Intertemporal Prosocial Choice: The Inconsistency Puzzle

Islam, Marco

2022

Document Version:
Other version

Link to publication

Citation for published version (APA):
Islam, M. (2022). *Intertemporal Prosocial Choice: The Inconsistency Puzzle*. (Working Papers; No. 2022:12).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2022:12

Department of Economics
School of Economics and Management

Intertemporal Prosocial Choice: The Inconsistency Puzzle

Marco Islam

July 2022

Intertemporal Prosocial Choice: The Inconsistency Puzzle

Marco Islam*

July 2022

Abstract

How does delay in the realization of a prosocial decision affect prosocial choice? This paper first provides a meta-analysis that collects existing evidence on the temporal consistency of prosocial behavior. I show that the evidence on the delay effect on prosocial choice is contradicting but appears reconcilable by a moderating factor: *repeated interaction.* Motivated by this finding, I conduct an intertemporal donation experiment to closely investigate this moderation effect. I design an experiment that mimics a telephone fundraiser and vary both the timing of the donation (*immediate* vs. *delayed*) and the frequency of interaction (*one-shot* vs. *repeated interaction*). The results reveal that both under repeated and one-time interaction delayed donations increase relative to immediate donations but the increase is not statistically significant. This evidence suggests that repeated interaction (via telephone) does not provide the conditions for delay to increase prosocial behavior.

JEL codes:D64, D90, C91

Keywords: intertemporal choice; prosocial behavior; charitable giving; repeated interaction

*Department of Economics, Lund University, email: marco.islam@posteo.de. I am very thankful to Marta Serra-Garcia who provided excellent advise throughout the entire study and who granted me access to the facilities and resources of UCSD. I also thank Arno Apffelstaedt, Thomas Epper, Jaromir Kovarik, Petra Thiemann and Erik Wengström for valuable comments and Sara L. Neff as well as Ehsan Amozegar for great research assistance. I acknowledge financial support from the Institute of Economic Research at Lund University. The experiment was approved by the Ethics Review Board of the University of California, San Diego (200053XX). It is preregistered at AsPredicted under the registration number 38628.

1 Introduction

Many economic decisions with repercussions for others are affected by time in that the decision outcome is delayed from the point of decision-making. For example, donation decisions are often made with foresight and fundraisers rely on the commitment of their donors to keep their earlier expressed willingness to donate. Time is also an important factor for long-term cooperation such as the provision of long-term public goods. Many public goods (e.g., climate protection measures) involve a considerable time lag until their implementation. This raises the important question: *How does time affect prosocial choice?*

Unfortunately, the answer to that question is not trivial. Given the relevance of the topic, a number of researchers have investigated the relation of time delay in the realization of prosocial choice. To establish causal inference, all of those studies experimentally varied the point of the prosocial act (e.g., the donation), that is, they deferred the realization of that act from the point of decision making. In the economic literature such a practice of deferral is referred to as intertemporal choice and common to analyze the role of time (Frederick et al., 2002). To investigate its impact on social preferences, the economic research, to date, has studied intertemporal choice for example in the context of charitable giving (Breman, 2011), Dictator Games (Kovarik, 2009) or Public Good Games (Kölle & Lauer, 2019). They have conducted laboratory (Andreoni & Serra-Garcia, 2021), online (Capra et al., 2019) and field experiments (Breman, 2011) and treated student samples (Noor & Ren, 2011), registered donors of charity organizations (Breman, 2011) and national representative samples of populations (Apffelstaedt & Pieters, 2019).

Considered individually, all of the existing studies in this field provide a holistic story including interesting treatment effects for delay and reasonable explanations for those effects. In the "big picture", however, the existing evidence poses a puzzle: While some studies find that delay has a negative effect on the willingness to give (e.g., Dreber et al., 2016; Kovarik, 2009), others claim the exact opposite (e.g., Andreoni & Serra-Garcia, 2021; Breman, 2011; Kölle & Wenner, 2021) and yet some others find no effect at all (Kölle & Lauer, 2019). Even the theoretical explanations brought forward to explain the findings appear inconsistent. For example, Dreber et al. (2016) extended the dual-self model by Fudenberg and Levine (2006), to account for altruistic preferences. They derived the prediction that delaying payments would reduce the utility of the short-run self with the consequence that the short-run self becomes tempted to give whereas the long-run self becomes tempted to keep. Andreoni and Serra-Garcia (2021), on the other hand, proposed a model where donated money is discounted less than money kept for oneself resulting in a prediction where short-run donors are tempted to keep and long-run donors are tempted to give. Focusing on the delay variation of the prosocial act in isolation, the existing analyses disregard that their detected treatment effects may have resulted from the underlying conditions, for example, from some specific design features that acted as moderators. From a meta perspective, their contradicting results do not suggest that the effect of delay on prosocial choice is clear. Instead, the delay effect seems to be context-dependent.

This study attempts to find a solution to this puzzle. It explores the research question whether the delay effect on prosocial choice could be dependent on conditions, which so far have been disregard in the experimental analyses. That is, it seeks to identify the underlying factors that provide the "fertile ground" for delay to cause time-inconsistent prosocial behavior.

To address this research question, the methodological approach of this study is twofold. First, I conducted a meta-analysis including all experimental studies that examined intertemporal prosocial behavior. As shown by Abeler et al. (2019) and Embrey et al. (2018), such an exercise can be a useful point of departure. For the purpose of solving the inconsistency puzzle, it can help to detect salient data patterns and identify design features that may be important in the context of intertemporal choice. Recall that unlike the individual studies, the purpose of this analysis is to explore background design features, which were kept constant within a study and thus can only be investigated in between-study comparisons.

My meta-analysis includes a total of 11 studies (14 experiments and 42 treatment conditions) and reveals three important findings. First, it highlights the conflicting evidence of the delay effect on prosocial choice in the literature. Second, it shows that the heterogeneity in the delay effects cannot be explained by different delay lengths employed in the experimental studies. Third, focusing on a number of different design features, which varied only across the experiments, it identifies the feature of repeated interaction as the most promising candidate to provide the necessary conditions for delay to trigger an increase in prosocial behavior. Relative to one-time interactions, the delay effect on prosocial behavior increases by about 0.4 to 0.65 standard deviations (sd) in experiments in which subjects interact repeatedly. These effect sizes are very sizable. The variable *repeated interaction* indicates that a subject interacts repeatedly either with the experimenter or with other subjects of the experiment and, because a subject can anticipate the future interaction, it likely serves as an external reminder of the delayed prosocial choice. In one-shot experiments, such an reminder does obviously not exist.

In a second step, I sought to test the identified feature – *repeated interaction* – more carefully by means of a donation experiment (N=201) conducted between April and May 2020. The prosocial act was a donation to a real charity organization (*GiveDirectly*). I employed an experimental setup that builds on individual telephone interviews in order to mimic a situation of a classic fundraiser. Moreover, the standardized procedure of the telephone calls and the personal interaction between the interviewer and the study participant allowed for relatively much experimental control. The experimental design was between-subject and included two treatment variations. The first variation was in the delay of the donation. While subjects in the Now conditions made their donations immediately, the donations of subjects assigned to the Later condition were implemented with a delay of seven days. The second variation was in the frequency of interaction. Subjects either participated once (One-Shot treatment) or twice over the period of a week (Repeated Interaction treatment). As a result, the experiment had a two-by-two design and constituted of four treatments conditions.

The results of my experiment do not support the evidence of the meta-study. I do not find that multiple interactions implemented through a longitudinal choice setup induce a donation behavior that is different from a One-shot setup. Although the delay effect under repeated interaction is positive, in line with my hypothesis, it is economically small (approximately 0.13 to 0.16 sd) and non-significant. If anything, the delay-induced increase in prosocial choice is larger for the One-Shot treatments. When pooling all data, these results remain largely unchanged. I do find that delay in general can cause an increase in donations by approximately 0.34 to 0.36 sd, but this result is only significant on the 10% level and, if anything, it is driven by the stronger delay effect in the One-Shot treatments.

Given my sample size and the observed variance in donations, I have 80% power to detect a delay effect of 0.34 sd captured by the difference-in-difference estimator in the pooled sample and an effect of approximately 0.45 sd when it is measured directly in a sample that is divided by interaction. These effect sizes are well below the effect size detected in the first experiment by Andreoni and Serra-Garcia (2021), which laid the foundation for my experimental design. I am therefore confident that my analysis is sufficiently powered.

My study makes two important contributions: First, to the best of my knowledge, this study is the first to collect the literature on intertemporal prosocial choice and bring together its seemingly contradicting findings about the effect of delay. To date, the body of literature which has studied this effect (Andreoni & Serra-Garcia, 2021; Apffelstaedt & Pieters, 2019; Breman, 2011; Capra et al., 2019; Dreber et al., 2016; Islam, 2016; Kim, 2019; Kölle & Lauer, 2019; Kölle & Wenner, 2021; Kovarik, 2009; Noor & Ren, 2011) shows striking similarities. Yet, there exists surprisingly little referencing among another and nearly all of the findings presented in the papers are evaluated without context to the related studies. This masks the fact that the detected findings may not be as robust or readily generalizable as suggested. My study helps to sort the existing evidence on this topic and contributes by searching for the missing puzzle peace that allows to better understand the "big picture".

Second, my analysis contributes to the more general strand of literature, which has investigated the factor *time* around prosocial choice. It has studied the impact of narrow bracketing (Adena & Huck, 2019), the discounting of others consumption (Albrecht et al., 2011; Howard, 2013; Shapiro, 2010), the discounting of social preferences (Shapiro, 2020) and interpersonal discounting (Chopra et al., 2021; Rong et al., 2018; Rong et al., 2019).[1] My work differs from this literature in that it focuses on the temporal consistency of prosocial choice. It seeks to understand the drivers of inconsistencies and learn about the potential of time delay to increase social welfare. Such knowledge is of first hand relevance for charity organizations and public policies.

The remainder of the paper is structured as follows: In section 2, I present the design and the results of my meta-study. In section 3, I introduce my experiment designed to test the design feature identified through the meta-study and present its results. In section 4, I provide a brief discussion of the analysis and conclude.

[1] More broadly, my work also relates to the literature that studies prosocial choice under uncertainty since a delay of outcomes may induce such uncertainty. The evidence from that literature suggests that under asymmetric uncertainty (i.e., uncertainty that only affects the prosocial outcome but not one's individual outcomes) prosociality can decrease because the uncertainty can create a "moral wiggle room" (e.g., Exley, 2016). Brañas-Garza et al. (2021), however, refine this result and show that under symmetric uncertainty, prosociality can increase.

Table 1: List of studies included in meta study

Study	Experiments	Treatments	N	Country	Pro-Social Task	Platform
Andreoni and Serra-Garcia (2021)	3	11	759	USA	Donation to charity	Lab
Apffelstaedt and Pieters (2019)	1	2	169	UK	Public Goods Game	Prolific
Breman (2011)	2	5	1134	Sweden	Donation to charity	Field
Capra et al. (2019)	1	2	318	USA	Donation to charity	MTurk
Dreber et al. (2016)	1	2	1417	USA	Dictator Game	MTurk
Islam (2016)	1	2	106	Denmark	Public Goods Game	Lab
Kim (2019)	1	2	226	USA	Prisoner's Dilemma	Lab
Kölle and Lauer (2019)	1	2	248	Germany	Public Goods Game	Lab
Kölle and Wenner (2021)	1	12	104	Germany	Dictator Game	Lab
Kovarik (2009)	1	7	204	Spain	Dictator Game	Lab
Noor and Ren (2011)	1	2	48	USA	Donation to charity	Lab

2 Meta-Analysis

2.1 Design

My meta-analysis comprises a total of 11 different studies including 14 separate experiments and 42 treatment conditions. All included studies involve experimental designs that vary the delay of a prosocial decision. I require that every experiment includes at least two comparable treatment conditions, between- or within-subject, which are identical in every aspect except for the timing of the prosocial act. In one treatment condition a subject takes a decision, which is implemented immediately, so that costs and benefits arise immediately. In another treatment, a subject takes a decision, which is implemented at a later point in time. This implies that both costs and benefits are delayed. This comparability of two otherwise identical decisions allows to study the temporal consistency of prosocial choice. Further, I only include experiments that use monetary incentives (i.e., no hypothetical decisions) and employ a random mechanism for treatment assignment if its design is between-subject. Finally, I only include studies that are publically accessible either as published or as working papers before the April 2020.

