of college athletic teams as an instrument for giving when "young", Meer (2013) shows that giving when young causes an increase in giving 20 years after graduation.

Self-perception theory serves as the guiding light to design an experiment capable of

[3] See Thaler and Sunstein (2003) and Sunstein and Thaler (2008) for a review of nudges.

[4] Cappelen et al. (2017) use a similar approach in a different context and find that incentivizing subjects to go to the gym increases the likelihood of exercise, which in turn, increases the subjects' academic performance.

[5]

[6] Thus, we also contribute to a relatively recent literature that examines the role of agency in charitable giving (Eckel, Herberich, and Meer, 2018; Kessler, Milkman, and Zhang, 2019), which finds that an increased sense of agency does not affect donation behaviour at the extensive margin, but does have a positive and significant effect on the intensive margin.

answering whether altruism begets altruism. We conduct an online experiment in which we nudged individuals to either donate to charity or to keep the cash for themselves by setting their default option to "donate" (henceforth: Default Charity) or to "keep" (henceforth: Default Cash), respectively. To avoid donating, subjects in the Default Charity condition must opt-out of giving to charity; by contrast, subjects in the Default Cash condition must opt-in to giving to charity and opt-out of keeping cash (Round 1). We chose a default option nudge for two reasons: (1) to generate exogenous variation in giving "now"; and (2) to maintain the active choice in Round 1 behavior. Consistent with past research on default option nudges (Benartzi and Thaler, 2007; Choi et al., 2003; Cronqvist and Thaler, 2004; Madrian and Shea, 2001; Kessler and Roth, 2012, 2014), we find that our nudge positively impacts charitable giving behavior .[7] Specifically, we find that subjects in the Default Charity condition are twice as likely to choose to donate in Round 1 than subjects in the Default Cash condition.

The critical part of the design is that at a later point in the experiment, "later", we ask subjects to make another donation to a different charity to test whether initial altruistic behavior increases altruism in the future (Round 2). Directly motivated by our model and experimental design, we estimate a local average treatment effect using our randomly assigned nudge as an instrument for Round 1 giving and find that the nudge-induced increase in choosing to give in Round 1 causes giving in Round 2 to increase by 200% or 40 percentage points. Overall, our experiment shows that the nudge-induced choice to be altruistic in Round 1 begets more altruism in Round 2, thus generating a virtuous cycle of altruism.

To provide more direct support for the validity of our identification strategy and to demonstrate the central causal role that *choice* in round one will have on the later donation choices, we ran additional treatments in which we randomly assigned subjects to a default position in Round 1, but do not give them the choice to opt-out of their default position; that is, they are forced to make a donation or are forced to keep the money in Round 1 (henceforth: No Choice Treatments).

The results from the additional treatments support our main finding. In particular, participants in Default Charity (No Choice) and Default Cash (No Choice) donate at equal rates in Round 2, emphasizing the criticality of *choice* in Round 1 in linking behavior inter-temporally.[8] In fact, we run three sets of No Choice Treatments, described in detail in

[7]Conceptually, setting a default option works by decreasing the marginal psychological costs of choosing the desired behavior.

[8]Gneezy et al. (2012) reports results from an experiment in which subjects who are randomly assigned to make a costly donation are more likely to behave honestly in a subsequent period than subjects who are randomly assigned to make a costless donation. Importantly, particularly in relation to our study, subjects in both the costly and costless treatment were forced to donate rather than having to choose whether to

Section 2.1.2, in which we vary the degree to which subjects reflect on the choice of donating or keeping money for themselves in Round 1 to ensure that the key aspect driving the path-dependence in the Choice Treatments is, indeed, the active choice in Round 1.

Our contribution is thus twofold. First, we provide evidence in favor of moral consistency; that is, we show that choosing to donate "now" causes an increase in altruism "later". To claim this causal relationship, we show that the exclusion restriction assumption holds and that the nudge itself is not responsible for the inter-temporal effect, but rather the choice to act altruistically that the nudge induces *now* causes the increase *later*. Therein lies our second contribution—we use experimental treatments to directly test that the theoretical assumptions behind our empirical test hold.

Last, we explore heterogeneous effects of moral consistency. Motivated by Bem (1972)'s self-perception theory and the model of Bénabou and Tirole (2011), we explore the role of identity and moral consistency, contributing to the growing literature on the role of identity in economics (Akerlof and Kranton, 2005) and, more specifically, in charitable giving (Benjamin, Choi, and Fisher, 2010; Kessler and Milkman, 2016). Consistent with both our model and Benabou and Tirole's (2011) model,[9] individuals whose past limited donations suggest that altruism is a weakly held facet of their identity behave in a significantly more morally consistent manner. For these individuals, choosing to behave altruistically in Round 1 causes an 83 percentage point (or 492%) increase in altruism in Round 2. Self-perception theory suggests that the altruistic behavior induced by the nudge is more informative for weak altruists than for strong altruists because weak altruists have a much sparser history that will make the Round 1 *choice* more salient.

Our results have several important implications for both public policy and corporate strategy. For example, tax policies that provide subsidies for individuals who give to charity may increase the direct amount of charitable giving (Gruber, 2004; Yörük, 2013), but also have the added benefit of increasing individuals' altruistic identities and thus leading to additional altruism. The moral consistency in giving is good news for philanthropic organizations who engage in fundraising. Of the 8775 non-profits reviewed on Charity Navigator,

behave altruistically. Thus, while a direct impact through salience is possible, Gneezy et al. (2012) prevents an indirect channel predicted by self-perception theory that we will explore here.

[9]In Benabou and Tirole's (2011) model, individuals for whom altruism is a weakly held facet of their identity are predicted to behave in a more morally consistent manner. On the other hand, their model also predicts that challenges to strongly-held aspects of identity "today" are met with contradictory responses "tomorrow". Thus, depending on whether an individual has a weakly or strongly held conviction towards altruism, Bénabou and Tirole (2011) predicts either moral consistency (Nisan, 1985; Nisan and Horenczyk, 1990) or moral licensing (Khan and Dhar, 2006; Monin and Miller, 2001; Ploner and Regner, 2013; Sachdeva, Iliev, and Medin, 2009) (also see Blanken, van de Ven, and Zeelenberg (2015) and Mullen and Monin (2016) for a review of this literature in psychology).

the average percentage of expenses spent on fundraising is 9% (or approximately 1.4 million USD).[10] Our results are also relevant to those firms who seek to establish more pro-social corporate cultures—promoting altruism among employees in one domain may initiate a habit of altruism due to moral consistency. This habit persistence may then have a positive effect on corporate culture and the firm's profitability through a variety of channels: increasing engagement in corporate social responsibility strategies, decreasing the costs associated with promoting social interactions among employees (Dur and Sol, 2010), and decreasing the costs associated with monitoring worker productivity (Sarkisian, 2017).

2 Design & Procedures

In this section, we describe our experimental design and the data generated by the experiment. We also present a model of consumption choice and motivate our hypotheses in Section 2.3.

2.1 Experimental Design

2.1.1 Calibrating Preferences

We ran a pre-experimental calibration exercise to gauge the amount that must be donated to the chosen charity for the average subject to be indifferent to giving up \$1. The calibration exercise is important to set the default options such that some subjects will prefer to donate, while other subjects will prefer to keep cash for themselves. By finding a the median point of indifference between donating to charity and keeping cash for self, we can be confident that the nudge towards charity or the nudge towards keeping cash will be on the appropriate margins.

To do the calibration, we used the same charity, CARE, that we will use in the Round 1 decision of the experiment. This exercise follows the calibration exercise in Exley (2015a) and presents subjects with a multiple price list. On each line, they are asked whether they prefer to keep a \$1 and give \$0 to the charity or keep \$0 and give \$x to the charity, where x$\in$ $\{\$0, \$0.1, ...\$3\}$. While Exley (2015a) uses a within-subject calibration, our calibration is taken as the median point of indifference across subjects, which was \$1 to self was utility-equivalent to \$1.50 to charity. This is how we chose the values in Round 1: subjects in the Default Cash condition were endowed with \$1 to keep for themselves and subjects in the Default Charity condition were endowed with making a \$1.50 donation to the charity. Sub-

[10] See Charity Navigator

jects in this calibration exercise were excluded from participating in any of the experimental conditions that follow.

