Discussion Paper No. 1273

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

PAY A LOT TO A FEW INSTEAD OF A BIT TO ALL! EVIDENCE FROM ONLINE DONATION EXPERIMENTS

Yohei Mitani
Nobuyuki Hanaki

February 2025

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Pay a lot to a few instead of a bit to all! Evidence from online donation experiments*

Yohei Mitani† Nobuyuki Hanaki‡

January 31, 2025

Abstract

We conduct an online donation dictator game experiment with over 1,300 participants, representative of the Japanese population, to investigate the relationship between the incentive scheme and prosocial behavior by systematically varying the stake size and probability of being paid, including those where the expected payments are controlled. We find that stake size is the main driver of donation decisions, even in the hypothetical scenario. Our result suggests that paying a large amount to a few participants incentivizes donation decisions better than paying a small amount to many in large-scale online experiments.

Keywords: Between-subject random incentive, Donation-dictator game, Social preferences

JEL Classification: C91, D90

*The experiment reported in this paper is approved by IRB for Economic Experiments of Kyoto University GSS Unit. (No. 202406), and pre-registered at aspredicted.org (# 204371) and available at https://aspredicted.org/sh47-xp2j.pdf. The financial aids from Japan Society for Promotion of Sciences, grant-in-aid (20H05631, 23H00055, 23K25497, 24K21413) are gratefully acknowledged.

†Corresponding Author: Graduate School of Agriculture, Kyoto University. E-mail: mitani.yohei.7w@kyoto-u.ac.jp

‡Institute of Social and Economic Research, University of Osaka, and University of Limassol. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

1 Introduction

The dictator game is the simplest and most widely used experimental tool for measuring and studying prosocial behavior in economics. In this game, a dictator decides how to divide a stake between herself and a paired recipient. Over the past three decades, experimental studies using the dictator game have advanced our understanding of prosocial behavior by exploring the conditions that affect generosity and by linking observed generosity to the socioeconomic and demographic characteristics of different populations (Engel, 2011). As economists continue to strive for a deeper understanding of the external validity dimensions, an increasing number of large-scale experiments and surveys have employed the dictator game and its variants to investigate generosity in real-world situations among representative populations (Falk et al., 2018, 2023; Chisadza et al., 2023; Cartwright and Thompson, 2023).

In the standard version of the game, participants are incentivized by using real money for the stakes, which generates a real cost of giving in the experiment with the aim of inducing real-world-like incentives. However, large-scale experiments with increasingly larger sample sizes pose challenges for researchers due to high financial and logistical implementation costs (Brañas-Garza et al., 2024; Clot et al., 2018; Charness et al., 2016). At the same time, the use of low stake sizes and hypothetical payments has gained popularity, particularly in online experiments and experimentally validated survey tools in economics (Brañas-Garza et al., 2024; Amir et al., 2012; Falk et al., 2018).[1] This trend, coupled with the demand from researchers with limited budgets, raises questions about whether, how, and to what extent incentivization impacts the measurement of prosocial behavior in dictator games.

Since a striking contrast emerged in transfer rates between real and hypothetical payments from the earliest study of the dictator game (Sefton, 1992; Forsythe et al., 1994),[2] more than

[1]For example, Falk et al. (2018, 2023) propose survey tools, the Global Preference Survey module, to measure economic preferences that are experimentally validated by selecting survey items that are correlated with choices in incentivized experiments. To measure altruism, they use the donation amount in a hypothetical dictator game with a charity as a recipient.

[2]In their study, the histograms of transfers in the ultimatum game are similar for real and hypothetical

a half dozen studies have explored the effects of incentivization on prosocial behavior in dictator games. In this literature, three incentivization methods have been examined: pay all, pay some, and pay none (Charness et al., 2016; Clot et al., 2018; Berlin et al., 2024). Pay all is the standard practice in experimental economics, where all participants are paid, whereas pay none corresponds to hypothetical payments in which none are paid. Earlier studies have analyzed the differences in transfer rates and egalitarian behavior between hypothetical payment and all-pay treatments, mostly confirming the original view that generosity diminishes with real money incentives.[3]