Despite the similar research purposes of some of the included studies, the conducted experiments are different. For example, some mimic a fundraiser setup with no strategic component while others employ games with strategic interaction among the study participants. Some experiments are conducted in the laboratory and involve rather high stakes while others only pay small incentives to subjects from the online platform Amazon Mechanical Turk. Due to this heterogeneity in the experimental designs, it is difficult to compare the initial outcome variables of the studies on equal terms. Moreover, since I am not in possession of most of the studies' raw data, I cannot conduct insightful analyses at the individual level. It is, however, possible to calculate standardized effect sizes across treatments and interpret their sizes as well as their directions in light of each experiment's specific design features. This is what I set out to do in this meta-study.

As my standardized measure for the effect size, I use *Cohen's d* (Cohen, 1988). This is the most commonly used measure for effect sizes. It is calculated as follows:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{(s_1^2 + s_2^2)/2}},$$

where μ_1 and μ_2 represent the mean of the control group (immediate prosocial behavior) and the treatment group (delayed prosocial behavior), respectively and s_1 and s_2 are the corresponding standard deviations. The interpretation of the measure is in units of sd. I am able to retrieve effect sizes either directly from the studies, by calculating them based on summary statistics published in the papers or by using the studies raw data, if it is available.

The included studies are conducted in six different countries. Most of the study participants are students, although a few experiments also used non-student samples from online platforms or donor lists of charity organizations. Moreover, the experiments employ different tasks to elicit prosocial behavior and are conducted using different platforms. Table 1 provides an overview of all included studies in this meta-study. Table A1 in the Appendix summarizes all design features that were meant to be constant within an individual study but can now be analyzed as exogenous variables in this meta-study.

2.2 Results

The approach for my analysis is threefold: First, I compare the treatment effects across the 11 studies of this meta-analysis to illustrate the point that the effect of delay on prosocial choice is not consistent. Second, I test if the manipulated variable – the delay length – could explain the different results of the studies. After all, the delay length employed in the studies varied from two days to one year. Third, I exploit the design differences of the experiments and study the role of other salient design features by performing meta-analysis and OLS regressions. Unlike each individual study, which kept those design features constant, this allows me to understand if the heterogeneity across studies can be explained by one or more design features.

Figure 1 displays a forest plot that shows the effect sizes for all included studies in the meta-study sorted by frequency of interaction. The effects represent the changes in prosocial behavior relative to the control condition of the respective experiment. Recall that the prosocial act in the control conditions immediately followed the decision, which implies that there existed no delay. If a study included different experiments or multiple treatments that all allow for comparisons to the control condition, the displayed effect size represents the average effect. Figure B1 in the Appendix shows an additional forest plot that lists all experiments and treatment conditions individually.

Finding 1. *The delay effect on prosocial choice is not consistent across studies.*

As illustrated in Figure 1, the documented effect size for the delay effect on prosocial choice of the 11

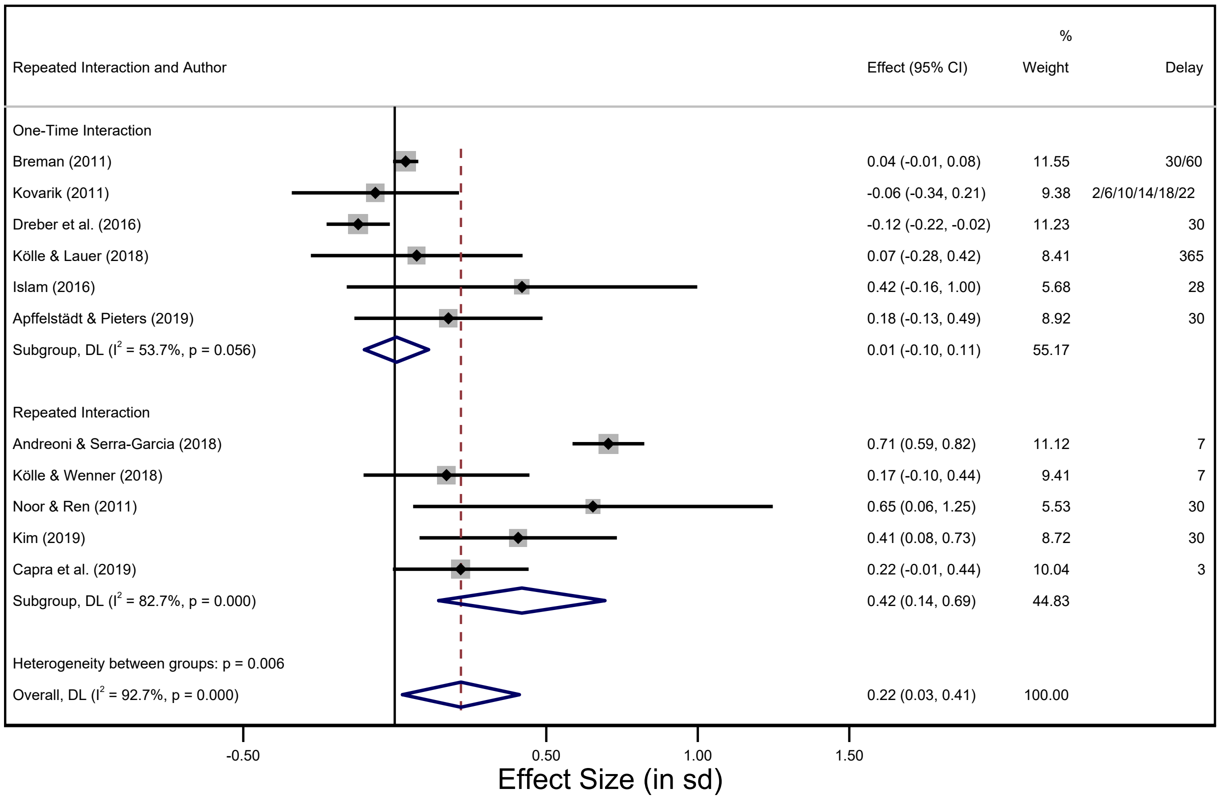


Figure 1: Effect Sizes by Delay

Notes: The figure shows the effect sizes and the respective confidence intervals for different treatment condition by interaction. Each data point represents one study. For studies, which include multiple experiments or multiple treatments, the effect sizes are pooled. Treatment effects are measured relative to the respective control condition (i.e., conditions with no delay). The weights are from a random effects model and reflect the confidence level of the results.

analyzed studies is very heterogeneous. Despite the obvious similarities in the experimental designs, the effects vary from -0.12 sd to 0.71 sd. While four studies (Andreoni & Serra-Garcia, 2021; Capra et al., 2019; Kim, 2019; Noor & Ren, 2011) find a significant positive effect, one study (Dreber et al., 2016) finds a significant negative effect and six studies (Apffelstaedt & Pieters, 2019; Breman, 2011; Islam, 2016; Kölle & Lauer, 2019; Kölle & Wenner, 2021; Kovarik, 2009) find non-significant results. Although the aggregate effect appears to be positive, this highly inconsistent evidence raises questions about the delay effect.

Finding 2. *Different delay lengths cannot explain the heterogeneity in effect sizes.*

In addition to the effect sizes, Figure 1 also reports the different delay lengths (in days) employed as experimental treatments. Note that some studies include multiple treatments with different delays. In those cases the figure illustrates the aggregate effect. The delay lengths vary from two days to 365 days. It is obvious that the delay length does neither seem to play a role for the sign of the treatment effect nor for its size. Positive as well as non-positive effect sizes exist along the entire range of the delay. This visual evidence is supported by linear regressions for which it is possible to exploit the full heterogeneity across treatments

Table 2: Meta Analysis - Regression Results

	Dependent Variable: Effect Size (in sd)					
	Meta-Reg			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Delay	-0.001		-0.000	-0.001		-0.000
	(0.001)		(0.001)	(0.001)		(0.001)
Repeated Interaction		0.407***	0.650**		0.443***	0.625**
		(0.112)	(0.303)		(0.115)	(0.224)
Lab Experiment			-0.040			-0.059
			(0.133)			(0.135)
Within-Subject			0.232			0.350*
			(0.177)			(0.188)
Strategic Game			0.139			0.202
			(0.204)			(0.225)
Random Pay			-0.102			-0.041
			(0.238)			(0.236)
Real Effort			-0.602**			-0.735**
			(0.235)			(0.260)
Continuous Choice			0.095			0.118
			(0.249)			(0.220)
Constant	0.291***	0.031	-0.068	0.290***	0.029	-0.082
	(0.075)	(0.083)	(0.252)	(0.078)	(0.084)	(0.236)
Observations	28	28	28	28	28	28
Adj. R-squared	-0.007	0.422	0.782	-0.02	0.34	0.63

Notes: Meta and OLS regressions for the effect sizes of all treatments included in the meta-study. The effect size quantifies the treatment effect of the delay manipulation relative to the control condition, which did not include any delay. The measure of the effect size is *Cohen's d*. All units are in standard deviations.Standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All design features included as exogenous variables occurred in at least 2 different studies. Of the 28 experimental treatments included for this regression, 15 used a design with repeated interaction, 22 were conducted in the lab 12 were within-subject, 4 employed a strategic setup, 8 a random payment method, in 7 subjects had to exert real effort and in 20 the subjects' choice was continuous.

and experiments (i.e., within-study heterogeneity). Columns (1) and (4) in Table 2 show the results of a meta-analysis regression and an OLS regression respectively, where the delay length is the only explanatory variable. Columns (3) and (6) show further regressions with additional design features as controls. All of the regression outcomes suggest that an increase in the delay length does not affect prosocial choice. The regression estimate for the variable *Delay* is virtually zero and the adjusted R^2 of the regression models (1) and (4) is negative. For robustness, I also study the effect of the delay length in logarithmic and quadratic form. Table A3 in the Appendix displays the results of the respective analyses. The interpretation that delay length does not affect prosocial choice remains unchanged.

Next, I turn to alternative design features and exploit the heterogeneity between studies to test if any of those features performs better at explaining the conflicting evidence of the delay effect. While it is easy to study the effect of each feature on the effect size individually, it becomes more complicated to test their impact simultaneously because a number of design features are strongly correlated. For example, many laboratory experiments use cash payments, but none of the experiments conduced online or in the field does so. To account for this, I first study the relationship between each design feature and the treatment effect individually and in a second step run regressions allowing for multiple exogenous variables at once. As exogenous variables, I include those design features, which appeared in at least three treatments in two different studies and for which their exist no issues of multicollinearity.

Finding 3. *Of all design features in the experiments, an environment in which participants interact repeatedly – either with each other or with the experimenter – provides the best conditions for a positive delay effect.*

The studies listed in Figure 1 are sorted by frequency of interaction. The upper part of the figure includes all treatment conditions for which the interaction in the experiment was one-time. The bottom part includes treatments with repeated interaction. Subjects who interacted repeatedly either participated in multiple parts of the experiment on different days (longitudinal studies) or in multiple rounds on one day. With the exception of the study by Kim (2019), all experiments, which meet the criterion of repeated interaction, are longitudinal. The feature likely acts as a reminder of a subject's prosocial choice and, because it is common knowledge at the point of decision-making, the reminder can be anticipated by the subject. Figure 1 suggests that repeated interaction can explain a sizable part of the differences in the observed treatment effects. First, it seems that effect sizes of treatments with repeated interaction are more likely to have a positive sign. In fact, all of the five studies and all of the treatment conditions (see Figure B1 in the Appendix) that employ repeated interaction, exhibit a positive average effect size. Four of the studies and nine of the treatments are statistically significant. Among the experiments with one-time interaction, only four of six experiments (and seven treatments) have a positive sign, the effect sizes are considerably smaller than under repeated interaction and positive effects are at best marginally significant (Breman, 2011). Although much of this difference is driven by the treatment conditions of the experiments by Andreoni and Serra-Garcia (2021), Figure B2 in the Appendix shows that the difference in the effect sizes remains statistically significant even if one excludes their experiments.