2.1.2 Main Treatments

The main experiment consists of two Rounds. In Round 1, subjects were randomly endowed with $1 cash (Default Cash condition) or endowed with a $1.50 donation to the charity CARE (Default Charity condition). Figures A1a and A1b display what the subjects saw if they were assigned to the Default Cash and Default Charity treatments, respectively. After providing their endowment, we took two additional steps to facilitate a sense of ownership among subjects of their default position. First, we asked subjects in the Default Charity condition to list three ways the charity CARE might spend this money and we asked subjects in the Default Cash condition to list three ways they might spend their cash endowment. Second, we asked subjects to complete a set of unrelated filler questions. These filler questions created a period over which the subject had ownership of their default position. Having subjects write about their endowment is a common technique in the psychology literature to increase the sense of ownership (Shu and Peck, 2011) and elongating the time of having ownership of one's endowment has been shown to increase the endowment effect (Strahilevitz and Loewenstein, 1998). Moreover, while completing the filler questions,[11] we reminded subjects of their default position by showing an image of their endowment to further reinforce the ownership of the default option they were given.

After completing the filler tasks, we asked subjects whether they would like to swap their position. Subjects assigned to the Default Cash treatment were asked if they wanted to give back their $1 to make a $1.50 donation to CARE while subjects assigned to the default donation treatment were asked if they wanted to not make the $1.50 donation to get $1 in cash. Figures A1c and A1d display the decisions faced by the subjects from the Default Cash and Default Charity treatments, respectively. When subjects made their Round 1 choice, they were unaware that there would be a Round 2 choice and we expect that their choices in Round 1 may have differed if they anticipated a Round 2 donation solicitation (Adena and Huck, 2019).

Next, in Round 2, we presented subjects with a multiple price list in which they had to choose one of 11 options. For each item, they could choose to add $X= $(0, 0.10, 0.20...1.00)$ to their bonus and donate $2×(1-X) to Save the Children (see Table A1). For example, in the first option, subjects could choose to add $1 to their bonus and donate $0 to Save the Children, while in the last option, subjects could choose to add $0 to their bonus and donate

[11] Please see the full experimental protocol here to see the filler tasks the subjects performed.

$2 to Save the Children. Subjects had to select one option from the list.We chose a new charity for the Round 2 decision to avoid a potential charity-specific wealth effect; that is, if some subjects donated to CARE in Round 1 (and others did not), then the marginal utilities of donating to CARE in Round 2 could differ by treatment assignment.

After completing the two rounds of decisions, we asked a brief series of demographic questions as well as questions about their past charitable giving behavior. We summarize and discuss these statistics below in Table 1.

2.1.3 No Choice Treatments

In addition to the two main choice treatments, we included three No Choice treatments. In the No Choice Treatments, subjects are assigned to either the Default Charity or Default Cash conditions, and follow the exact procedures as in the Choice Treatments, including ownership questions and filler tasks, and are then asked to make the same Round 2 donation decision as the Choice treatments. The only difference between the No Choice treatments and the Choice treatments is that subjects in the No Choice treatments are not given the opportunity to switch their Round 1 default position. Thus, subjects in the Default Charity (No Choice) treatment are forced to make a donation in Round 1, while subjects in the Default Cash (No Choice) treatment are forced to keep the cash in Round 1.

For robustness and to isolate the importance of *choice* in driving spill-overs, we ran three variations of the No Choice treatments: (1) **Hypothetical Choice** in which they are asked to indicate what their Round 1 Choice would have been if given the opportunity; (2) **Hypothetical Scenario** in which they are asked to reflect upon, but not indicate, what their Round 1 choice would have been if given the opportunity; and (3) **No Information** in which they are not informed of any alternative to their assigned default position. The Hypothetical Choice treatment is identical to our Main treatment *except* for the choice is not executed and is thus not an active choice.

Self-perception theory argues that an active choice is an important component linking behavior and identity, noting that rejected alternatives reinforce the inferences an individual can make from their choice about their identity (Bem, 1972; Zanna, 1972). For example, an individual learns less about his altruistic identity if he is forced to donate rather than having had an active choice to keep the money for himself.

In a final control treatment (henceforth: Round 2 Only), subjects do not make a Round 1 decision and instead begin the experiment with the filler tasks and are then asked to make a Round 2 donation decision that is identical to the other treatments. The Round 2 Only treatment allows us to examine how subjects make Round 2 donation choices when they are

only asked to give once.

2.2 Data

The data from our experiment come from Amazon's Mechanical Turk (Mturk) and was completed by 1801 Mturk workers from the United States who have HIT approval rate greater than 99% and have had more than 10000 HITs approved. Table 1 displays the summary statistics for the subjects in our experiment. Approximately half of the subjects are female, the majority work full-time and 77% have donated to charity at least once in the last year.

To proxy for each subject's identity towards altruism, we asked whether they had donated money to a charity 0, 1, 2, 3 or 4 or more times in the past year (not including the donation made during the experiment). Using this variable, we classify subjects as having a strong conviction towards altruism if they indicated that they have given 4 or more times in the past year. Otherwise, a subject is classified as having a weak identity towards altruism.

Table 1: Summary Statistics

		Treatment Conditions				
	All	Default Charity (Choice)	Default Cash (Choice)	Default Charity (No Choice)	Default Cash (No Choice)	Round 2 Only
Altruism Strongly Held Value Past donations ≥ 4	.32 (.47)	.38 (.49)	.31 (.46)	.26 (.44)	.34 (.47)	.38 (.49)
Female	.52 (.50)	.52 (.50)	.54 (.50)	.51 (.50)	.52 (.50)	.54 (.50)
Age	37.91 (12.52)	40.51 (11.82)	38.38 (11.56)	37.46 (14.31)	37.22 (11.63)	38.36 (11.24)
Unemployed	.07 (.26)	.05 (.22)	.09 (.29)	.09 (.28)	.06 (.24)	.08 (.26)
Employed full-time	.63 (.48)	.64 (.48)	.58 (.49)	.64 (.48)	.64 (.48)	.59 (.49)
Employed part-time	.16 (.36)	.14 (.35)	.18 (.38)	.14 (.35)	.17 (.35)	.17 (.37)
Retired	.03 (.18)	.05 (.21)	.04 (.20)	.03 (.18)	.03 (.16)	.02 (.14)
Income < \$10,000	.05 (.21)	.07 (.25)	.05 (.23)	.05 (.21)	.04 (.19)	.05 (.22)
Income > \$150,000	.04 (.19)	.04 (.19)	.02 (.15)	.04 (.19)	.04 (.19)	.05 (.21)
Observations	1802	191	224	558	630	199

Means reported with standard deviations in parentheses.

2.3 Model, Hypotheses and Empirical Strategy

Next, we turn to modeling the choice to donate at t, given previous donation choices, to examine whether altruism beget altruism. To formalize this question, we consider an individual who has preferences over two goods at time t, private consumption (c_t) and charitable

giving (A_t). The individual's preferences can be represented by a utility function with the following form,

$$U(c, A) = u(c_t, c_{t-1}(\Theta_c)) + \alpha v(A_t, A_{t-1}(\Theta_A)) \tag{1}$$

where Θ_c and Θ_A represent a composite of private consumption and charitable giving up to and including time $t-2$, respectively. Thus, today's utility depends on the choices the individual makes today as well as all past choices. The parameter $\alpha \in [0, 1]$ governs the intensity of the individual's preference for altruism and warm glow. The functions $u(\cdot)$ and $v(\cdot)$ are concave in consumption and donations to charity, respectively. At time t, a subject chooses (c_t, A_t) taking their previous choices, (c_{t-1}, A_{t-1}), as given.