As large-scale experiments have been adopted, economists have turned their attention to the third mechanism, pay some, where only a subset of participants are paid for real, with the hope that it could reduce financial and logistical costs for implementation while still providing real incentivization (Charness et al., 2013; Clot et al., 2018; Umer, 2023). This mechanism is known as a "between-subject random incentive system" because participants are paid with a certain probability.[4] Meta-analyses show no clear systematic difference in transfer rates between random incentives and the standard practice where all participants play once and receive full payment (Engel, 2011; Grech et al., 2022; Umer, 2023). However, these findings are inconclusive due to a lack of sufficient observations on the between-subject random incentive systems.[5] Among these individual studies, five have directly examined the performance of between-subject random incentives by comparing one specific probability of being selected for payment (0.1, with two exceptions of 0.25 and 0.5) against pay-all or pay-none (hypothetical)

payments. However, in the dictator game, the share of zero transfers has significantly decreased, and the share of half transfers has sharply increased in hypothetical payments compared to real payments. This contrast between real and hypothetical payments in the dictator game indicates that socially desirable transfers incur costs only with real payments.

[3]These studies include Forsythe et al. (1994), Dana et al. (2007), Ben-Ner et al. (2008), Amir et al. (2012), and Bühren and Kundt (2015).

[4]This term was introduced to distinguish it from the within-subject random incentive system, where one of the choices made by each participant is randomly selected for payment (Baltussen et al., 2012).

[5]Engel (2011) and Grech et al. (2022) combine between-subject and within-subject random incentives into a single random incentive. Umer (2023) distinguishes between these two random protocols but has only 12 observations for the between-subject random incentives, which is only 3%, resulting in low statistical power to detect a difference.

treatments at a certain level of stake size, and the results are somewhat mixed.[6] While Brañas-Garza et al. (2024) do not find significant differences between pay-none, pay-all, and pay-some treatments, others find more giving in pay-none treatment compared to incentivized ones. Between pay-all and pay-some, on one hand, Clot et al. (2018) do not find significant differences, while Chisadza et al. (2023), on the other hand, do.

Because these experiments differ not only in stake size but also in sample pools, a systematic analysis that varies both stake size and the probability of being paid, using the same sample pool, is needed to better understand the impact of the between-subject random incentive system on the observed behavior. Furthermore, these studies do not consider smaller probabilities, such as 5% and 1%, of being selected for payment. Because online experiments involving more than 100 participants are increasingly common and a low probability of being paid, such as 1%, has become more relevant (Ahles et al., 2024), it is of interest to understand the trade-off, if any, of further increasing the stake size while reducing the probability of being paid while keeping the expected value of payment constant.[7]

We fill this gap in the literature by considering a broader range of stake sizes and selection probabilities than the existing papers in a unified experiment on the donation dictator game where the receipient is a charity rather than another participant from the same subject pool (Cartwright and Thompson, 2023; Diederich et al., 2022). First, we employ five stake sizes: 100, 500, 1000, 2000, and 10000 JPY (comparable to 1, 5, 10, 20, and 100 USD in terms of purchasing power), with a focus on two common practices of standard and low stake sizes. 1000 JPY corresponds to the standard stake size of 10 USD, which is commonly used

[6]These five studies are Sefton (1992) (using selection probabilities of 0, 0.25, and 1 with a stake size of 5 USD), Clot et al. (2018) (using probabilities of 0, 0.1, and 1 with a stake size of 10 or 100 EUR), Walkowitz (2021) (using probabilities of 0.5 and 1 with a stake size of 10 EUR), Chisadza et al. (2023) (using probabilities of 0, 0.1, and 1 with a stake size of 100 or 500 ZAR), and Brañas-Garza et al. (2024) (using probabilities of 0, 0.1, and 1 with a stake size of 1 GBP).

[7]Ahles et al. (2024) investigate whether paying participants with 1% probability result in a different outcome than paying them with 10% or 100% (pay-all) probabilities in the context of a willingness-to-pay (WTP) elictation experiment. They report no statistically significant differences in the elicited WTP between pay-all, 10%, and 1% probability of being paid, which are all significantly lower than the WTP elicited in the hypothetical treatment.

in laboratory settings (Engel, 2011), while 100 JPY corresponds to a low stake size of 1 USD, which is becoming a new standard in large-scale online experiments (Amir et al., 2012; Brañas-Garza et al., 2024).

Second, we consider five selection probabilities: 0, 0.01, 0.05, 0.1, 0.2, and 1, focusing particularly on the small probability of 0.01 along with three common practices: 0, 0.1, and 1. In particular, we have four probabilities (0, 0.01, 0.1, and 1) for each of the 1 and 10 USD stake sizes, generating a full factorial combination of the two stake sizes and four selection probabilities. This experimental framework enables us to isolate the driving factors that influence donation decisions and test incentive effects in varying stake sizes.