The regression results presented in columns (2), (3), (5) and (6) of Table 2 provide further support for this evidence. They suggest that relative to one-shot experiments, the effect sizes of experiments with repeated interaction increase by 0.4 to 0.65 standard deviations on average. To ensure that this estimated effect is not driven by one individual study, I conducted 11 additional regressions, where for each regression one study was excluded from the analysis (see Figure B3 in the Appendix). While the size of the coefficient naturally reacts to the exclusion of some particular studies (i.e., Andreoni & Serra-Garcia, 2021; Kölle & Lauer, 2019), the qualitative interpretation of the regression remains unchanged. For all of the 11 robustness regressions, the effect of repeated interaction is statistically significant.

My analyses in columns (3) and (6) only distinguishes between whether the excluded experiment had a strategic component or not. I can, however, also control for the exact nature of the task used to elicit prosocial behavior. The tasks employed in the experiments include donations to charities, dictator games, public good games and prisoner's dilemmas. The regression results in Table A4 in the Appendix reveal that controlling for the prosocial choice task does not change the treatment effect estimated for the variable *repeated interaction*.

Considering the regression results in columns (5) and (6), one might also get the impression that real effort could provide the conditions for a significant negative treatment effect of delay. While this appears to be an interesting finding on the first glance, it needs to be treated with caution. My meta-analysis only includes two studies that examined real effort, thus the power to critically reflect on this finding is low. Moreover, the significant effect of real effort only exists when the variable is examined in a set of independent variables. Studied individually, I am no longer able to reject the hypothesis that real effort causes an effect that is significantly different from zero (see Table A2 in the Appendix).

Finally, I stress the fact that a meta-study consisting of 11 studies cannot provide conclusive summarizing insights. However, allowing to contrast different experimental designs, my meta-analysis does serve as a useful exercise to identify potential design features that could explain the heterogeneous evidence across experiments. The main take-away of this meta-study is thus the identification of repeated interaction as one *potentially* influential feature.

3 Experiment

In the following, I investigate the role of repeated interaction more systematically. I conduct a novel donation experiment designed to test the following hypothesis:

Hypothesis. *Repeated interaction, generated in longitudinal experiments causes an increase in altruism with delay.*[2]

This hypothesis is a direct outcome of the meta-analysis, conducted in the previous section. The goal of my experiment is to test if repeated interaction in the form of a longitudinal design provides the necessary fertile ground for delay to increase prosocial behavior.

3.1 Design

To study the effect of delay on prosocial choice, I conducted an intertemporal donation experiment. I employed a simple individual choice task in order to avoid strategic interaction among the study participants. The experiment was executed over telephone with 201 students from the University of California, San Diego between April and May 2020. The calls were conduced by two research assistants. The reason I decided to call the subjects rather than setting up a classic computerized experiment was to stress the interaction between the participant and the experimenter (or interviewer) and to emulate a more realistic situation of a fundraiser. All donations in this study were made to the charity organization *GiveDirectly*.

[2]The preregistered version of the hypothesis included the additional specification that repeated interaction induces "dynamic image concerns". However, since image concerns are not studied in my experiment I dropped this specification retrospectively.

Table 3: Experimental Treatments

Interaction	Payment Date	
	Immediate	One week delay
One-Shot	OS - Now	OS - Later
Repeated Interaction	RI - Now	RI - Later

As illustrated in Table 3, the experiment involved four between-subject treatments, which varied in the amount of interaction and in the date of the donation. Depending on treatment assignment, subjects received either one phone call (One-Shot treatments) or two phone calls (Repeated Interaction treatments) over the period of exactly seven days. Independent of the treatment assignment, however, each subject received two endowments of 6 US$; the first endowment on the day of the first call (which was compulsory for everyone), and the second endowment seven days later. This implies that even subjects assigned to the One-Shot treatments received two endowments and thus two potential payments. This was important in order to keep the payment and donation conditions identical across treatments. The reason I conducted only two and not more phone calls in the condition with repeated interaction was to keep attrition to a minimum and not allow subjects to learn about the nature of interaction. Moreover, I intended to keep my experimental design similar to the experiments by Andreoni and Serra-Garcia (2021) and Dreber et al. (2016).

The second treatment variation affected the timing of the payoffs. I assigned each subject either to a Now or to a Later treatment, where in the Now treatments, subjects were asked to donate immediately (i.e., out of their first endowment) and in the Later treatments subjects were asked to donate with delay (i.e., out of their second endowment). Note that subjects only made one donation decision, which implies that one of the two endowments always remained untouched. Importantly, all donation decisions, both in the Now treatments and in the Later treatments, were made during the first phone call and subjects assigned to the Later treatments had no opportunity to revise their decisions, even if they were called a second time. All these information were given in the beginning of the first phone call, so every participant knew that she would receive two endowments and that her decision during the first phone call was final. As payment method I used the mobile payment system *Venmo*, which allows to make payments in real time. All donations were made via bank transfer on the assigned day for the donation. All donation decisions were realized.

The main part of the experiment took place during the first phone call (which, for one half of the subjects, also happened to be the only phone call). It consisted of five short blocks. Block one included my main task, the donation decision. Being informed about her two endowments, the only decision the subject had to make was to choose how much of her first endowment (Now-Treatments) or her second endowment (Later-Treatments) she would like to donate.

In block two, the participants answered two incentivized belief questions in random order. Question one asked how much a subject believed other participants assigned to the same treatment donated on average.

For question two, the subject was familiarized with the conditions of the opposite delay treatment and asked how much she thought subjects assigned to that treatment had donated on average. To ensure incentive compatibility, every subject was informed that she could earn additional 10 US$ if her responses lied within +/- 25 cents of the actual average amount donated. One of the two responses in this task was chosen for payment with a probability of five percent.

Block three consisted of a time preference task. The experimental technique employed to elicit the time preferences was a multiple price list. Each subject participated in two sets of eight questions in which she had to choose between a smaller earlier payment and a larger later payment. For one set, the earlier payment was an immediate payment (i.e., it was paid out on the same day) and the late payment a payment with a delay of one week. For the other set, the earlier payment had a delay of one week and the later payment a delay of two weeks. The order of sets was randomized. The early option remained constant at a value of 20 US$ throughout the eight questions within each set. The late option increased in value (see screenshot C11 in the Appendix). As in block two, each subject had a five percent chance that one of her 16 decisions in this task was chosen for payment.

In block four, I elicited a subject's risk preferences. The procedure of the task was similar to the task in block three in that I used a multiple price list as elicitation technique. However, the task consisted of only one set of five questions and rather than choosing between two differently dated payments, the subject had to choose between a certain payment and a lottery, which paid 20 US$ with 50% probability and 0 US$ otherwise. While the lottery was consistent throughout the five questions in this block, the certain option decreased in its value (see screenshot C14 in the Appendix). Once again, each subject had a five percent chance that one of her five decisions in this task was chosen for payment.

Block five consisted of a short sociodemographic questionnaire. Note that blocks two to five were identical across treatments. Only for my donation question in block one, subjects were exposed to a treatment variation.

Subjects assigned to treatments with repeated interaction received a second call by one of the interviewers exactly seven days after the first call. This second call only lasted between two and five minutes and did not include any tasks of importance for this study. However, before receiving the second call, subjects did not know anything about the content of that call. They only knew that the second call would not last longer than five minutes and that they would not be able to revise their donation decision from call one.

3.2 Results

Of the 201 subjects who participated in this experiment, 52 were randomly assigned to the OS-Now treatment, 52 were assigned to the OS-Later treatment, 49 were assigned to the RI-Now treatment and 48 were assigned to the RI-Later treatment.[3] Table A5 in the Appendix provides an overview of important summary

[3] One of the subjects assigned to the RI-Later treatment did not return her second phone call. She was therefore excluded from the data set.

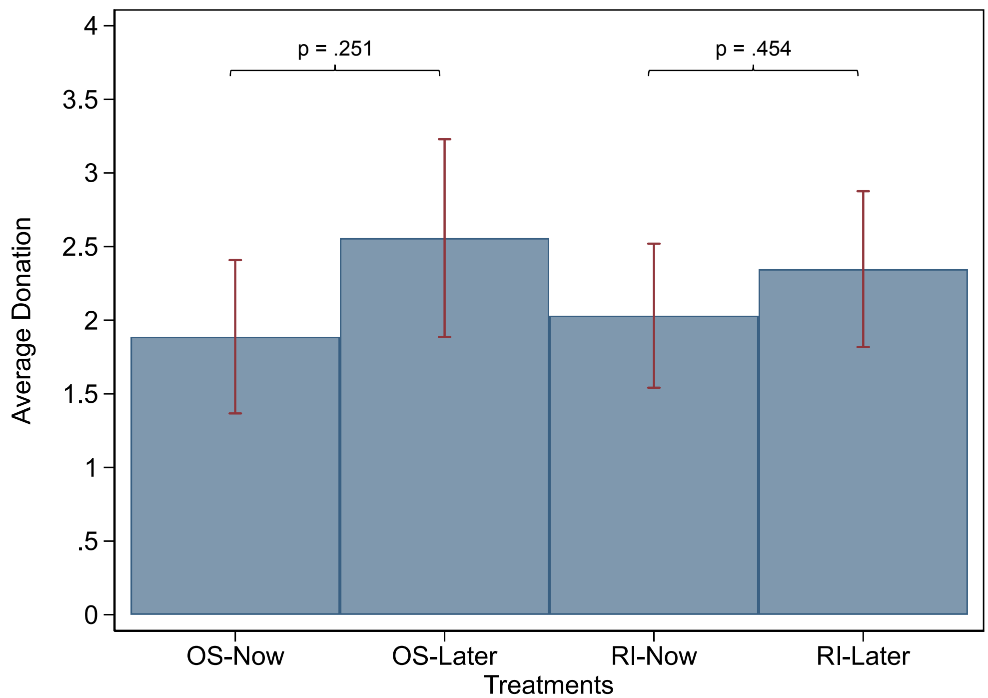


Figure 2: Average Donation by Treatment

Notes: The figure shows the average donations (in US$) in the four treatments. The values above the confidence intervals show p-values of non-parametric Wilcoxon rank-sum tests for the comparisons of means of the Now and Later treatments.

statistics sorted by treatment. For none of the displayed variables, there exist any significant treatment differences. Therefore, I assume that the random assignment of subjects to the treatments was successful.

Figure 2 shows the average donations by treatment.[4] Although it appears that subjects who donated with delay donated more on average, I do not find support for my hypothesis that the delay effect on prosocial choice would be stronger under repeated interaction. For subjects assigned to treatments with repeated interaction, I do not find any significant treatment differences that can be attributed to the delay of the donation ($\mu_{RI-Now} = 2.031$ ($sd = .243$), $\mu_{RI-Later} = 2.347$ ($sd = .263$), $diff = -.316$, Wilcoxon rank-sum test: $p = .454$). In fact, the treatment difference is even smaller than the difference for subjects exposed to the One-Shot treatments, although that difference is not statistically significant either ($\mu_{OS-Now} = 1.88$ ($sd = .259$), $\mu_{OS-Later} = 2.56$ ($sd = .335$), $diff = -.670$, Wilcoxon rank-sum test: $p = .251$). Figure B6 and Figure B7 in the Appendix provide additional insights about the heterogeneity of donations across treatments as they show the distributions of donations and cumulative distribution functions, respectively. However, performing Kolmogorov-Smirnov to tests for differences in the distributions, I do not find that those differ significantly.[5]

[4] Figure B4 and Figure B5 in the Appendix show the probability of giving by treatment and the average donations conditional on giving, respectively.

[5] A Kolmogorov-Smirnov test to compare the distributions of the One-Shot treatments yields $p = 0.417$ and a Kolmogorov-Smirnov test to compare the distributions of the Repeated Interaction treatments yields $p = 0.900$.