$$\max_{c_t, A_t} U(c_t, A_t \mid \bar{c}, \bar{A}) = \max_{c_t, A_t} u(c_t - \gamma_c c_{t-1}(\Theta_c)) + \alpha v(A_t - \gamma_A A_{t-1}(\Theta_A)) \text{ subject to } I = c_t + p \times A_t \tag{2}$$

where the parameter γ_c and $\gamma_A \in \mathbb{R}$ represent the intensity of the past consumption choices $(c_{t-1}(\Theta_c),\ A_{t-1}(\Theta_A))$ on today's utility and will pin down whether there are negative, positive or no path-dependence. I is income and p is the relative price of making a donation. We want to compare the optimal choices at time t of individuals nudged towards altruism versus subjects nudged towards selfishness at time $t-1$. Let $A_t(Z)$ and $A_{t-1}(\Theta_A, Z)$ represent the choices at t and $t-1$, respectively, for an individual who receives nudge $Z \in 0, 1$, where $Z = 1$ indicates the subject was nudged towards altruism and $Z = 0$ indicates the individual was nudged towards selfishness. From the first order conditions we find that

$$A_t(Z = 0) - \gamma_A A_{t-1}(\Theta_A, Z = 0) = A_t(Z = 1) - \gamma_A A_{t-1}(\Theta_A, Z = 1) \tag{3}$$

We assume that $\frac{\partial A_{t-1}}{\partial \Theta_A}|_{Z=1} \geq \frac{\partial A_{t-1}}{\partial \Theta_A}|_{Z=0}$. Rearranging and taking expectations of equation 3, we obtain

$$\frac{E\left[A_t \mid Z = 1\right] - E\left[A_t \mid Z = 0\right]}{E\left[A_{t-1} \mid \Theta_A, Z = 1\right] - E\left[A_{t-1} \mid \Theta_A, Z = 0\right]} = \gamma_A \tag{4}$$

The left-hand-side of equation 4 is the equation for an instrumental variable estimand, β^{IV}. Thus, we propose to test for positive or negative spillovers by estimating the local average treatment effect (Imbens and Angrist, 1994)using instrumental variables (Angrist, Imbens, and Rubin, 1996).

Our identification strategy relies on four assumptions. First, the instrument, Z, is ran-

domly assigned. We satisfy this assumption in our experimental design. Second, the effect of the instrument, Z, must be monotonic in that a subject in the Default Charity condition must be at least as likely to donate in Round 1 than he would have been had he been assigned to the Default Cash condition. The monotonicity assumption is related to the denominator of equation 4, which is the first stage of our IV estimate. Further, there is ample evidence that default option nudges are effective (Benartzi and Thaler, 2007; Choi et al., 2003; Cronqvist and Thaler, 2004; Madrian and Shea, 2001; Kessler and Roth, 2012, 2014). Thus, we hypothesize that, on average, subjects in the Default Charity condition will be more likely to donate in Round 1 then subjects in the Default Cash condition; that is, $E\left[A_{t-1} \mid \Theta_A, Z=1\right] - E\left[A_{t-1} \mid \Theta_A, Z=0\right] > 0$. We test and provide support for this hypothesis in Section 3.1.

Hypothesis 1. ***Default Option Hypothesis****: Participants who are defaulted into making a donation are more likely to donate in Round 1 than participants who are defaulted into keeping cash.*

Third, given the large experimental literature that find *positive* spill overs in pro-social behaviour (Shang and Croson, 2009; Cairns and Slonim, 2011; Gneezy et al., 2012; Cassar, d'Adda, and Grosjean, 2014; Peysakhovich and Rand, 2016; Castillo, Petrie, and Samek, 2017; Engl, Riedl, and Weber, 2018), we hypothesize that subjects in the Default Charity condition will *also* be more likely to donate in Round 2 than subjects in the Default Cash condition; that is, $E\left[A_t \mid Z=1\right] - E\left[A_t \mid Z=0\right] > 0$

Hypothesis 2. ***Positive Spill Over Hypothesis****: Subjects in the Default Charity treatment, $Z=1$, will be more likely to donate in Round 2 than subjects in the Default Cash condition, $Z=0$.*

Fourth, the exclusion restriction states that the instrument only affects outcome A_t through A_{t-1} (i.e., Round 1 donation behavior) and does not directly affect outcomes, A_t. Consistent with self-perception theory, this requires that it is the *choice* of donating in Round 1, which is influenced by the subject's default position (i.e., the instrument, Z), rather than the default position itself that affects Round 2 donation decisions (i.e., outcomes, A_t). This assumption is the hardest to justify without some evidence and is the motivation for our No Choice treatments described in Section 2.1.3. To demonstrate that *choice*, rather than treatment assignment, is the driving force behind Round 2 behaviour, we hypothesize and show that: (1) $E\left[A_t \mid Z=1, NoChoice\right] - E\left[A_t \mid Z=0, NoChoice\right] = 0$ and (2) $E\left[A_t \mid Z=1\right] - E\left[A_t \mid Z=0\right] >E[A_t \mid Z=1, NoChoice] - E\left[A_t \mid Z=0, NoChoice\right] = 0$. We go one step further and We test and provide support for this hypothesis in Section 3.2

Hypothesis 3. ***Exclusion Restriction Hypothesis****: The default option treatment Z does not directly affect the decision to donate in Round 2. Instead, any effect of Z on Round 2 donation choices operates solely through the choice to donate in Round 1.*

Finally, we turn to the main hypothesis about the sign of γ_A. As a direct consequence of hypotheses 1 and 2, we hypothesize that $\gamma_A > 0$, implying positive path-dependence. We interpret a positive path-dependence as moral consistency since $\gamma_A > 0$ implies that the Default Charity condition exogenously increases altruism in Round 1 and that this nudge-induced increase in altruism in Round 1 causes an increase in altruism in Round 2.[12]

Hypothesis 4. ***Moral Consistency Hypothesis****: $\beta^{IV} > 0$, implying that $\gamma_A > 0$ which means that an increased propensity to choose to donate in Round 1 will:*
(i) increase the propensity to donate in Round 2;
(ii) increase the amount donated in Round 2.

To analyze the experimental data, we estimate a two-stage least squares instrumental variable regression in which standard errors are adjusted to account for the two-stage estimation procedure.[13] In the first stage, we estimate the effect of the treatment assignment, Z_i, on Round 1 donation behavior, $A_{i,t-1}$.

$$A_{i,t-1} = \rho_0 + \rho_1 \mathbf{1}\left[Z_i = 1\right] + \rho_2 \mathbf{X} + \nu_i, \tag{5}$$

where $\mathbf{X}$ is a vector of demographic controls, including gender, employment status, and income. In the second stage, we use the predicted values of Round 1 donation behavior, $\widehat{A_{i,t-1}}$, to obtain the causal effect of donating in Round 1 on donating in Round 2, Y_i, and adjust the standard errors to account for the stage-stage estimation procedure. The interpretation of the coefficient, β^{IV}, is the change in Round 2 donation rates that are caused by the *treatment-induced* donation choice in Round 1.

$$A_{i,t} = \beta_0 + \beta^{IV}\widehat{A_{i,t-1}} + \beta_2 \mathbf{X} + \varepsilon_i, \tag{6}$$

We also explore heterogenous effects stemming from identity towards altruism, the details of which are found in Section 3.4 and Appendix B. We continue to draw from self-perception theory by positing that identity is inferred from past choices. Thus, the composite of past charitable giving, Θ_A, serves as a proxy for the facet of identity related to altruism. In our

[12] In Appendix B we consider immoral consistency, that is whether keeping the money in Round 1 causes people to behave more selfishly in Round 2.

[13] Because our model is just-identified, the two stage least square estimator and the limited-information maximum likelihood estimator are equivalent. We thus only report the two-stage least squares model.

model, it is straightforward to show that $\frac{\partial \gamma_A}{\partial \Theta_A} <$, thus moral consistency is *decreasing* in the strength of the individual's altruistic identity (Θ_A).

3 Results

Our results are consistent with self-perception theory and moral consistency; that is, exogenously-induced altruism in one period causes an increase in altruism in a subsequent period. In other words, altruism begets altruism. To establish our main result, we first demonstrate that we have a strong instrument (result 1) and, importantly, that the exclusion restriction holds (result 2), in Sections 3.1 and 3.2, respectively. We then establish our main result, that altruism begets altruism, result 3, in Section 3.3. For each result, we begin with a statement of the result, followed by the evidence to support it.