Furthermore, our design manipulates the selection probability and stake size while controlling for expected payments. This means that we clearly distinguish between the stake size conditional on being selected for payment and the expected payment, which is the product of the conditional stake size and the selection probability. The expected payment acts as a confounding factor if we change the probability or stake while keeping one constant. Only Clot et al. (2018) addresses this issue to isolate a determinant of prosocial behavior in dictator games among selection probability, stake size, and expected payment.[8] We keep the expected payoff constant at 100 JPY (1USD) and vary the probability from 1, 0.2, 0.1, 0.05, to 0.01 to test whether the expected payoff is the primary determinant of prosocial behavior.

Using this experimental setup, we ask the following research questions regarding the impact of incentivization on prosocial behavior within donation dictator games. Answering these questions provides insight on how experimenters should incentivize participants under a limited budget.

RQ 1 *Do expected payoffs solely explain prosocial behavior?*

If the answer is yes, the result suggests that researchers should focus on determining

[8] Anderson et al. (2023) also use a similar design as Clot et al. (2018) to test incentive effects in the context of eliciting risk preferences.

the ideal level of expected payments and subsequently modify other parameters to reduce logistical implementation costs. Otherwise, the result suggests that either or both the selection probability and stake size conditional on being selected shall induce different prosocial behavior, leading to the second question.

RQ 2 *Which factor drives behavioral changes: selection probability or stake size?*

The answer offers valuable insight into the central question of "pay one or pay all" (Charness et al., 2016) in the context of donation dictator games.

RQ 3 *Do small incentives matter for donation decisions?*

Large-scale experiments prompt economists to consider small incentives due to constraints in research budgets (Amir et al., 2012; Brañas-Garza et al., 2024). We explore whether providing a small incentive creates a significant difference in measured prosocial behavior compared to a hypothetical payment. Specifically, we first examine the impact of small expected payments against hypothetical payments by asking:

RQ 3-1 *Does the effect of small-but-actual or hypothetical incentives on prosocial behavior vary with probability and stake size while keeping the expected payment constant?*

Next, we assess the effects of a small 1% probability of being paid compared hypothetical payments by asking:

RQ 3-2 *Do the effects of actual incentives with a small probability fluctuate with stake size?*

Finally, we compare hypothetical or probabilistic payments with sure payments in the same way as previous studies, aiming to replicate incentive effects at different stake sizes using a unified experiment setup.

RQ 4 *Can we replicate the incentive effects identified in the same manner as previous studies that compare hypothetical or probabilistic payments to sure payments at different stake sizes?*

We conduct an online donation dictator game with over 1,300 participants from an internet panel to implement this comprehensive test. Participants are a representative sample of the Japanese population in terms of gender, age, marital status, education, household income, and region of residence. Since most previous studies testing incentive effects in dictator games use the standard version of dictator games with student recipients,[9] this study also serves as a robustness check with charity recipients and representative participants.

We find that stake size is the primary driver of the decision to donate. Increasing it diminishes the observed generosity as shown by the meta-regression results by Engel (2011) and Larney et al. (2019). The stake size matters even in hypothetical scenarios. Furthermore, for the low stake size ($\approx$ 1 USD), the probability of being paid does not matter, echoing the finding of Brañas-Garza et al. (2024). For a larger stake size, there is a significant decline in the donation rate once participants are paid probabilistically. Reducing the payment probability from 100% to 10% results in about 12% decline in the donation rate in the standard stake size ($\approx$ 10 USD). This is consistent with most existing studies with similar stake sizes (Sefton, 1992; Chisadza et al., 2023).

What our results indicate is that even when the probability of being selected for payment is as low as 1%, if the stake size is as large as 100 USD, the observed prosocial behavior differs significantly from the hypothetical experiment with the same stake size. In contrast, the differences become statistically indistinguishable for stake sizes of 10 USD or less. Thus, even small incentives, in terms of the expected payment, can influence the observed generosity compared to the hypothetical experiment if the stake size is sufficiently large.