Table 4: OLS Regression Results - Donations

	One-Shot		Repeated Interaction		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A. Average donation			
Delay	0.670	0.709	0.316	0.260	0.670	0.716*
	(0.423)	(0.442)	(0.358)	(0.359)	(0.423)	(0.427)
Repeated Interaction					0.143	0.075
					(0.356)	(0.383)
Delay × Rep. Interaction					-0.354	-0.325
					(0.555)	(0.560)
Constant	1.887***	2.300	2.031***	0.975	1.887***	2.571
	(0.259)	(2.214)	(0.243)	(2.987)	(0.259)	(1.623)
Observations	104	104	97	97	201	201
R-squared	0.02	0.13	0.01	0.11	0.02	0.09
			Panel B. Probability of giving			
Delay	0.019	0.067	0.037	0.042	0.019	0.040
	(0.091)	(0.099)	(0.083)	(0.086)	(0.091)	(0.094)
Repeated Interaction					0.083	0.075
					(0.088)	(0.096)
Delay × Rep. Interaction					0.018	0.002
					(0.123)	(0.129)
Constant	0.692***	0.463	0.776***	1.411**	0.692***	1.045**
	(0.065)	(0.502)	(0.060)	(0.683)	(0.065)	(0.414)
Observations	104	104	97	97	201	201
R-squared	0.00	0.09	0.00	0.11	0.01	0.07
			Panel C. Average donation conditional on giving			
Delay	0.868*	0.686	0.270	0.055	0.868*	0.793*
	(0.442)	(0.498)	(0.349)	(0.367)	(0.442)	(0.468)
Repeated Interaction					-0.108	-0.191
					(0.366)	(0.406)
Delay × Rep. Interaction					-0.598	-0.481
					(0.563)	(0.624)
Constant	2.726***	4.232	2.618***	-1.681	2.726***	2.394
	(0.276)	(2.559)	(0.240)	(2.688)	(0.276)	(1.816)
Observations	73	73	77	77	150	150
R-squared	0.05	0.20	0.01	0.13	0.05	0.12
Demographics		✓		✓		✓
Interviewer FE		✓		✓		✓

Notes: OLS regressions for donations in US$ (Panel A), probability of giving (Panel B) and average donations in US$ conditional on giving (Panel C). Columns (1) and (2) show the results for subjects assigned to the One-Shot treatments. Columns (3) and (4) show the results for subjects assigned to the Repeated Interaction treatments. Columns (5) and (6) show the results for the full sample. Demographic controls include the gender of the subjects, age, nationality, a dummy, which indicates if the subject is a student or not, the student's monthly expenditure on food, transport, leisure and activities and two dummies which indicate whether the subject has previously volunteered at charitable organizations and whether she has donated money within the previous year. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Next, I turn to my regression analysis, which allows me to test the treatment effects while controlling for sociodemographic variables and fixed effects induced by the telephone interviewer. Table 4 shows the results of six OLS regressions for three constructs of the dependent variable capturing prosocial choice. In columns (1) to (2) and (3) to (4) the sample is divided by frequency of interaction. Columns (5) and (6) show regression results for the full sample. In Panel A, the dependent variable of the regression is the unconditional average donation. In Panel B, I focus on the probability of giving, that is, whether a subject donated some positive amount or not. Finally, in Panel C, the dependent variable is the average donation conditional on giving.

All regression results support the evidence that under repeated interaction donations do not increase when they are made with delay. Although the direction of the delay effect is consistent with the hypothesis derived from the meta-study, independent of the model specification and the form of the dependent variable, none of the results from the regressions (3) and (4) suggest that the treatment effect could be statistically significant. The regression estimates are also economically small ranging from 26 to 32 Cents (0.13 to 0.16 sd) for the unconditional donations and from 6 to 27 Cents (0.03 to 0.14 sd) for the conditional donations. If anything, the delay effect on average donations (conditional and unconditional) appears stronger for subjects assigned to the One-Shot treatments (regressions (1) and (2)), although that effect remains mostly non-significant, too. For the probability of giving the estimates suggest a delay-induced increase of about four percentage points (pp) under repeated interaction. Again, this is surprisingly low compared to the 15pp increase found in the first experiment by Andreoni and Serra-Garcia (2021).

Focusing on the regressions results from the full sample (regressions (5) and (6)), the evidence regarding the delay effect remains unchanged. A positive general delay effect of about 67 to 72 Cents (approximately 0.34 to 0.36 sd) is only marginally significant at best and it is only visible when focusing on the average donations (Panel A and Panel C). Importantly, however, the interaction term, which contains the relevant information with regard to the relation between delay and repeated interaction, does not support my hypothesis. Considering the results of Panel A and Panel C, it appears that the regression estimates of the interaction term are not only non-significant, they are also negative implying a decrease of delayed donations under repeated interaction (relative to delayed donations in a one-shot setting). For the probability of giving (Panel B) the estimate of the interaction is virtually zero indicating no difference between the delay effect under one-time and repeated interactions.

To demonstrate the robustness of these results, I provide regression results of nine Tobit regressions in Table A6 in the Appendix. The exercise is similar to the OLS analysis conducted in this section for the unconditional average donation (Panel A), but the Tobit regressions allow to censor the regression models at 0 US$ and 6 US$. The estimates of the Tobit regressions suggest that the effect of the delay treatment is slightly larger than that estimated with OLS. However, the results are consistent with the OLS models in that they reveal that the delay effect under repeated interaction is not statistically significant.

Finally, I shed light on the results of my secondary tasks, i.e., the incentivized belief task, the time and risk preference tasks. Those tasks had been conducted to investigate interesting relations between the hypothesized treatment effect and belief or preference measures. Not surprisingly, both belief measures, i.e., for the subject's own Payment Date treatment and the alternative Payment Date treatment, are highly correlated with her own donation. The correlation between the expected average donation in one's own treatment and the own donation is $\rho = 0.664$ ($p = 0.000$) and the correlation between the expected average donation in the alternative Payment Date treatment and the own donation is $\rho = 0.550$ ($p = 0.000$). More interesting are the belief differences induced by the delay and the interaction treatments. Recall, that after the main task, each subject was asked what she believed subjects donated on average in her own and in the alternative Payment Date condition. Figure B8 in the Appendix depicts the average beliefs by treatment. Interestingly, in both Later treatments, subjects seemed to expect higher donations in the Later conditions than in the Now conditions, while in the two Now treatments, these belief differences cannot be observed. That is, it appears the delay effect was successful in inducing different beliefs about the donation activity across the two Payment Date conditions. I do not, however, observe such differences for the two interaction conditions.

To study the relationship between preference parameters and the donations in this study, I conducted additional OLS regressions. The elicited parameters indicate the degree of present bias (*Beta*), standard time discounting (*Delta*) and risk aversion. All parameters are standardized, and thus, can be interpreted in sd. However, recall that the preference parameters were elicited after the main task. For that reason, they might not be exogenous to my treatment interventions and thus should not be treated as classic control variables. The regression results are presented in Table 5.

I do not find that any of the preference parameters correlate with the donations. Neither the time preference parameters *Beta* and *Delta* nor *risk aversion* seem to be predictive for subjects' donations. Moreover, the results do not suggest that there exists any significant interaction effect between the preference parameters and the treatment variable *delay*.

4 Discussion and Concluding Remarks

This paper presents a meta-analysis documenting substantial contradicting evidence on the effect of time delay on prosocial choice in 11 independent experimental studies. While the evidence cannot be explained by different delay lengths, my analysis suggests that the heterogeneity in the delay effect appears reconcilable by a moderating factor. Analyzing a set of different experimental design features, my meta-study identifies *repeated interaction* as the most promising candidate to evoke a positive delay effect on prosocial behavior. To further test this empirical observation, I conducted a donation experiment implemented via telephone calls in which I varied both the date of the donation and the frequency of interaction. The results of this experiment, however, do not support the evidence from the meta-study. While I do find a weakly significant general delay effect on prosocial choice – an increase in donations by about 0.34 to 0.36 sd – I do not find that this effect is moderated by repeated interaction. If anything, the delay-induced increase in donations is stronger in the treatments with one-time interaction than in those with repeated interaction.

Table 5: Donations and Preference Parameters

	Dependent Variable: Donations			
	Full Sample			
	(1)	(2)	(3)	(4)
Beta	-0.104	0.062	-0.087	-0.061
	(0.134)	(0.167)	(0.135)	(0.138)
Delta	0.018	0.030	-0.057	0.060
	(0.176)	(0.177)	(0.217)	(0.181)
Risk Aversion	0.005	-0.008		-0.183
	(0.150)	(0.150)		(0.206)
Delay		0.455	0.444	0.454
		(0.283)	(0.283)	(0.283)
Delay × Beta		-0.294		
		(0.227)		
Delay × Delta			0.173	
			(0.310)	
Delay × Risk Aversion				0.371
				(0.303)
Constant	2.225***	1.991***	1.981***	1.997***
	(0.141)	(0.182)	(0.388)	(0.182)
Observations	196	196	196	196
R-squared	0.00	0.02	0.02	0.03

Notes: OLS regressions for donations using elicited preference parameters as independent variables. All preference parameters are standardized (z-scores), so units are in sd. The variable *Delay* captures the delay of payments (in days) in the treatment relative to the control group. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

One potential concern with my study might be the validity of the meta-analysis, which comprises 11 studies and 14 experiments. Such an analysis might lack the necessary power to provide conclusive evidence. However, the goal of the meta-analysis was not to provide conclusive summarizing evidence for the literature that it covers. Rather, it sought to a) reveal and visualize the inconsistencies across studies and b) identify potential design features that could explain the heterogeneous state of evidence. Despite its low power, the analysis was successful in doing so.

The main tool of analysis to test the prediction derived from my meta-study was my experiment. It resembled the situation of a telephone fundraiser. It was intentionally kept similar to a number of experiments included in the meta-study. In particular, it closely resembled the first (motivational) experiment conducted by Andreoni and Serra-Garcia (2021). The only relevant differences of my experiment were the introduction of continuous choice (the decisions in Andreoni and Serra-Garcia (2021) were binary) and the implementation of the experiment by means of phone calls (Andreoni and Serra-Garcia (2021) conducted a lab experiment). To design my study and determine the number of participants to invite for each treatment, I therefore relied on the effect size of their experiment's treatment effect ($d = 0.65$). I conducted a pre-experimental power simulation and found that a sample size of 50 subjects per treatment would yield more than 90% power to detect a significant effect of the delay treatment for the treatments in which participants interact repeatedly. Even with an effect size of $d = 0.5$, or treatment sizes slightly below 50 subjects, the remaining power would have satisfied the standard criterion of 80%. Using my own experimental data after the execution of my experiment, I can confirm that these pre-experimental simulations were accurate. Given the sample size and the observed variance in the four treatments, I have 80% power to detect an effect size of $d = 0.45$, which is well below the effect size in Andreoni and Serra-Garcia (2021). For that reason, I do not believe that the failure of detecting a significant delay effect for subjects exposed to repeated interaction is a result of a lack of power. The treatment difference induced by the delay manipulation in the treatments with repeated interaction is not even close to passing standard significance tests. Moreover, I do not find any evidence that the delay manipulation in the One-Shot treatments triggers a donation behavior that is different from that in the Repeated Interaction treatments. I acknowledge that my experimental design with only two interactions between the experimenter and the subject could be considered the weakest form of repeated interaction. While I cannot exclude that a stronger manipulation (i.e., more than two phone calls) would have increased donations, given the similarities of my experimental design and those in other studies, I do not think that the intensity of repeated interaction explains my findings.

The goal of this paper was to reveal the conflicting evidence of 11 closely related studies and to identify and test the role of a design feature that could potentially link this evidence. After my analysis, however, two questions remain: I. Why do the mechanisms, which caused an increase in donations in the experiment by Andreoni and Serra-Garcia (2021), fail to be effective in my treatments with repeated interaction? II. Why is the theoretical prediction of the extended dual-self model provided by Dreber et al. (2016) to explain a delay-induced decrease of giving not valid for the results in my One-Shot treatments? Since the attempt of my study failed to provide an explanation for the discrepancy of the existing evidence, it remains a task for future research to solve the inconsistency puzzle and provide robustness for the suggested explanations. One possible avenue to move forward could be to analyze the motives underlying prosocial choice more carefully, that is, study how those motives differ depending on the decision context and how they are affected by time.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153.

Adena, M., & Huck, S. (2019). Giving once, giving twice: A two-period field experiment on intertemporal crowding in charitable giving. *Journal of Public Economics*, *172*, 127–134.

Albrecht, K., Volz, K. G., Sutter, M., Laibson, D. I., & Von Cramon, D. Y. (2011). What is for me is not for you: Brain correlates of intertemporal choice for self and other. *Social cognitive and affective neuroscience*, *6*(2), 218–225.