3.1 Round 1 Decisions

Result 1. *Our nudge significantly affects round 1 donation choices. Subjects who are defaulted into giving to charity are significantly more likely to choose to donate in Round 1 than subjects who are defaulted to keep a utility-equivalent amount of cash.*

In Table 2, we show that the Default Charity treatment increases the propensity to give in Round 1 by 19 percentage points compared to the Default Cash treatment. Importantly, Table 2 also provides evidence of a strong first stage—the F-statistic in our first stage is 18.96, which exceeds the "rule of thumb" for a relevant instrument proposed by Staiger and Stock (1997).[14]

3.2 Round 2 Decisions and the Importance of Choice

The purpose of this section is to provide evidence of the importance of *choice* in Round 1 in driving Round 2 choices. We do this by showing significant differences in Round 2 behavior between the Default Charity and Default Cash for the Choice Treatments, but no differences in Round 2 behavior between the Default Charity and Default Cash in the No Choice treatments. Thus, we argue that the Round 1 *choice*, rather then the treatment assignment, drives the positive path-dependence and can thus be interpreted as moral consistency in Section 3.3.

[14] In column (2), the F-statistic drops due to the inclusion of additional control variables that have little predictive power in the Round 1 donation behaviour.

Table 2: Relevance of Instrument: Effect of Default Treatment on Round 1 Donation Rates

	(1)	(2)
Default Charity	0.19***	0.18***
	(0.04)	(0.04)
Constant	0.17***	0.13**
	(0.02)	(0.06)
Observations	415	415
R^2	0.05	0.05
F statistic	18.96	3.57
Demographic Controls	No	Yes

OLS regression estimates. Demographic controls include gender, employment status (Full-time, part-time, retired or unemployed), and income. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. Given that we have a directional hypothesis for the effect of default charity nudge, the test of the hypothesis on the coefficient for Default Charity are one-tailed.

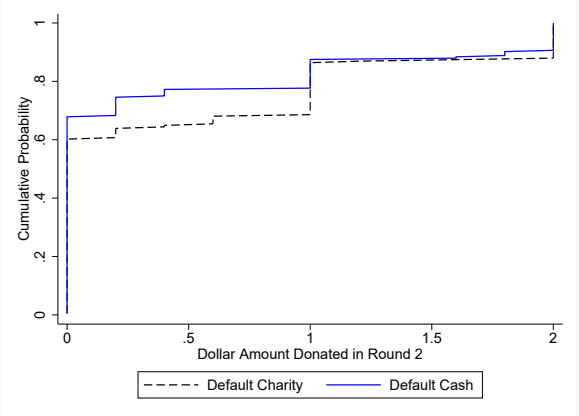
Recall, that subjects in the No Choice treatments were assigned to either a Default Charity or Default Cash option in one of three treatment arms: (1) **Hypothetical Choice**, (2) **Hypothetical Scenario**, and (3) **No Information**. By including these three variations of the No Choice treatments, we are able to carefully rule out several competing explanations. The Hypothetical Choice treatment is nearly identical to the Choice Treatments, except the choice made in Round 1 is not executed. Thus, any difference between the Hypothetical Choice and the Choice Treatments can be attributed to the "active choice" and not experimenter demand effects or salience of charitable giving. Interestingly, we find no significant differences in Round 2 behavior between the Hypothetical Choice, Hypothetical Scenario and No Information treatments and thus for the purpose of this section we pool them together as the No Choice Treatments.

Result 2. *We find that the treatment itself has no direct effect on Round 2 choices; that is, subjects in the Default Charity (No Choice) treatments do not behave significantly different in Round 2 than subjects in the Default Cash (No Choice) treatments.*

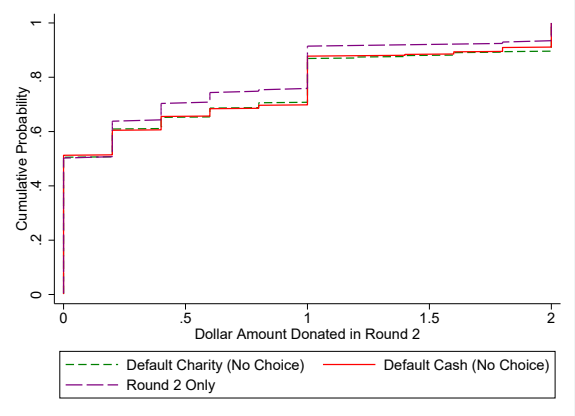
Figure 1 shows the distributions of donation amounts in Round 2 by treatment assignment for the Choice and No Choice treatments. Figure 1a shows that the distribution of donations in Round 2 between the Choice Treatment conditions—Default Cash and Default Charity—appear quite different, particularly at \$0 and \$1. The difference between these distributions is verified with a Kruksall-Wallis test of distributions (One-tailed test: $\chi^2 = 3.01$, p-value=0.04).

On the other hand, Figure 1b shows that the distributions of donations for the No Choice treatments are strikingly similar. The Kruksall-Wallis test allows for a multi-sample

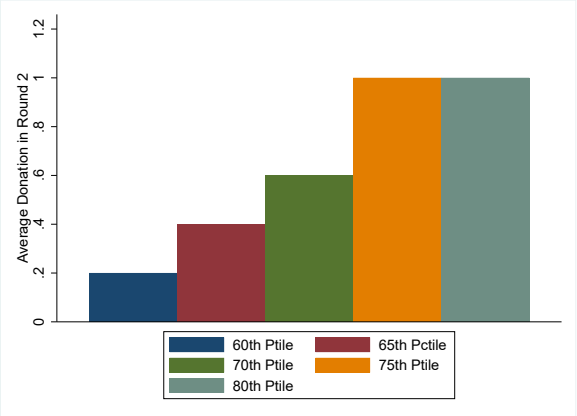
FIGURE 1: DISTRIBUTION OF ROUND 2 DONATION AMOUNTS, BY TREATMENT ASSIGNMENT



(A) Choice Treatments



(B) No Choice Treatments



(C) All Treatments

Distribution of donation amounts in Round 2 by treatment assignment. Figure 1a shows the distribution of donation amounts for the Choice Treatments and Figure 1b shows the distributions for the No Choice treatments. Figure 1c pools all the data from our experiment and shows the average donation amount in Round 2 at various percentiles.

comparison and finds no statistical differences between the distributions of Round 2 donation amounts for the No Choice Treatments (One-tailed test: $\chi^2 = 0.51$, p-value=0.39). In fact, the Round 2 donation behaviour of subjects in the Default Charity and Default Cash No Choice treatments is not statistically different from the Round 2 behaviour of subjects in the Round 2 Only treatment; that is, subjects in the No Choice conditions behave *as if* they never participated in Round 1.

A close investigation of Figure 1a shows that most of the differences in Round 2 donation behaviour between the Choice treatments occurs in the middle of the distribution–a large fraction of subject do not give in Round 2 in both of the Choice treatments (68% in the Default Cash and 60% in the Default Charity) and very few subjects give more than a $1 donation (14% in Default Cash and Default Charity). Thus, in the region where there is a difference between the Default Charity and Default Cash Choice treatments, we are interested to know whether this difference is significantly larger than in the No Choice treatments.

Table 3: The Importance of Choice in Driving Positive Spill overs: Round 2 Donation Amounts

	Donation Amount in Round 2 Percentiles					
	60th	65th	70th	75th	80th	Tobit
	(1)	(2)	(3)	(4)	(5)	(6)
Default Charity × Round 1 Choice	00	0.6***	1.00***	0.6***	00	0.29*
	(0.13)	(0.27)	(0.26)	(0.13)	(0.05)	(0.21)
Default Charity	00	00	-0.2	00	00	0.03
	(0.07)	(0.14)	(0.13)	(0.07)	(0.02)	(0.1)
Round 1 Choice	-0.2**	-0.4**	-0.8***	-0.6***	00	-0.45***
	(0.09)	(0.19)	(0.18)	(0.09)	(0.03)	(0.14)
Round 2 Only	00	00	-0.6***	-0.2**	00	-0.09
	(0.09)	(0.2)	(0.18)	(0.09)	(0.03)	(0.14)
Constant	0.2***	0.4***	1.00***	1.00***	1.00***	-0.1
	(0.05)	(0.1)	(0.09)	(0.05)	(0.02)	(0.07)
Observations	1802	1802	1802	1802	1802	1802

Columns (1)-(5) show quantile regression estimates and column (6) shows tobit regression estimates censored below at 0 and above at $2. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. Given that we have a directional hypothesis for effects of spill overs, the estimates on the coefficient for Default Charity × R1 Choice are one-tailed, all other tests are two-tailed.