Furthermore, we find that increasing the stake size, while reducing the payment probability to keep the expected payoff equal to 1 USD, reduces the difference in observed generosity compared to the conventional laboratory setting, where all participants receive the standard

[9]Exceptions are that Amir et al. (2012) use MTurk to recruit their participants, Chisadza et al. (2023) use an online donation dictator game experiment with South African residents, and Brañas-Garza et al. (2024) use an interactive protocol with role duality and Prolific Academic to recruit their participants.

stake of 10 USD. Therefore, our results suggest that paying a large amount to a few participants incentivizes donation decisions more effectively than paying a small amount to many in large-scale donation experiments.

2 The Experiment

2.1 Donation Dictator Game

We use a real-donation dictator game, a variant of dictator games in which the recipient is a charity organization instead of another participant from the same subject pool (Diederich et al., 2022; Umer et al., 2022; Cartwright and Thompson, 2023).

We use a donation dictator game experiment for the following reasons. First, the donation dictator game is one of the most widely applied variants, with numerous applications in charitable giving and survey research (Cartwright and Thompson, 2023; Falk et al., 2018). Second, only half of the participants can be dictators and be included in the analysis in the standard version of the dictator game, while the other half are entirely passive recipients. To address this design problem, many studies utilize protocols with interactive roles or role uncertainty (Grech and Nax, 2020; Grech et al., 2022; Brañas-Garza et al., 2024).[10] However, interactive protocols alter rational choice benchmarks, which may result in measured generosity that differs systematically from the standard version (Grech and Nax, 2020; Grech et al., 2022). Third, the differences in measured prosocial behavior between standard and donation dictator games are well documented in the literature. Meta-analyses confirm that transfer rates tend to be higher when using charity or deserving recipients compared to standard recipients (Engel, 2011; Grech et al., 2022; Umer et al., 2022).

[10] In the interactive protocol, all participants act as both dictators and recipients (Grech et al., 2022). In the role uncertainty protocol, all participants make a dictator decision ex-ante, but only half are selected as dictators with a 50% chance, while the other half remain recipients (Grech et al., 2022; Brañas-Garza et al., 2024).

In our experiment, participants are asked how much money from an endowment they want to donate to a charity campaign.[11] We chose the Japan UNICEF Association's "UNICEF Fundraising" as a recipient because of its popularity and trustworthiness in Japan.[12]

The instructions describing the incentives in the donation dictator game, that is, how and how much can be paid, are as follows.

- All Pay treatments: "*All participants in this question will be paid Y JPY as additional points.*"
- Probabilistic Pay treatments: "*Among the 100 people participating in this question, X will be selected by lottery to receive Y JPY as additional points.*"
- Hypothetical Pay treatments: "*This is a hypothetical question, but please answer it as if it were as described. Let us assume that you have been paid Y JPY as additional points.*"

Following information on additional payments that vary with treatments, the charity's information is presented as in the official website of UNICEF Japan. Across treatments, the donation decision is elicited by the following question: "*How much of the Y JPY would you like to donate to the UNICEF Fundraising? Please enter the amount you want to donate in units of 1 JPY between 0 and Y in single-byte numbers.*" Participants decide their donation amount in increments of 1 JPY. Appendix C provides an English translation of the instructions.

[11] Money endowment is delivered as an extra reward in addition to a fixed participation fee. Only information on the amount of the fixed participation fee and the fact that there is an opportunity for an additional reward is explained in the informed consent form at the beginning of the Web survey experiment (see Appendix C), so participants would not know the chance to be selected for payment and its amount until they start the donation dictator game part.

[12] The Japan UNICEF Association is one of the National Committees for United Nations Children's Fund.

2.2 Treatments

Table 1 shows the thirteen treatment conditions. Treatments are indicated as T^{Prob}_{Stake}. The superscript $Prob$ indicates the probability of being selected for payment, which is calculated as the number of winners (X) divided by the number of participants (N). The subscript $Stake$ indicates the stake size of the payment (Y) in JPY, conditional on being selected.[13]

Table 1: Experimental Design (Treatments)

#	T	Prob	Stake	E[Payoff]	X	N
1	T^{1}_{1000}	1	1000	1000	100	100
2	T^{1}_{100}	1	100	100	100	100
3	$T^{0.2}_{500}$	0.2	500	100	20	100
4	$T^{0.1}_{1000}$	0.1	1000	100	10	100
5	$T^{0.05}_{2000}$	0.05	2000	100	5	100
6	$T^{0.01}_{10000}$	0.01	10000	100	1	100
7	$T^{0.01}_{1000}$	0.01	1000	10	1	100
8	$T^{0.1}_{100}$	0.1	100	10	10	100
9	$T^{0.01}_{100}$	0.01	100	1	1	100
10	T^{0}_{100}	0	100	0	0	100
11	T^{0}_{1000}	0	1000	0	0	100
12	T^{0}_{2000}	0	2000	0	0	100
13	T^{0}_{10000}	0	10000	0	0	100