Andreoni, J., & Serra-Garcia, M. (2021). Time inconsistent charitable giving. *Journal of Public Economics*, *198*, 104391.

Apffelstaedt, A., & Pieters, A. (2019). Time-inconsistent behavior in social dilemma situations: The role of impure altruism. *Unpublished manuscript.*

Brañas-Garza, P., Jorrat, D., Kovářık, J., & López, M. C. (2021). Hyper-altruistic behavior vanishes with high stakes. *PloS one*, *16*(8), e0255668.

Breman, A. (2011). Give more tomorrow: Two field experiments on altruism and intertemporal choice. *Journal of Public Economics*, *95*(11-12), 1349–1357.

Capra, C. M., Jiang, B., & Su, Y. (2019). Volunteer now or later: The effects of effort time allocation on donations. *Bing and Su, Yuxin, Volunteer Now or Later: The Effects of Effort Time Allocation on Donations (August 2019).*

Chopra, F., Eisenhauer, P., Falk, A., & Graeber, T. (2021). Intertemporal altruism.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences new york. *NY: Academic*, 54.

Dreber, A., Fudenberg, D., Levine, D. K., & Rand, D. G. (2016). Self-control, social preferences and the effect of delayed payments. *Social Preferences and the Effect of Delayed Payments (February 17, 2016).*

Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the finitely repeated prisoner's dilemma. *The Quarterly Journal of Economics*, *133*(1), 509–551.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*, *83*(2), 587–628.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351–401.

Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *American economic review*, *96*(5), 1449–1476.

Howard, G. (2013). Discounting for personal and social payments: Patience for others, impatience for ourselves. *Journal of Environmental Economics and Management*, *66*(3), 583–597.

Islam, M. (2016). Public goods and intertemporal choice: Evidence from a public goods game. *Unpublished manuscript.*

Kim, J. (2019). *The effects of time preferences on cooperation: Experimental evidence from infinitely repeated games* (tech. rep.). Mimeo.

Kölle, F., & Lauer, T. (2019). Cooperation, discounting, and the effects of delayed costs and benefits. *Available at SSRN 3260626*.

Kölle, F., & Wenner, L. (2021). Is generosity time-inconsistent? present bias across individual and social contexts. *The Review of Economics and Statistics*, 1–45.

Kovarik, J. (2009). Giving it now or later: Altruism and discounting. *Economics Letters*, *102*(3), 152–154.

Noor, J., & Ren, L. (2011). Temptation and social preference. *Unpublished manuscript*.

Rong, R., Gnagey, M., & Grijalva, T. (2018). the less you discount, the more it shows you really care": Interpersonal discounting in households. *Journal of Economic Behavior & Organization*, *154*, 1–23.

Rong, R., Grijalva, T. C., Lusk, J., & Shaw, W. D. (2019). Interpersonal discounting. *Journal of Risk and Uncertainty*, *58*(1), 17–42.

Shapiro, J. (2020). *Giving now and later: Discounting of altruistic and warm glow utility* (tech. rep.). Working Paper.

Shapiro, J. (2010). *Discounting for you, me and we* (tech. rep.). Working paper.

Appendices

A Additional Tables

Table A1: Overview of Key Design Features for Studies included in Meta Analysis

Study	Repeated interaction	Lab experiment	Within-subject	Strategic game	Random payment	Real effort	Continuous choice
Andreoni & Serra-Garcia 2019	yes	yes	yes/no	no	no	no	no
Apffelstädt & Pieters 2019	no	no	no	yes	no	no	yes
Breman 2011	no	no	no	no	no	no	yes
Capra et al. 2019	yes	no	yes	no	no	yes	no
Dreber et al. 2016	no	no	no	no	no	no	yes
Islam 2016	no	yes	no	yes	yes	no	yes
Kim 2019	yes	yes	no	yes	no	no	no
Kölle & Lauer 2018	no	yes	no	yes	yes	no	yes
Kölle & Wenner 2018	yes	yes	yes	no	yes	yes	yes
Kovarik 2011	no	yes	no	no	no	no	yes
Noor & Ren 2011	yes	yes	no	no	no	no	yes

Table A2: Meta Analysis - Alternative Design Features

	Dependent Variable: Effect Size (in sd)					
	Meta-Regression					
	(1)	(2)	(3)	(4)	(5)	(6)
Lab Experiment	0.261*					
	(0.146)					
Within-Subject		0.287**				
		(0.127)				
Random Pay			-0.069			
			(0.151)			
Strategic Game				-0.008		
				(0.203)		
Real Effort					-0.128	
					(0.153)	
Continuous Choice						-0.537***
						(0.099)
Constant	0.073	0.132	0.285***	0.266***	0.300***	0.627***
	(0.125)	(0.086)	(0.083)	(0.075)	(0.081)	(0.083)
Observations	28	28	28	28	28	28
Adj. R-squared	0.162	0.160	-0.044	-0.053	-0.021	0.707

Notes: Meta regressions for the effect sizes of all treatments included in the meta-study. The effect size quantifies the treatment effect of the delay manipulation relative to the control condition, which did not include any delay. The measure of the effect size is *Cohen's d*. All units are in standard deviations. Standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Meta Analysis - Robustness Results for Delay Effect

	Dependent Variable: Effect Size (in sd)							
	Meta-Reg				OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln_Delay	-0.086		-0.003		-0.071		-0.023	
	(0.062)		(0.071)		(0.067)		(0.067)	
$Delay^2$		-0.000		-0.000		-0.000		-0.000
		(0.000)		(0.000)		(0.000)		(0.000)
Repeated Interaction			0.652**	0.649**			0.646**	0.622**
			(0.310)	(0.303)			(0.236)	(0.223)
Lab Experiment			-0.044	-0.031			-0.087	-0.049
			(0.141)	(0.141)			(0.142)	(0.143)
Within-Subject			0.231	0.232			0.346*	0.353*
			(0.177)	(0.177)			(0.187)	(0.189)
Strategic Game			0.132	0.154			0.196	0.210
			(0.189)	(0.210)			(0.201)	(0.231)
Random Pay			-0.093	-0.118			-0.014	-0.052
			(0.215)	(0.249)			(0.203)	(0.247)
Real Effort			-0.609**	-0.591**			-0.784***	-0.723**
			(0.258)	(0.242)			(0.271)	(0.268)
Continuous Choice			0.094	0.100			0.139	0.116
			(0.269)	(0.247)			(0.240)	(0.217)
Constant	0.489**	0.273***	-0.060	-0.078	0.450**	0.275***	-0.028	-0.092
	(0.176)	(0.071)	(0.283)	(0.257)	(0.185)	(0.073)	(0.251)	(0.243)
Observations	28	28	28	28	28	28	28	28
Adj. R-squared	0.08	-0.04	0.78	0.78	0.01	-0.03	0.63	0.63

Notes: Meta and OLS regressions for different delay specifications. Model (1), (3) (5) and (7) include the delay variable in logarithmic form. The remaining models include the delay variable in quadratic form. The effect size quantifies the treatment effect of the delay manipulation relative to the control condition, which did not include any delay. The measure of the effect size is *Cohen's d*. All units are in standard deviations. Standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Meta Analysis – Game Dummies

	Dependent Variable: Effect Size (in sd)	
	Meta-Regression	OLS
	(1)	(2)
Repeated Interaction	0.410***	0.381***
	(0.100)	(0.107)
Delay	-0.001	-0.001
	(0.001)	(0.001)
Dictator Game	-0.362***	-0.405***
	(0.096)	(0.105)
Public Goods Game	0.104	0.070
	(0.205)	(0.199)
Prisoner's Dilemma'	-0.173	-0.224
	(0.252)	(0.257)
Constant	0.191*	0.276**
	(0.095)	(0.111)
Observations	28	28
Adj. R-squared	0.69	0.57

Notes: Meta and OLS regressions using different game dummies as control variables. The excluded dummy and benchmark is the variable *Charity*. All units are in standard deviations. Standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Summary Statistics by Treatment

	One-Shot - Now					One-Shot - Later				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Female	52	.615	.491	0	1	52	.615	.491	0	1
Age	52	21.26	2.44	18	30	52	21.75	3.76	18	41
Student	52	.981	.139	0	1	52	.885	.323	0	1
Volunteer	52	.519	.505	0	1	52	.519	.505	0	1
Donated Past	52	.5	.505	0	1	52	.462	.503	0	1
Beta	52	1.041	.118	.706	1.416	52	1.063	.121	.891	1.5
Delta	52	.786	.174	0	.976	52	.781	.131	.494	.976
	Repeated Interaction - Now					Repeated Interaction - Later				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Female	49	.673	.474	0	1	48	.646	.483	0	1
Age	49	21.02	1.98	18	28	48	21.19	2.18	18	29
Student	49	.918	.277	0	1	48	.938	.245	0	1
Volunteer	49	.490	.505	0	1	48	.563	.501	0	1
Donated Past	49	.612	.492	0	1	48	.646	.483	0	1
Beta	49	1.059	.102	.899	1.313	48	1.01	.099	.75	1.357
Delta	49	.780	.173	0	.976	48	.803	.145	.488	.976

Table A6: Tobit Regression Results – Unconditional Average Donations

	Dependent Variable: Average Donations								
	One-Shot			Repeated Interaction			Full Sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Donation Delay	1.133	1.470*	1.225	0.430	0.322	0.340	1.020	1.166*	1.049
	(0.792)	(0.833)	(0.810)	(0.496)	(0.471)	(0.482)	(0.695)	(0.701)	(0.689)
Repeated Interaction							0.278	0.284	0.164
							(0.570)	(0.607)	(0.613)
Delay × Rep. Interaction							-0.548	-0.718	-0.492
							(0.881)	(0.888)	(0.886)
Constant	1.271**	2.480	2.389	1.783***	1.278	1.297	1.417***	3.398	3.368
	(0.493)	(4.555)	(4.157)	(0.340)	(4.140)	(4.141)	(0.430)	(2.872)	(2.772)
Observations	104	104	104	97	97	97	201	201	201
Pseudo R-squared	0.005	0.031	0.040	0.002	0.033	0.033	0.004	0.022	0.025
Demographics		✓	✓		✓	✓		✓	✓
Interviewer FE			✓			✓			✓

Notes: Tobit regressions for donations censored at 0US$ and 6US$. Columns (1) to (3) show the results for subjects assigned to the One-Shot treatments. Columns (4) to (6) show the results for subjects assigned to the Repeated Interaction treatments. Columns (7) to (9) show the results for the full sample. Demographic controls include the gender of the subjects, age, nationality, a dummy, which indicates if the subject is a student or not, the student's monthly expenditure on food, transport, leisure and activities and two dummies which indicate whether the subject has previously volunteered at charitable organizations and whether she has donated money within the previous year. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Additional Figures