To formally test this, we conduct a series of quantile regressions at the 60th, 65th, 70th, 75th and 80th percentiles to further demonstrate the importance of choice in driving Round

2 behaviour. Table 3 pools together the Choice and No Choice treatments, using the Default Charity No Choice condition as the omitted category, and interacts Default Charity and Choice. The coefficient of interest in this table is the interaction term, Default Charity × R1 Choice, which shows the *additional* Round 2 donation made by subjects in the Default Charity condition who also had the *choice* to donate in Round 1 (i.e., the Choice treatments). Consistent with Figures 1a and 1b, we find that the interaction is insignificant at the 60th percentile (and not shown, for all quantiles below the 60th percentile, indicating that in the Default Charity and Default Cash Choice treatments all subjects are donating \$0.00, while in the Default Charity and Default Cash No Choice treatments subjects are donating the same amounts—50 percent in both conditions donating \$0.00 and an additional 10% donating \$0.30. However, as we move up the distribution, a gap emerges in the amount given in the Choice conditions whereas no gap occurs in the No Choice conditions. For example, in the Choice conditions (see Figure 1a) 8 percent more subjects give at least \$0.20 in Default Charity (68%) than Default Cash (60%), and approximately 15 percent more subjects give at least \$0.80 in Default Charity (78%) than Default Cash (63%). And then as we move further up the distribution we see that in the Choice Conditions the gap disappears at \$1.00 of giving, where in both conditions we see that about 83% both give at least \$1.00 and almost everyone else gives \$2.00. In contrast, inspection of Figure 1b shows that in the No Choice conditions there is never a gap in the distribution between Default Cash and Default Charity. The interaction estimates presented in Table 3 confirm that the gap between Default Charity and Default Cash in the Choice conditions is significantly larger than the (lack of any) gap in the No Choice conditions at the 65th, 70th and 75th percentiles. In column (5), we show the results from a Tobit regression model over the entire distribution of donations and we find qualitatively similar, but smaller and less precise results—the Default Charity is only effective at increasing donations in Round 2 relative to Default Cash when there was a *choice* in Round 1.

3.3 Main Results: Self-Perception & Moral Consistency

Result 3. *Consistent with Hypothesis 4, we find evidence of moral consistency, i.e.,* $\beta^{IV} > 0$*, implying that* $\gamma_A > 0$*; that is, choosing to give in Round 1 increases giving in Round 2. In particular, choosing to give in Round 1 increases the probability of giving in Round 2 by 200% (40 percentage points) and increases the amount given in Round 2 by \$0.59.*

In Table 4, we estimate the effect that the nudge-induced increase in choosing to give in Round 1 has on charitable giving in Round 2 using the instrumental variable approach discussed in Section 2.3. Column (1) and column (2) indicates that giving in Round 1 causes

a 41 percentage point and 44 percentage point (200% increase above the baseline) increase in the propensity to give in Round 2, respectively. In columns (3) and (4) the dependent variable is donation amount in Round 2. Column (3) and (4) indicate that giving in Round 1 causes subjects to increase their giving by \$0.59 (200%) in Round 2. In sum, altruism begets altruism.

Table 4: Moral Consistency: Round 2 Donation Rates & Amounts

	Propensity to Donate		Donation Amount	
	(1)	(2)	(3)	(4)
$\widehat{A_{i,1}}$	0.41**	0.44**	0.59**	0.62**
	(0.23)	(0.23)	(0.32)	(0.33)
Constant	0.25***	0.28***	0.27***	0.31***
	(0.06)	(0.08)	(0.08)	(0.11)
Observations	415	415	415	415
R^2	0.2	0.21	0.21	0.22
Demographic Controls	No	Yes	No	Yes

Columns (1)-(4) present two-stage least square regression estimates and columns (5)-(8) present limited-information maximum likelihood estimates. Demographic controls include gender, employment status (Full-time, part-time, retired or unemployed), and income. Table B2 presents the analogous results for immoral consistency. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. Given that we have a directional hypothesis for the effect of moral consistency, the estimates on the coefficient for $\widehat{A_{i,1}}$ are one-tailed.

The results in Table 4 suggest that nudging virtuous behavior "now" may promote virtuous behavior "later". In other words, the nudge successfully crowds people into giving in Round 2, who would likely not have given in Round 2, by nudging them to give in Round 1.

3.4 Heterogenous Effects: Identity towards Altruism

In this section, we explore the heterogenous effects of identity on moral consistency. In our model, it is straightforward to show that $\frac{\partial \gamma_A}{\partial \Theta_A} < 0$, that is, moral consistency is *decreasing* in the strength of the individual's altruistic identity (Θ_A). Thus, the magnitude of the local average treatment effect, $\beta^I V$, will be greater for those with a weak identity than for those with a strong identity towards altruism.[15]

In columns (1) and (2) of Table 5 our dependent variable is the probability of donating in

[15]Benabou & Tirole's (2011) model, also drawing heavily from self-perception theory, predicts that when weakly-held values are encouraged, individuals respond in a confirmatory way (i.e., morally consistent). In other words, individuals for whom altruism is a weak facet of their identity will behave in a morally consistent way later, when nudged towards altruism now. Thus, the predictions from our model and the Bénabou and Tirole (2011) model about individuals with weak identities towards altruism are similar when $\gamma_A > 0$.

TABLE 5: MORAL CONSISTENCY & IDENTITY: ROUND 2 DONATION RATES & AMOUNTS

Propensity to Donate	Donation Amount			
	(1)	(2)	(3)	(4)
$\widehat{A_{i,1}} \times StrongIdentity$	-0.18	-0.14	0.04	0.1
	(0.39)	(0.39)	(0.53)	(0.53)
$\widehat{A_{i,1}} \times WeakIdentity$	0.83***	0.85***	0.95***	0.95***
	(0.36)	(0.35)	(0.49)	(0.47)
Strong Identity	0.36**	0.36**	0.37*	0.37*
	(0.15)	(0.15)	(0.21)	(0.2)
Constant	0.14*	0.17*	0.15	0.21*
	(0.08)	(0.09)	(0.1)	(0.13)
Observations	415	415	415	415
R^2	0.03	0.04	0.12	0.14
Demographic Controls	No	Yes	No	Yes
χ^2 test				
$\widehat{A_i, 1}\times$ Strong $=\widehat{A_i, 1}\times$ Weak	3.59	3.58	1.58	1.44

OLS regression estimates. Demographic controls include gender, employment status (Full-time, part-time, retired or unemployed), and income. Table B2 presents the analogous results for immoral consistency. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. Given that we have a directional hypothesis for the effect of identity on moral consistency, the estimates on the coefficient for $\widehat{A_{i,1}} \times Weak$ and $\widehat{A_{i,1}} \times Strong$ are one-tailed.

Round 2 and consistent with our model we find that subjects for whom altruism is a weakly held value are significantly more morally consistent than subjects for whom altruism is a strongly held value. However, in columns (3) and (4) our dependent variable is donation amount in Round 2 and we find that the difference between strong and weak altruists, while large in effect size, not significant.

One potential concern is that an individual who has only given to charity once or twice in the past year, but gave a large sum of money, would be classified as having a weak identity towards altruism under our definition. To address this potential problem, in **??** Table B3, we re-classify a weak identity towards altruism as those who report giving 0 times in the past year. Again, we find that they are significantly more morally consistent than subjects who have given 4 or more times in the past year.