Note: Treatments are expressed as T^{Prob}_{Stake} where the subscript $Stake$ indicates the stake size of payment (in JPY) conditional on being selected, and the superscript $Prob$ indicates the probability of being selected for payment. X is the number of winners, while N is the number of participants in a treatment. 1000 JPY $\approx$ 6.5 USD at the time of the experiment, but note that 1000 JPY is equivalent to about 10 USD (converted using the purchasing power parity).

Two All Pay treatments (#1 and #2 in Table 1) serve as baseline conditions. T^{1}_{1000} is the ideal experimental setting in which we highly reward all participants, and is the standard dictator game in a laboratory setting (Engel, 2011). T^{1}_{100} pays all participants but a low stake, which is gaining popularity among online economic experiments (Brañas-Garza et al.,

[13] 1000 JPY $\approx$ 6.5 USD at the time of the experiment. Note that 1000 JPY is comparable to 10 USD for Japanese residents and is 10.56 USD according to the purchasing power parity calcurater `https://pppcalculator.pro/` checked on January 24, 2025.

2024; Amir et al., 2012).

To address RQ1, while keeping the expected value of the payment constant at 100 JPY (comparable to 1 USD), we compare T^{1}_{100} with four Probabilistic Pay treatments (from #3 to #6 in Table 1). As we move from $T^{0.2}_{500}$ to $T^{0.01}_{10000}$, the probability of being selected for payment decreases from 0.2, 0.1, 0.05, to 0.01, while the stake size conditional on winning increases from 500, 1000, 2000, to 10000. To our knowledge, no study has investigated the impact of between-subject random incentives with selection probabilities of 5% or less on prosociality in dictator games, although such low probabilities for between-subject random incentivized methods are increasingly relevant for large-scale online experiments.

To address RQ2 and further explore the impact of low probability, we add two more treatments with a 1% probability of standard and low stake payments (#7 and #9 in Table 1): $T^{0.01}_{1000}$ and $T^{0.01}_{100}$. We use the seven All and Probabilistic Pay treatments, with stake sizes of 100, 500, or 1000, to explore which factors among selection probability, stake size, and expected payment likely drive behavioral changes in the donation dictator game. Previous studies have limited variations in probabilities and stake sizes. Most studies compare Hypothetical Pay, Probabilistic Pay, and All Pay treatments for a given stake size, while Clot et al. (2018) compare two Probabilistic Pay treatments of standard and high stakes with an All Pay treatment of standard stake. We use rich variations from the full factorial combination of probability and stake size to isolate driving factors in donation decisions.

We also have four Hypothetical Pay treatments to understand how monetary incentives shape prosocial behaviors observed through dictator game experiments (from #10 to #13 in Table 1) by addressing RQ3-1 and RQ3-2. Existing studies use one Hypothetical Pay treatment of a specific stake size and compare it with All Pay and Probabilistic Pay with relatively high probabilities of 10% or greater (Clot et al., 2018; Brañas-Garza et al., 2024), but not with a lower probability.

2.3 Implementation

We conducted the experiment online in December 2024 using the self-administered survey platform provided by Qualtrics. Online experiments have been increasingly used to investigate social preferences with dictator games (Brañas-Garza et al., 2024; Diederich et al., 2022; Chisadza et al., 2023) and other economic preferences (Brañas-Garza et al., 2023; Chapman et al., 2024b; Hanaki et al., 2024).[14]

About a week before the scheduled survey experiment, a Japanese private research institution (MyVoice Communications, Inc.) conducted a pre-screening survey in which 49,475 Japanese residents aged 20 or older were randomly recruited from the institution's Internet panel.[15] The pre-screening survey provided minimum information on the survey experiment, such as its scheduled period, expected length, and completion fee. Then, age, gender, and willingness to participate in the experiment were asked.