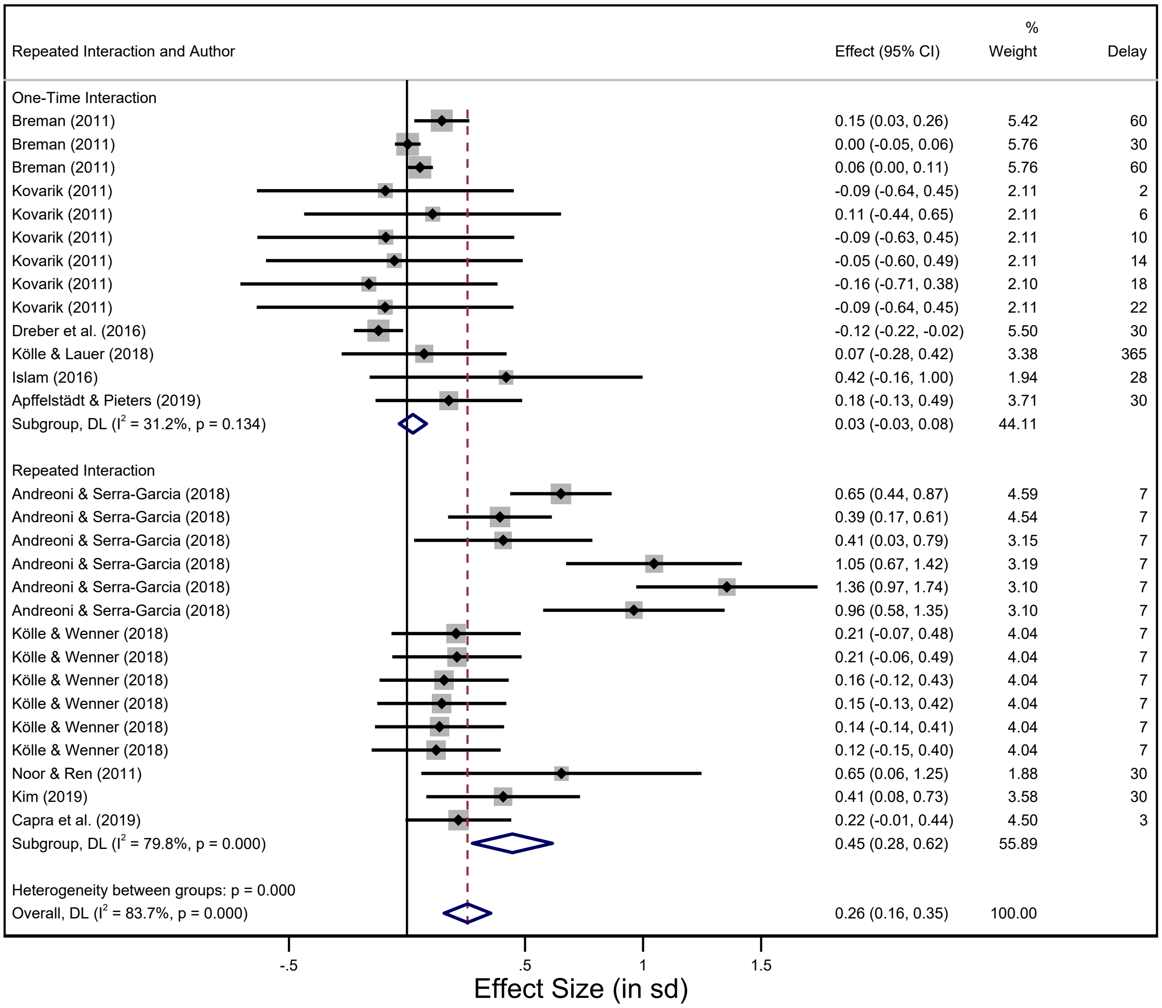


Figure B1: Effect Sizes by Interaction

Notes: The figure shows the effect sizes and the respective confidence intervals for different treatment condition by interaction. Each data point represents one treatment condition. The treatment effects are measured relative to the respective control condition (i.e., conditions with no delay) of the experiment. The weights are from a random effects model and reflect the confidence level of the results.

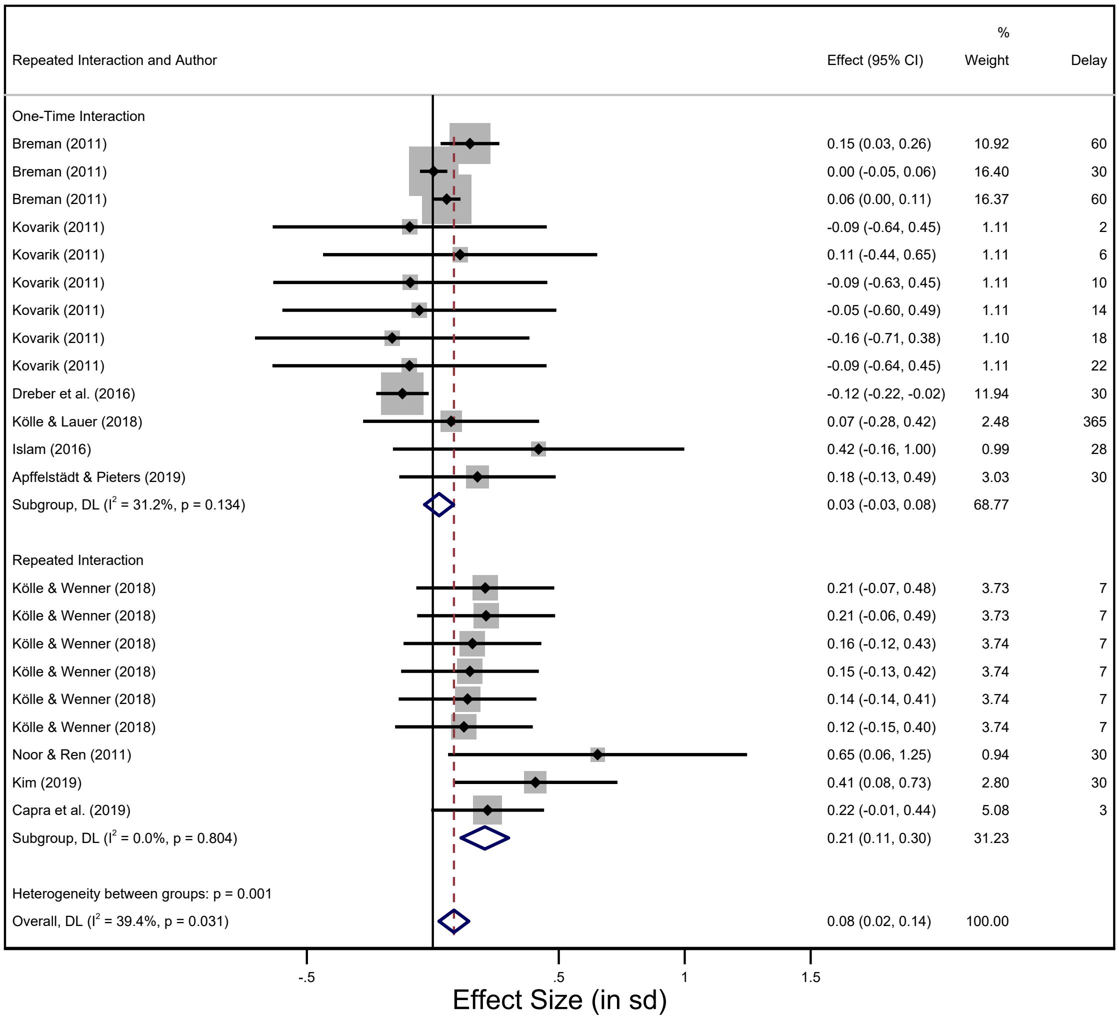


Figure B2: Effect Sizes by Interaction Excluding Andreoni and Serra-Garcia (2021)

Notes: The figure shows the effect sizes and the respective confidence intervals for different treatment condition by interaction. Each data point represents one treatment condition. The treatment effects are measured relative to the respective control condition (i.e., conditions with no delay) of the experiment. The weights are from a random effects model and reflect the confidence level of the results.

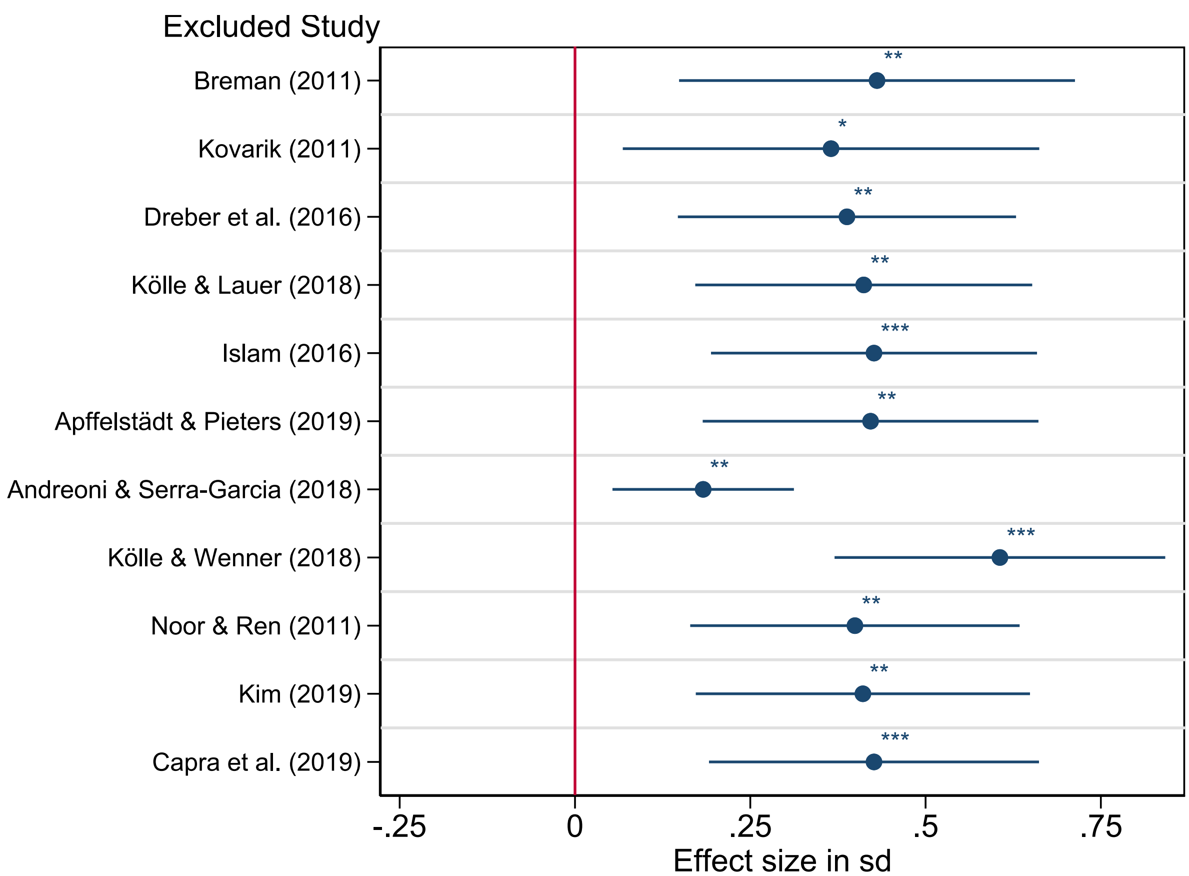


Figure B3: Regression Estimates for Repeated Interaction – Robustness Analysis

Notes: The figure shows the estimates for the dependent variable Repeated Interaction for 11 meta-regressions. For each regression, one study was excluded from the analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

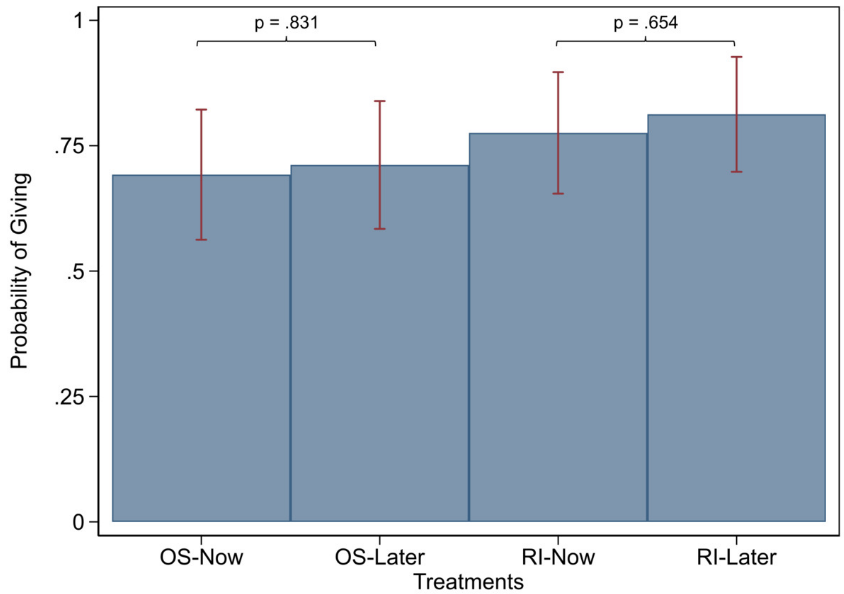


Figure B4: Probability of Giving by Treatment

Notes: The figure shows the probability of giving (i.e. Donations > 0) by treatment. The values above the confidence intervals show p-values of non-parametric Wilcoxon rank-sum tests for the comparisons of means of the Now and Later treatments