Moral consistency also helps to overcome decreases in giving that are typically associated with ask fatigue and multiple donation solicitations. On average, subjects in the Default Charity condition give \$0.53 in Round 1 and \$0.48 in Round 2 for a total average donation of \$1.01. Using the donation rates from the Round 2 Only condition, we know the average donation amount is \$.41 if an individual is only asked to give once. Thus, if there was no moral consistency and the Round 1 and Round 2 decisions were instead independent, then

subjects in the Default Charity condition would donate 15% less. Of course, this difference is magnified more if we consider those subjects who have a weak identity towards altruism. When Round 1 and Round 2 decisions are linked through moral consistency, individuals in the Default Charity condition with a weak identity towards altruism donate a total of $0.98 to charity. However, if Round 1 and Round 2 decisions were independent, then these same subjects would donate 27% less.

4 Conclusion

In this paper, we conducted an experiment to provide evidence that altruism begets altruism. We estimate a local average treatment effect, which is directly informed by self-perception and our model of history dependence. We go beyond the existing literature and show that an increase in choosing to be altruistic now *causes* an increase in altruistic behavior later.

We also believe the findings in this paper generate interesting questions for future research. For example, one interesting question for future research may study whether different types of nudges or a longer length of time between asks result in similar patterns of moral consistency. We obtain exogenous variation in our Round 1 giving by using a default option nudge, but studying whether reminding individuals about social norms around giving, also a popular nudge, also generates moral consistency would be of great academic and practical interest.

References

Adena, Maja and Steffen Huck. 2019. "Giving once, giving twice: A two-period field experiment on intertemporal crowding in charitable giving." *Journal of Public Economics* 172:127–134.

Akerlof, George A and Rachel E Kranton. 2005. "Identity and the Economics of Organizations." *Journal of Economic perspectives* 19 (1):9–32.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91 (434):444–455.

Anik, Lalin, Lara B Aknin, Michael I Norton, and Elizabeth W Dunn. 2009. "Feeling good about giving: The benefits (and costs) of self-interested charitable behavior." *Harvard Business School Marketing Unit Working Paper* (10-012).

Bem, Daryl J. 1972. "Self-perception theory." In *Advances in experimental social psychology*, vol. 6. Elsevier, 1–62.

Bénabou, Roland and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics* 126 (2):805–855.

Benartzi, Shlomo and Richard H Thaler. 2007. "Heuristics and biases in retirement savings behavior." *The journal of economic perspectives* 21 (3):81–104.

Benjamin, Daniel J, James J Choi, and Geoffrey W Fisher. 2010. "Religious identity and economic behavior." Tech. rep., National Bureau of Economic Research.

Blanken, Irene, Niels van de Ven, and Marcel Zeelenberg. 2015. "A meta-analytic review of moral licensing." *Personality and Social Psychology Bulletin* 41 (4):540–558.

Cairns, Jason and Robert Slonim. 2011. "Substitution effects across charitable donations." *Economics Letters* 111 (2):173–175.

Cappelen, Alexander W, Gary Charness, Mathias Ekström, Uri Gneezy, and Bertil Tungodden. 2017. "Exercise improves academic performance." .

Cassar, Alessandra, Giovanna d'Adda, and Pauline Grosjean. 2014. "Institutional quality, culture, and norms of cooperation: Evidence from behavioral field experiments." *The Journal of Law and Economics* 57 (3):821–863.

Castillo, Marco, Ragan Petrie, and Anya Samek. 2017. "Time to Give: A Field Experiment on Intertemporal Charitable Giving." .

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2003. "Optimal defaults." *The American Economic Review* 93 (2):180–185.

Clotfelter, Charles T. 1980. "Tax incentives and charitable giving: Evidence from a panel of taxpayers." *Journal of Public Economics* 13 (3):319–340.

Cronqvist, Henrik and Richard H Thaler. 2004. "Design choices in privatized social-security systems: Learning from the Swedish experience." *The American Economic Review* 94 (2):424–428.

Damgaard, Mette Trier and Christina Gravert. 2018. "The hidden costs of nudging: Experimental evidence from reminders in fundraising." *Journal of Public Economics* 157:15–26.

Dur, Robert and Joeri Sol. 2010. "Social interaction, co-worker altruism, and incentives." *Games and Economic Behavior* 69 (2):293–301.

Eckel, Catherine C, David H Herberich, and Jonathan Meer. 2018. "It's not the thought that counts: A field experiment on gift exchange and giving at a public university." *The Economics of Philanthropy* .

Engl, Florian, Arno Riedl, and Roberto A Weber. 2018. "Spillover effects of institutions on cooperative behavior, preferences, and beliefs." .

Exley, Christine L. 2015a. "Excusing selfishness in charitable giving: The role of risk." *The Review of Economic Studies* 83 (2):587–628.

———. 2015b. "Using charity performance metrics as an excuse not to give." Tech. rep., Working paper.

Feldman, Naomi E. 2010. "Time is money: Choosing between charitable activities." *American Economic Journal: Economic Policy* 2 (1):103–30.

Frey, Bruno S and Stephan Meier. 2004. "Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment." *American Economic Review* 94 (5):1717–1722.

Gee, Laura and Jonathan Meer. 2019. "The Altruism Budget: Measuring and Encouraging Charitable Giving." *NBER Working Paper* (w25938).

Gneezy, Ayelet, Alex Imas, Amber Brown, Leif D Nelson, and Michael I Norton. 2012. "Paying to be nice: Consistency and costly prosocial behavior." *Management Science* 58 (1):179–187.

Gneezy, Uri, Elizabeth A Keenan, and Ayelet Gneezy. 2014. "Avoiding overhead aversion in charity." *Science* 346 (6209):632–635.

Gruber, Jonathan. 2004. "Pay or pray? The impact of charitable subsidies on religious attendance." *Journal of Public Economics* 88 (12):2635–2655.

Huck, Steffen, Imran Rasul, and Andrew Shephard. 2015. "Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model." *American Economic Journal: Economic Policy* 7 (2):326–69.

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.

Karlan, Dean and John A List. 2007. "Does price matter in charitable giving? Evidence from a large-scale natural field experiment." *American Economic Review* 97 (5):1774–1793.

Kessler, Judd B and Katherine L Milkman. 2016. "Identity in charitable giving." *Management Science* 64 (2):845–859.

Kessler, Judd B, Katherine L Milkman, and C Yiwei Zhang. 2019. "Getting the rich and powerful to give." *Management Science* 65 (9):4049–4062.

Kessler, Judd B and Alvin E Roth. 2012. "Organ allocation policy and the decision to donate." *The American Economic Review* 102 (5):2018–2047.

———. 2014. "Getting more organs for transplantation." *The American Economic Review* 104 (5):425–430.

Khan, Uzma and Ravi Dhar. 2006. "Licensing effect in consumer choice." *Journal of Marketing Research* 43 (2):259–266.

Landry, Craig E, Andreas Lange, John A List, Michael K Price, and Nicholas G Rupp. 2010. "Is a donor in hand better than two in the bush? Evidence from a natural field experiment." *American Economic Review* 100 (3):958–83.

List, John A and David Lucking-Reiley. 2002. "The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign." *Journal of political Economy* 110 (1):215–233.

Madrian, Brigitte C and Dennis F Shea. 2001. "THE POWER OF SUGGESTION: INERTIA IN 401 (k) PARTICIPATION AND SAVINGS BEHAVIOR." *Quarterly Journal of Economics* 116 (4).

Meer, Jonathan. 2013. "The habit of giving." *Economic Inquiry* 51 (4):2002–2017.

Meer, Jonathan and Benjamin A Priday. 2019. "Tax Prices and Charitable Giving: Projected Changes in Donations Under the 2017 TCJA." Tech. rep., National Bureau of Economic Research.

Meier, Stephan. 2007. "Do subsidies increase charitable giving in the long run? Matching donations in a field experiment." *Journal of the European Economic Association* 5 (6):1203–1222.

Monin, Benoit and Dale T Miller. 2001. "Moral credentials and the expression of prejudice." *Journal of personality and social psychology* 81 (1):33.