Approximately 2300 respondents who declared their willingness received an invitation email from the institute with a link that directed them to the survey experiment on Qualtrics. Participants were informed that they would earn 230 JPY for completing a 20-30 minute academic survey on economic behaviors and that each participant was allowed to participate only once. Participants who confirmed informed consent were randomly assigned to one of the thirteen treatments. The survey experiment consisted of non-incentivized economic preference modules (Chapman et al., 2024b; Falk et al., 2018), a questionnaire on food and environment-related topics, the donation experiment, and a 3-item cognitive reflection test (Frederick, 2005). The order of the first two parts before the donation experiment was randomized, while the last two parts were given always in this order.

[14]For comparisons between online and laboratory experiments, see, among others, Arechar et al. (2018), Hergueux and Jacquemet (2015), Snowberg and Yariv (2021), Ozono and Nakama (2022), Prissé and Jorrat (2022), and Schmelz and Ziegelmeyer (2020).

[15]The Internet panel comprises more than 1 million active Japanese residents with a wide range of age and other demographic characteristics. The institution constantly monitors and trucks the panel's activities as well as updates the panel's demographic characteristics twice a year.

Table 2: Sample Characteristics by Treatments

	Age	Male	CRT	ρ	λ	δ	β	N
T^{1}_{1000}	47.28	0.52	1.35	0.80	2.36	0.83	0.89	100
T^{1}_{100}	49.84	0.56	1.45	0.75	2.45	0.86	0.91	101
$T^{0.2}_{500}$	51.91	0.46	1.16	0.76	2.58	0.87	0.90	100
$T^{0.1}_{1000}$	49.17	0.50	1.27	0.77	2.54	0.86	0.88	100
$T^{0.05}_{2000}$	53.79	0.50	1.09	0.73	2.42	0.87	0.89	100
$T^{0.01}_{10000}$	46.62	0.50	1.41	0.77	2.64	0.85	0.90	101
$T^{0.01}_{1000}$	50.47	0.47	1.18	0.75	2.27	0.83	0.90	100
$T^{0.1}_{100}$	47.63	0.46	1.35	0.72	2.43	0.86	0.90	100
$T^{0.01}_{100}$	49.17	0.53	1.26	0.73	2.57	0.86	0.90	100
T^{0}_{100}	49.64	0.49	1.18	0.72	2.49	0.85	0.88	107
T^{0}_{1000}	50.20	0.53	1.10	0.73	2.67	0.83	0.88	112
T^{0}_{2000}	48.20	0.47	1.17	0.73	2.49	0.86	0.89	109
T^{0}_{10000}	48.19	0.49	1.21	0.77	2.44	0.88	0.90	111
Mean	49.38	0.50	1.24	0.75	2.49	0.85	0.89	
SD	16.84	0.50	1.12	0.33	1.11	0.18	0.12	
N	1,341	1,341	1,341	1,341	1,341	1,341	1,341	
p-value	0.17	0.97	0.33	0.77	0.60	0.66	0.45	

Note: CRT is Frederick (2005)'s 3-item cognitive reflection test score from 0 to 3. Risk aversion $(1-\rho)$, loss aversion (λ), less discounting (δ), and patience (β) are elicited by DOSE method (Chapman et al., 2024b). p-value is from the Kruskal-Wallis test of differences among all treatment groups for each characteristic.

2.4 Sample

A total of 1,341 people participated in the donation dictator game experiment. The sample is fairly representative of the population in terms of gender, age, marriage status, education, and household income. Half of the participants are female, and the average age is 49.4. Slightly more than half of the participants are married, and the median annual household income is about 6 million JPY. The participants have 15 years of education on average.

Table 2 reports the participants' cognitive test score and economic preferences, as well as age and gender. CRT is Frederick (2005)'s 3-item cognitive reflection test score, which takes a value between 0 and 3. Risk aversion $(1-\rho)$, loss aversion (λ), less discounting (δ), and patience (β) are elicited by the 20-question version of the dynamically optimized sequential

Table 3: Descriptive Results by Treatments

	Rate		Indicators		
	Mean	SD	Zero	Half	Full
T^{1}_{1000}	0.18	0.24	0.44	0.19	0.03
T^{1}_{100}	0.34	0.40	0.43	0.22	0.22
$T^{0.2}_{500}$	0.30	0.36	0.36	0.10	0.17
$T^{0.1}_{1000}$	0.30	0.32	0.31	0.30	0.09
$T^{0.05}_{2000}$	0.24	0.26	0.35	0.24	0.05
$T^{0.01}_{10000}$	0.23	0.30	0.29	0.14	0.07
$T^{0.01}_{1000}$	0.24	0.33	0.40	0.17	0.11
$T^{0.1}_{100}$	0.34	0.42	0.42	0.10	0.25
$T^{0.01}_{100}$	0.42	0.43	0.42	0.18	0.29
T^{0}_{100}	0.37	0.41	0.36	0.19	0.24
T^{0}_{1000}	0.30	0.37	0.38	0.22	0.14
T^{0}_{2000}	0.20	0.28	0.40	0.11	0.06
T^{0}_{10000}	0.16	0.28	0.45	0.11	0.05
Total	0.28	0.34	0.39	0.17	0.14

experimentation (DOSE) method (Chapman et al., 2024a,b).[16]