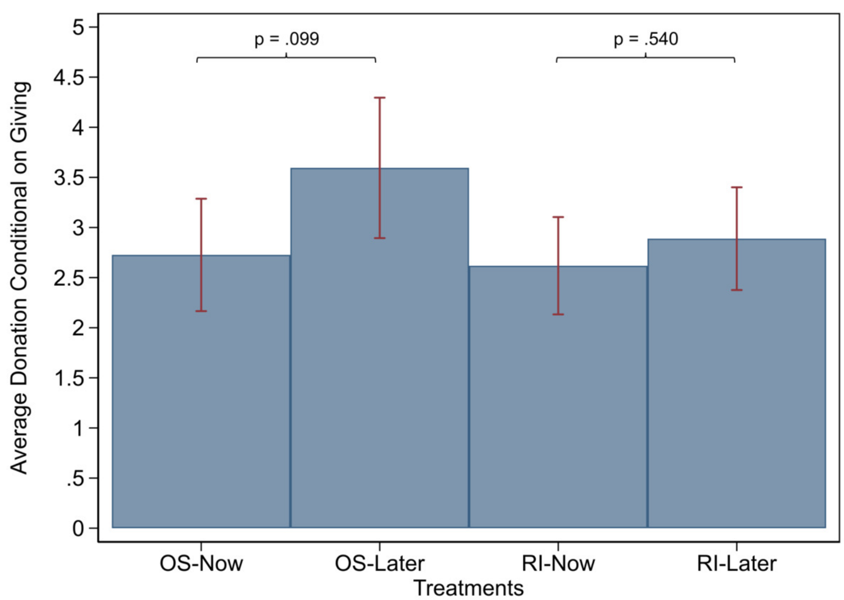


Figure B5: Average Donation Conditional on Giving

Notes: The figure shows the average donation (in US$) conditional on giving (i.e. Donations > 0) by treatment. The values above the confidence intervals show p-values of non-parametric Wilcoxon rank-sum tests for the comparisons of means of the Now and Later treatments

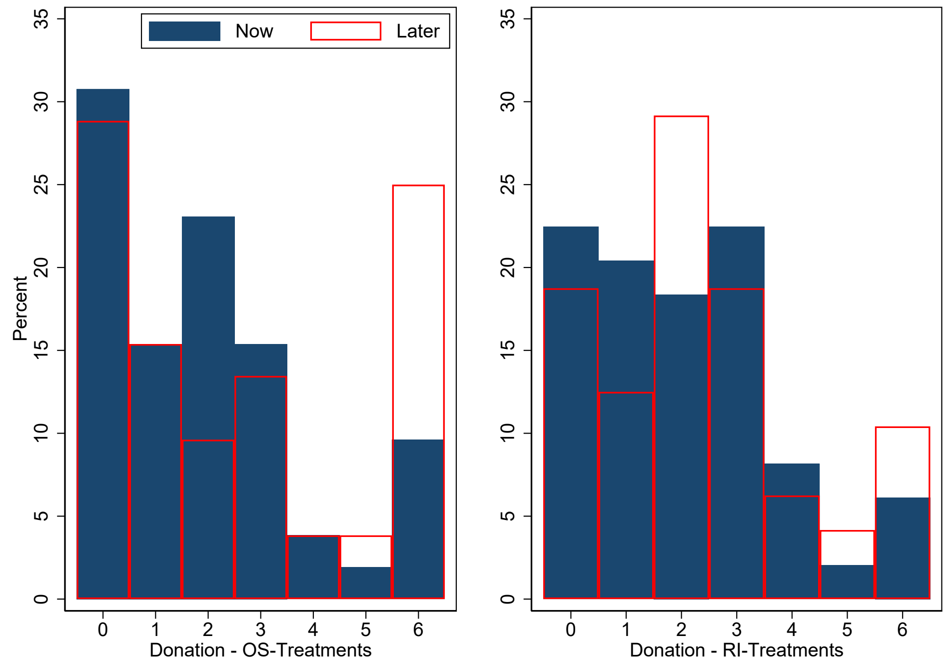


Figure B6: Distribution of Donation by Treatment

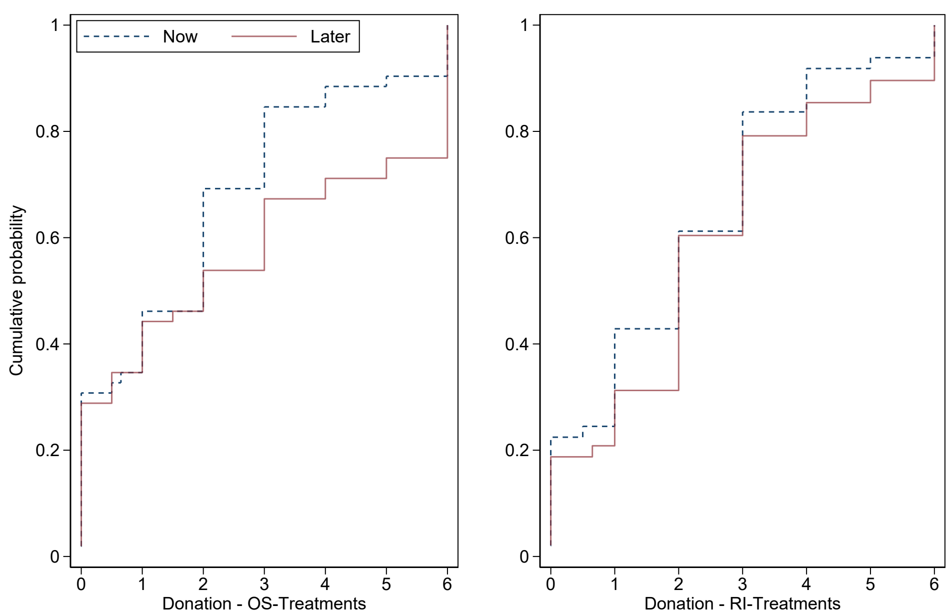


Figure B7: Cumulative Distribution of Donation by Treatment

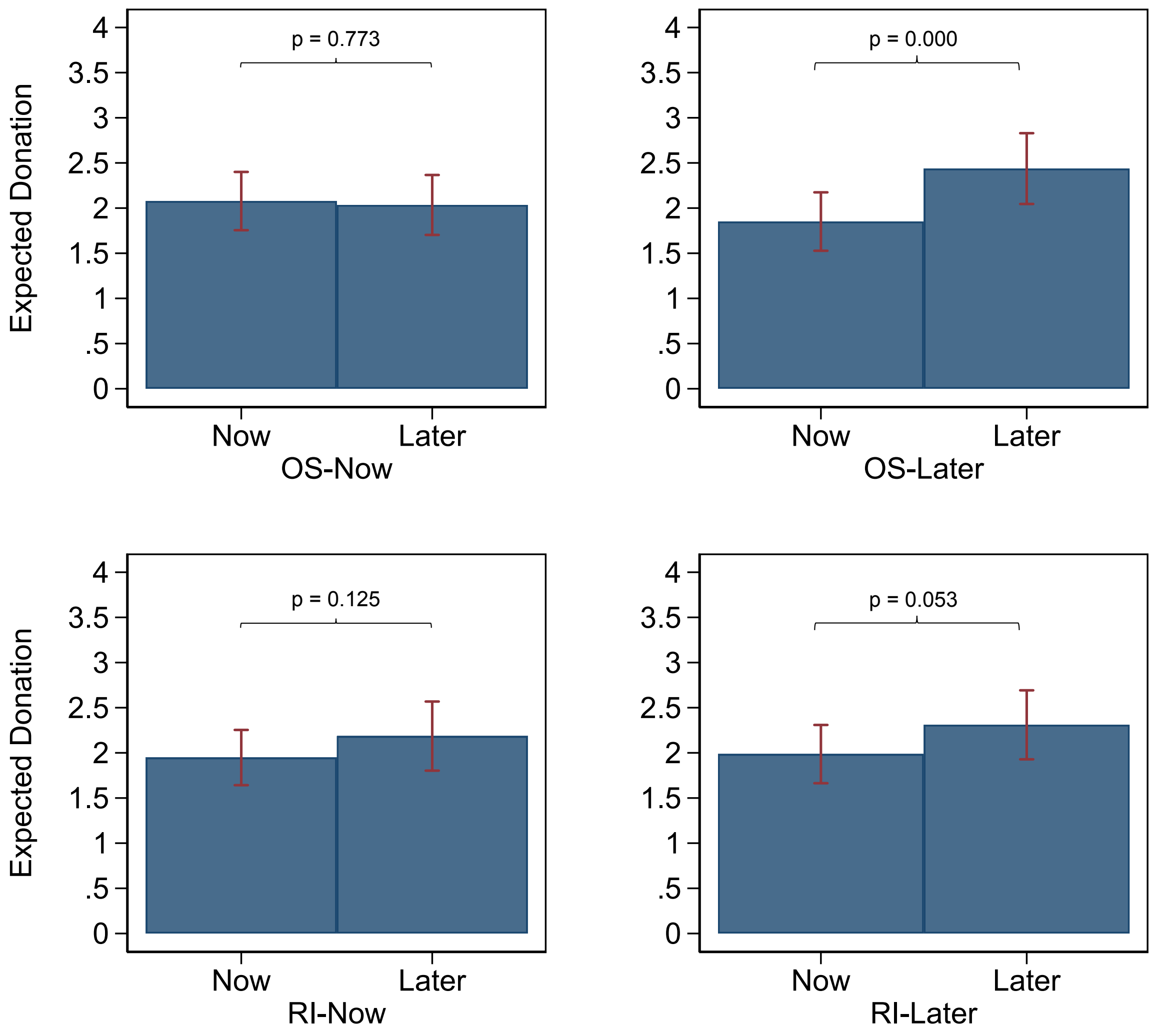


Figure B8: Beliefs by Treatment

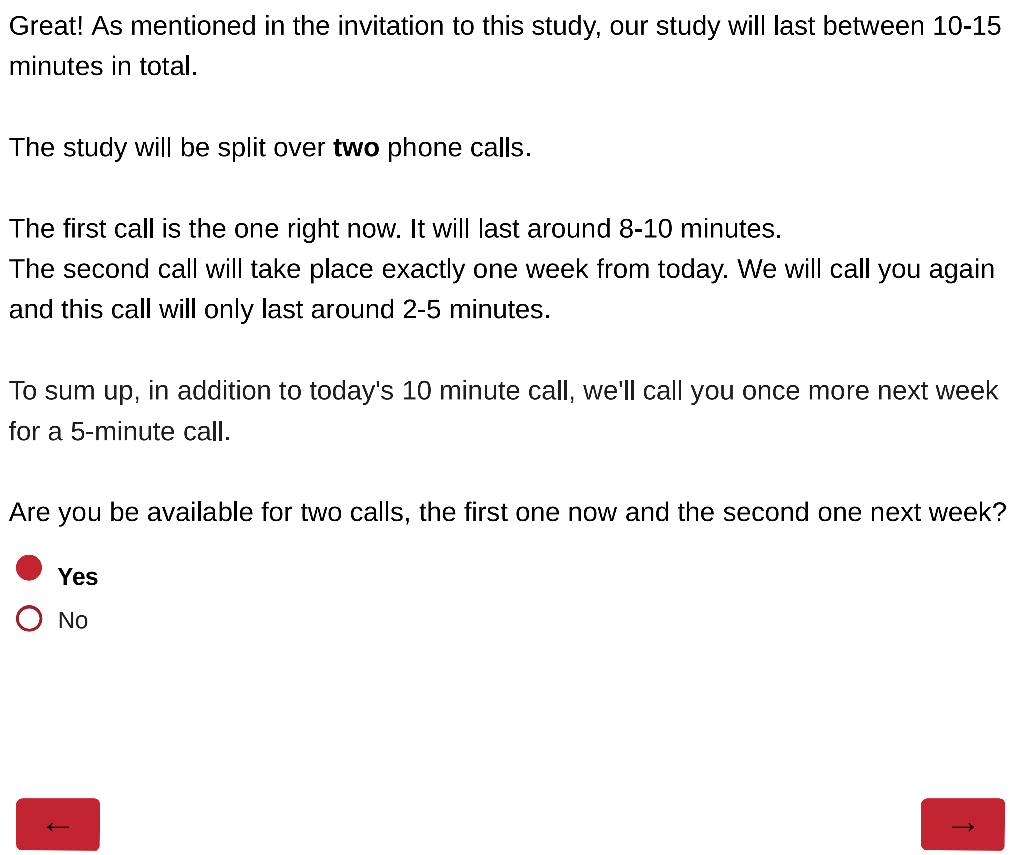
C Experimental Script for Telephone Interviewers

Welcome, my name is *[name],* I work at the Rady Incentives lab at the Rady School.Thanks for allowing us to call you for our study. We emailed you a short consent form and I hope you had some time to look at it. In order to proceed, we have to know whether you agree with the conditions of that form. So, please let us know if you do or if you don't.