Mullen, Elizabeth and Benoît Monin. 2016. "Consistency versus licensing effects of past moral behavior." *Annual review of psychology* 67.

Nisan, Mordecai and Gaby Horenczyk. 1990. "Moral balance: The effect of prior behaviour on decision in moral conflict." *British journal of social psychology* 29 (1):29–42.

Nisan, Mordechai. 1985. "Limited morality: A concept and its educational implications." *Moral education: Theory and application* :403–420.

Peysakhovich, Alexander and David G Rand. 2016. "Habits of virtue: Creating norms of cooperation and defection in the laboratory." *Management Science* 62 (3):631–647.

Ploner, Matteo and Tobias Regner. 2013. "Self-image and moral balancing: An experimental analysis." *Journal of Economic Behavior & Organization* 93:374–383.

Pollak, Robert A. 1970. "Habit formation and dynamic demand functions." *Journal of political Economy* 78 (4, Part 1):745–763.

Sachdeva, Sonya, Rumen Iliev, and Douglas L Medin. 2009. "Sinning saints and saintly sinners the paradox of moral self-regulation." *Psychological science* 20 (4):523–528.

Sarkisian, Roberto. 2017. "Team Incentives under Moral and Altruistic Preferences: Which Team to Choose?" *Games* 8 (3):37.

Shang, Jen and Rachel Croson. 2009. "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods." *The Economic Journal* 119 (540):1422–1439.

Shu, Suzanne B and Joann Peck. 2011. "Psychological ownership and affective reaction: Emotional attachment process variables and the endowment effect." *Journal of Consumer Psychology* 21 (4):439–452.

Staiger, Douglas and James H Stock. 1997. "Instrument variables regression with weak instruments." *Econometrica* 65 (3):557–586.

Strahilevitz, Michal A and George Loewenstein. 1998. "The effect of ownership history on the valuation of objects." *Journal of consumer research* 25 (3):276–289.

Sunstein, Cass and Richard Thaler. 2008. "Nudge." *The politics of libertarian paternalism. New Haven* .

Thaler, Richard H and Cass R Sunstein. 2003. "Libertarian paternalism." *The American Economic Review* 93 (2):175–179.

Yörük, Barış K. 2013. "The impact of charitable subsidies on religious giving and attendance: Evidence from panel data." *Review of Economics and Statistics* 95 (5):1708–1721.

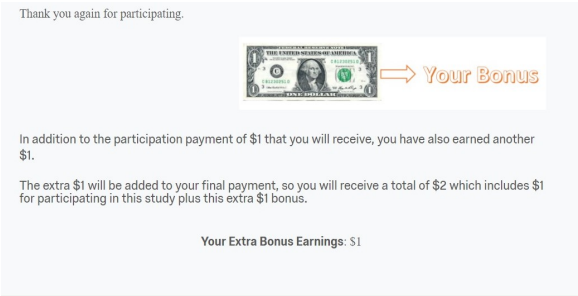
———. 2014. "Does giving to charity lead to better health? Evidence from tax subsidies for charitable giving." *Journal of Economic Psychology* 45:71–83.

Zanna, Mark P. 1972. "Inference of belief from rejection of an alternative action." *Representative Research in Social Psychology* .

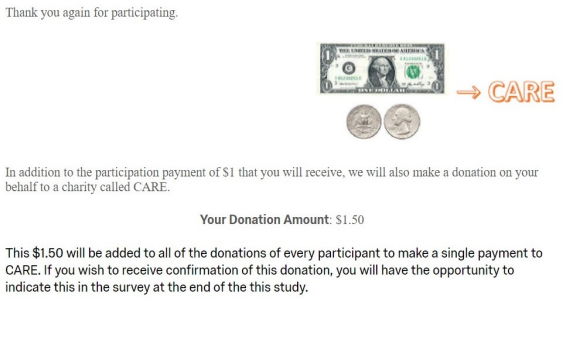
Appendix A Appendix A: Tables and Figures

This section is meant for online publication only.

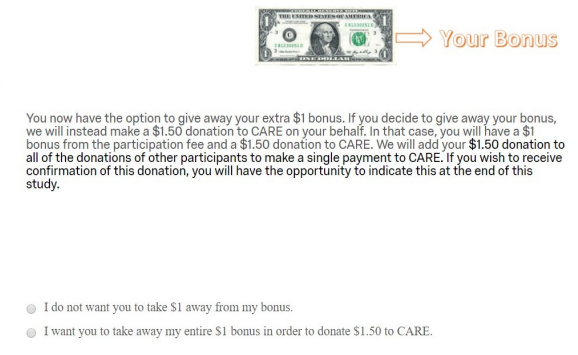
FIGURE A1: DONATION EXPERIMENT SCREENSHOTS



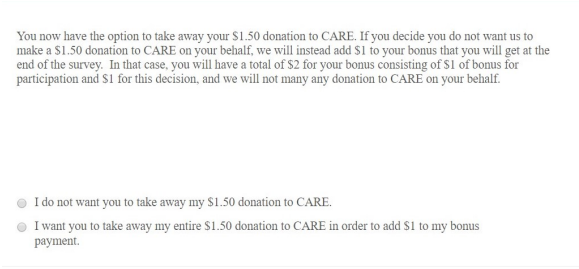
(A) ROUND 1, CASH ENDOWMENT



(B) ROUND 1, CHARITY ENDOWMENT

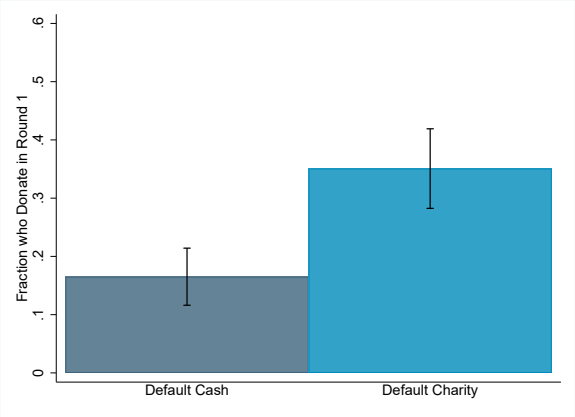


(C) SWAP CASH FOR DONATION

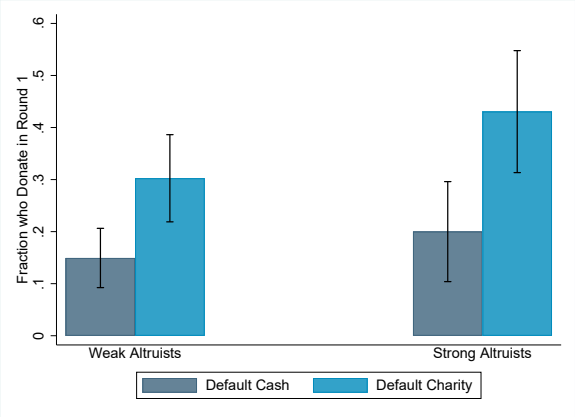


(D) SWAP DONATION FOR CASH

Figure A2: First-Stage: Round 1 Donation Rates



(a) Round 1 Rates



(b) Round 1 Rates by Convictions

Average donation rates in Round 1 by treatment assignment with 95% confidence intervals. Figure A2a shows that subjects assigned to the Default Charity treatment are 19 percentage points more likely (more than a 100 percent increase) to donate to charity in Round 1 than subjects assigned to the Default Cash treatment (t-test: p-value<.0001). In Figure A2b we look at the effect of the default option nudge by subjects' number of charitable donations in the past 12 months. In general, we find that the Default Charity condition significantly increases donation rates regardless of the subjects' past donation history. For example, subjects with strongly-held (weakly-held) views towards altruism are 23 (15) percentage points more likely to give under the Default Charity than Default Cash condition (t-test: p-value=.003 and p-value=.002, respectively).