- ○ Agree
- ○ Don't agree

← →

Figure C1: Script Page 1



Great! As mentioned in the invitation to this study, our study will last between 10-15 minutes in total.

The study will be split over **two** phone calls.

The first call is the one right now. It will last around 8-10 minutes.
The second call will take place exactly one week from today. We will call you again and this call will only last around 2-5 minutes.

To sum up, in addition to today's 10 minute call, we'll call you once more next week for a 5-minute call.

Are you be available for two calls, the first one now and the second one next week?

- ● **Yes**
- ○ No

← →

Figure C2: Script Page 2

Perfect! Then, let me briefly introduce you to our procedure.

Our call today consists of three parts and a quick survey. In each part, you will be asked to answer some questions or make some choices. All parts will be conducted over the phone and you will have the opportunity to earn money.

The purpose of our call next week is to conduct a follow-up survey.

It is very important for us that you participate in both phone calls, today and next week, so that we can complete our study.

At the end of this call today, we will ask you when exactly we should call you next week, but please keep it in mind already now and make sure you are available.

For participating in this study you will receive $12, which will be split in two equal payments.

The first $6, you will receive today, right after this phone call.
The second $6 you will receive after our second call, exactly one week from today, on Tuesday, September 7th.

Please note that all payments will be made via *Venmo*. The only requirement is that you answer all questions in all parts of the experiment. As a test, I will Venmo you $0.01 already now. Please give me a few seconds to do that. OK, I just did that. Can you see that you received the payment?

○ Yes
○ No

← →

Figure C3: Script Page 3

Great, now that *Venmo* seems to work fine, let's start with part 1 of our call today.

←   →

Figure C4: Script Page 4

We would now like to ask you whether you would like to donate some amount out of the $6 payment **you will receive next week** to GiveDirectly? You can give any amount you like. If you donate, your donation will be deducted from your $6 payment **next week** after our second call. Otherwise, you will keep the full $6 and no donation will be made.

Please note that your donation decision will be final today.

So, would you like to donate any amount out of your $6 payment **next week**? If so, how much?

________ $

← →

Figure C5: Script Page 5

We would now like to ask you whether you would like to donate some amount out of your $6 payment **you will receive next week** to GiveDirectly? You can give any amount you like. If you donate, your donation will be deducted from your $6 payment **next week**. Otherwise, you will keep the full $6 and no donation will be made.
Please note that your donation decision will be final today.

So, would you like to donate any amount out of your $6 payment **next week**? If so, how much?

$

← →

Figure C6: Script Page 6

In this part, we will ask you two questions. For your answers, you may receive an additional payment.

Specifically, one out of 20 participants will be selected at random. One of the two questions below will also be selected at random and used to determine the additional payment, as we describe below.

So, make your decisions carefully as these could significantly affect your payment.

← →

Figure C7: Script Page 7

Like you, other participants in this study, were asked to donate out of the $6 payment they will receive next week. What do you think, how much did other study participants donate from the $6 payment they will receive next week on average?

If you are selected for payment and this question is drawn for payment, you will receive $10 if your estimate lies within +/- 25 cents of the actual average amount donated.

→

Figure C8: Script Page 8

A different group of participants in this study was asked to donate out of the $6 payment they receive today. What do you think, how much did those participants donate from the $6 payment they receive today on average?

If you are selected for payment and this question is drawn for payment, you will receive $10 if your estimate lies within +/- 25 cents of the actual average amount donated by those participants.

Figure C9: Script Page 9

Thanks, this was part 2.

In the next part, you will have to choose between two options. First, you will be asked to choose between receiving a payment at an earlier date or a payment at a later date. The payment at the early date is the same in each of these situations. The payment at the late date is different in every situation. For each of these situations, we would like to know which of the payments – early or late – you would like to choose.
Let's go through an example:

Do you prefer $30 today or $35 in 1 month from today?

[Wait for answer]

Great! It's as easy as that. As you can see, our questions are not designed to test you – there are no "correct" or "incorrect" answers.

We will present to you 16 questions like this one. Then, we will ask you about 5 situations that involve money today paid via a lottery or not.

Now let me explain the payment. One out of 20 participants will be selected at random and will receive payment based on his/her decision in the question that is chosen for payment. The question that counts is determined randomly, so please think thoroughly because every answer could be the one that counts. In the example question, you chose to receive [repeat answer that was given], which means, if that question were selected for payment, you would receive additional payment of [repeat answer that was given].

Do you have any questions?

[Wait for answer]

Then, let's start with the task.

→

Figure C10: Script Page 10

Would you rather like to receive $20 in one month from today or $20.50 in two months from today?

$20 in one month	$20.50 in two months
○	○

Would you rather like to receive $20 in one month from today or $21.50 in two months from today?

$20 in one month	$21.50 in two months
○	○

Would you rather like to receive $20 in one month from today or $23 in two months from today?

$20 in one month	$23 in two months
○	○

Would you rather like to receive $20 in one month from today or $25 in two months from today?

$20 in one month	$25 in two months
○	○

Would you rather like to receive $20 in one month from today or $27 in two months from today?

$20 in one month ○ $27 in two months ○

Would you rather like to receive $20 in one month from today or $30 in two months from today?

$20 in one month ○ $30 in two months ○

Would you rather like to receive $20 in one month from today or $34 in two months from today?

$20 in one month ○ $34 in two months ○

Would you rather like to receive $20 in one month from today or $40 in two months from today?

$20 in one month ○ $40 in two months ○

→

Figure C11: Script Page 11

Would you rather like to receive $20 today or $20.50 in 1 month from today?

$20 today	$20.50 in one month
○	○

Would you rather like to receive $20 today or $21.50 in 1 month from today?

$20 today	$21.50 in one month
○	○

Would you rather like to receive $20 today or $23 in 1 month from today?

$20 today	$23 in one month
○	○

Would you rather like to receive $20 today or $25 in 1 month from today?

$20 today	$25 in one month
○	○

Would you rather like to receive $20 today or $27 in 1 month from today?

$20 today	$27 in one month
○	○

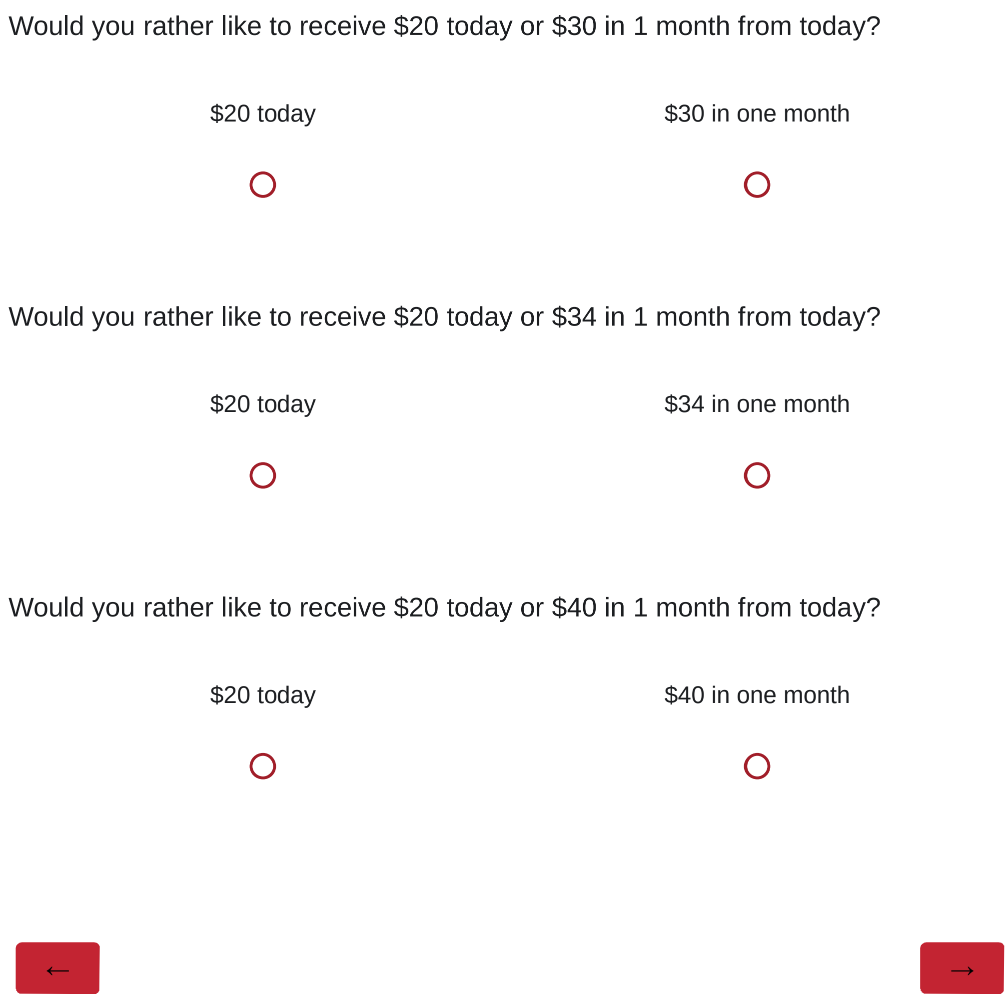


Figure C12: Script Page 12

Thanks for your answers. The next 5 situations will not involve time. Instead, you will be asked to choose between a lottery and a guaranteed payoff, both of which will be paid today if they are chosen for payment.

For example, you could be asked the following:

Do you prefer a lottery, which will pay you $20 with 50% chance and $0 with 50% chance or $12 for sure?

As you can see, the task follows a similar structure as the task before. The lottery option will always remain the same but the alternative option will vary from question to question. The only thing you have to do is to tell me if you prefer the lottery or the alternative.

Any questions?

[Wait for answer]

Then, let's start with the task.

→

Figure C13: Script Page 13

Do you prefer a lottery that will pay you $20 with 50% chance and $0 with 50% chance or $15 for sure?

Lottery	$15 guaranteed
○	○

Do you prefer a lottery that will pay you $20 with 50% chance and $0 with 50% chance or $10 for sure?

Lottery	$10 guaranteed
○	○

Do you prefer a lottery that will pay you $20 with 50% chance and $0 with 50% chance or $9 for sure?

Lottery	$9 guaranteed
○	○

Do you prefer a lottery that will pay you $20 with 50% chance and $0 with 50% chance or $7.50 for sure?

Lottery	$7.50 guaranteed
○	○

Do you prefer a lottery that will pay you $20 with 50% chance and $0 with 50% chance or $5 for sure?

Lottery	$5 guaranteed
○	○

← →

Figure C14: Script Page 14

Great! Let us now move to our final part, a short survey.

What is your gender?

- ○ male
- ○ female
- ○ other

How old are you?

[ ]

What is your nationality

[ ]

Are you a student?

- ○ Yes
- ○ No

What is your study major?

[ ]

How much money do you spend on average per week (on food, transport, leisure and other activities)?

[ ]

Have you volunteered with charitable organizations within the past year? By volunteering we mean activities such as coaching or helping kids at school, delivering food or fundraising for charity for no pay.

- ○ Yes
- ○ No

Have you donated to charity within the past year?

- ○ Yes
- ○ No

→

Figure C15: Script Page 15

Thank you for participating in our study! We are now done with all of our questions and I will venmo you your payment of $**6** for today's session right after our call. I will also call you next week. I will then venmo your payment for next week after our second call next week.

Unfortunately, you were not selected for payment in part 2 and part 3.

Before I say goodbye I would quickly like to discuss your availability next week? As I said before, we will call you exactly one week from today, which would be on Tuesday, September 7th. Would you be available at the same time as today?

[If not, what time would you be available so that we could call you again?]

[Please note date (dd.mm.yy) and time (hh:mm) you have agreed on. Leave field empty if you could not agree on a time]

←　　→

Figure C16: Script Page 16