Table A1: Round 2: Multiple Price List for Donation Experiment

Option 1:	Add \$1.00 to your bonus and Donate \$0 to Save the Children.
Option 2:	Add \$.90 to your bonus and Donate \$.20 to Save the Children.
Option 3:	Add \$.80 to your bonus and Donate \$.40 to Save the Children.
Option 4:	Add \$.70 to your bonus and Donate \$.60 to Save the Children.
Option 5:	Add \$.60 to your bonus and Donate \$.80 to Save the Children.
Option 6:	Add \$.50 to your bonus and Donate \$1.00 to Save the Children.
Option 7:	Add \$.40 to your bonus and Donate \$1.20 to Save the Children.
Option 8:	Add \$.30 to your bonus and Donate \$1.40 to Save the Children.
Option 9:	Add \$.20 to your bonus and Donate \$1.60 to Save the Children.
Option 10:	Add \$.10 to your bonus and Donate \$1.80 to Save the Children.
Option 11:	Add \$0 to your bonus and Donate \$2.00 to Save the Children.

Appendix B Additional Moral Consistency Results

Appendix B.1 Immoral Consistency

To estimate the causal effect of keeping the money in Round 1 ($c_{i,1}$) on the likelihood of also keeping the money in Round 2 ($c_{i,2}$) we instrument for keeping the money in Round 1 using the assignment to the Default Cash condition. We use a similar interaction as in equation 8.

$$c_{i,2} = \lambda_0 + \lambda_1^{IV}\widehat{c_{i,1}} \times \mathbf{1}\left[\Theta_A = 0\right] + \lambda_2^{IV}\widehat{c_{i,1}} \times \mathbf{1}\left[\Theta_A = 1\right] + \lambda_3\mathbf{X} + \varepsilon_i, \tag{7}$$

In Table **??**, we estimate equation 7 to examine whether there is evidence of immoral consistency; that is, does keeping the cash in Round 1 cause an increase in keeping the cash in Round 2. We do not find evidence consistent with immoral consistency on the extensive margin (columns (1) & (2)), but columns (3)& (4) show that keeping the cash in Round 1 causes subjects to keep more cash in Round 2.

Result 4. *We find no evidence that subjects who have a strong identity towards altruism behave in a morally balanced manner when nudged towards selfishness.*

Columns (2) & (4) test for the second part of the Bénabou and Tirole (2011) hypothesis, which states that subjects who are nudged away from a strongly-held value will respond in a contradictory manner. Thus, we hypothesized that subjects for whom altruism is a strongly-held value, but are nudged towards selfishness, would less selfish (or more altruistic) in Round 2. However, we do not find support for this hypothesis.

Appendix B.2 Identity & Moral Consistency

To test these hypotheses about identity from section 3.4, we will use the same specification as equation 6, but with an interaction between endogenous regressor (A_{t-1}) and the strength of conviction towards altruism, either weak ($\Theta_A = 0$) or strong ($\Theta_A = 1$), and instrument for Round 1 donation behavior using the assignment to the Default Charity treatment interacted with the strength of the conviction. Our specification for this hypothesis is therefore given by

$$A_{i,2} = \delta_0 + \delta_1^{IV}\widehat{A_{i,1}} \times \mathbf{1}\left[\Theta_A = 0\right] + \delta_2^{IV}\widehat{A_{i,1}} \times \mathbf{1}\left[\Theta_A = 1\right] + \delta_3\mathbf{X} + \varepsilon_i, \tag{8}$$

Hypothesis 5. ***Altruism as a Weak Facet of Identity I****:*
$\delta_1^{IV} > 0$*: individuals who hold altruism as a weak facet of their identity will behave morally consistently.*

Table B2: Local Average Treatment Effects for Immoral Consistency: Round 2 Donation Rates & Amounts

	Propensity to Donate				Donation Amount			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8))
$\widehat{c_{i,1}}$	0.15	0.15	.	.	0.3*	0.31*	.	.
	(0.16)	(0.16)			(0.16)	(0.16)		
$\widehat{c_{i,1}} \times StrongIdentity$	.	.	0.17	0.2	.	.	0.02	0.05
			(0.24)	(0.25)			(0.27)	(0.26)
$\widehat{c_{i,1}} \times WeakIdentity$	.	.	0.11	0.09	.	.	0.48**	0.47**
			(0.23)	(0.22)			(0.24)	(0.24)
Strong Identity	.	.	-0.09	-0.13	.	.	0.27	0.24
			(0.25)	(0.24)			(0.27)	(0.26)
Constant	0.78***	0.72***	0.82***	0.78***	0.57***	0.53***	0.45**	0.42**
	(0.12)	(0.14)	(0.18)	(0.19)	(0.12)	(0.14)	(0.19)	(0.2)
Observations	415	415	415	415	415	415	415	415
R^2	0.08	0.1	0.09	0.1	0.21	0.22	0.12	0.14
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
χ^2 test								
$\widehat{c_i, 1}\times$ Strong $=\widehat{c_i, 1}\times$ Weak			.03	.11			1.58	1.44
(p-value)			(.86)	(.74)			(.21)	(.23)

Demographic controls include gender, employment status (Full-time, part-time, retired or unemployed), and income. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Hypothesis 6. ***Altruism as a Weak Facet of Identity II****: if $\gamma_A > 0$, then individuals who hold altruism as a weak facet of their identity will behave more morally consistent than individuals who hold altruism as a strong facet of their identity; that is, $\delta_1^{IV} > \delta_2^{IV}$.*

Bénabou and Tirole (2011) also predict that when strongly-held convictions are challenged, individuals will be more likely to respond in a contradictory way to the challenge to restore their self-image. This means that for those individuals who have a strong identity towards altruism but are nudged towards selfishness (i.e., the Default Cash condition), Bénabou and Tirole (2011) predicts that individuals will respond by being more altruistic in the future. We formally state this in Hypothesis 7.

Hypothesis 7. ***Altruism as a Strong Facet of Identity****:*
(i) $\lambda_2^{IV} < 0$: individuals who hold altruism as a strongly-held facet of their identity will respond in a contradictory way to a nudge towards selfishness.

TABLE B3: LOCAL AVERAGE TREATMENT EFFECTS: ROUND 2 DONATION RATES & AMOUNTS

Panel A: Moral Consistency	Propensity to Donate		Donation Amount	
	(1)	(2)	(3)	(4)
$\widehat{A_{i,1}} \times StrongValue$	-0.18	-0.13	0.04	0.12
	(0.39)	(0.39)	(0.53)	(0.53)
$\widehat{A_{i,1}} \times WeakValue$	1.17**	1.26**	1.50*	1.58*
	(0.63)	(0.67)	(0.96)	(0.99)
Strong Value	0.42***	0.42***	0.43*	0.43**
	(0.15)	(0.15)	(0.22)	(0.22)
Constant	0.07	0.09	0.1	0.08
	(0.08)	(0.1)	(0.12)	(0.16)
Observations	245	245	245	245
R^2	.	.	0.002	0.03
χ^2 test				
$\widehat{A_i, 1}\times$ Strong $=\widehat{A_i, 1}\times$ Weak	3.30*	3.21*	1.75	1.69

Panel B: Immoral Consistency	Propensity to Keep		Keep Amount	
	(1)	(2)	(3)	(4)
$\widehat{c_{i,1}} \times StrongValue$	0.17	0.19	0.02	0.06
	(0.24)	(0.25)	(0.27)	(0.26)
$\widehat{c_{i,1}} \times WeakValue$	0.46	0.46	0.75	0.79
	(0.46)	(0.45)	(0.48)	(0.49)
Strong Value	0.23	0.2	0.52	0.52
	(0.45)	(0.44)	(0.47)	(0.48)
Constant	0.51	0.52	0.2	0.17
	(0.42)	(0.41)	(0.43)	(0.44)
Observations	245	245	245	245
R^2	0.03	0.05	0.002	0.03
χ^2 test				
$\widehat{c_i, 1}\times$ Strong $=\widehat{c_i, 1}\times$ Weak	.32	.27	1.75	1.69

This table replicates Table 5 and Table B2 but we have redefined a Weak Identity towards altruism as those subjects who report that they gave 0 donations in the past year. Robust standard errors in parentheses and *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. Given that we have a directional hypothesis for the effect of identity on moral consistency, the estimates on the coefficient for $\widehat{A_{i,1}} \times Weak$ and $\widehat{A_{i,1}} \times Strong$ are one-tailed.