The last column of Table 2 presents the number of participants in each treatment. Kruskal–Wallis tests do not show significant differences among all treatment groups for each characteristics (see p-value reported in the last row of Table 2). However, given the diversity of individual characteristics,[17] we control these characteristics in the following analyses.

3 Results

The outcome variables of interest are the donation rate ($Rate$), the share of zero donation ($Zero$), the share of half donation ($Half$), and the share of full donation ($Full$).[18]

[16]The risk module of DOSE elicits the preference parameters ρ and λ in $v = x^{\rho}$ for $x \geq 0$, and $v = -\lambda(-x)^{\rho}$ for $x < 0$. The time module uses ρ from the risk module to estimate the preference parameters δ and β in $u = \beta\delta^{t}x_{t}^{\rho}$.

[17]For example, the participants' age ranges from 20 to 82.

[18]$Rate$ (r_i) is calculated as $r_i = \frac{c_i}{Y}$, where c_i is observed donation amount and Y is the stake size. $Zero$ is a selfish share, defined as $Pr[r_i = 0] = \frac{\sum_{i=1}^{N} I[r_i=0]}{N}$ where N is the number of participants. $Half$ is a egalitarian share, defined as $Pr[r_i = 0.5] = \frac{\sum_{i=1}^{n} I[r_i=0.5]}{n}$. $Full$ is a hyper-altruistic share, defined as $Pr[r_i = 1] = \frac{\sum_{i=1}^{n} I[r_i=1]}{n}$ (Branas-Garza et al., 2021).

Table 3 presents descriptive results of prosocial behavior measured by our outcome variables. The first two columns report the means and standard deviations of the donation rate (*Rate*). The overall average donation rate is 27.8%, which is very similar to the meta-estimates from laboratory experiments, that is, 28.4% by Engel (2011) and 30.2% by Umer (2023). The overall fractions of participants who donate nothing and half of the stake are 38.6% and 17.4%, which are also similar to those reported by Engel (2011), 36.1% and 16.7%. The overall share of full donation is 13.6%, which is higher than 5.4% reported by Engel (2011). It is surprising that overall donation decisions are as generous as meta-estimates because we use a charity organization as a recipient and several hypothetical-pay treatments are included, in which the literature suggests higher generosity for charity recipients and hypothetical payments (Engel, 2011; Umer et al., 2022; Forsythe et al., 1994; Sefton, 1992).[19]

3.1 Prosocial behavior under the same expected value of the payment

To test whether the expected payoff is the primary determinant of prosocial behavior, we vary the probability from 1 to 0.01 and the stake from 100 to 10,000 while keeping the expected value of the payment constant at 100 JPY. Thus, we compare the five incentivized treatments: T^{1}_{100}, $T^{0.2}_{500}$, $T^{0.1}_{1000}$, $T^{0.05}_{2000}$, and $T^{0.01}_{10000}$. Table 3 indicates that the donation rate, standard deviation, zero donation, and full donation tend to decrease by approximately 10 to 15% as the probability of payment decreases from 1 (i.e., T^{1}_{100}) to 0.01 (i.e., $T^{0.01}_{10000}$), while no clear trend is detected in half donation.

Figure 1 shows histograms of donation rates in these five incentivized treatments.[20] The distribution has characteristics similar to one of the individual transfer rates shown in Engel

[19]Decomposing by three incentivized mechanisms, although those are not directly comparable, *Rate*, *Zero*, *Half*, and *Full* are respectively 0.26, 0.43, 0.20, 0.12 in the All Pay treatments (N=201), 0.30, 0.36, 0.18, 0.15 in the Probabilistic Pay treatments (N=701), and 0.26, 0.40, 0.16, 0.13 in the Hypothetical Pay treatments (N=439).

[20]Figure B1 shows histograms of the rest of the treatments.

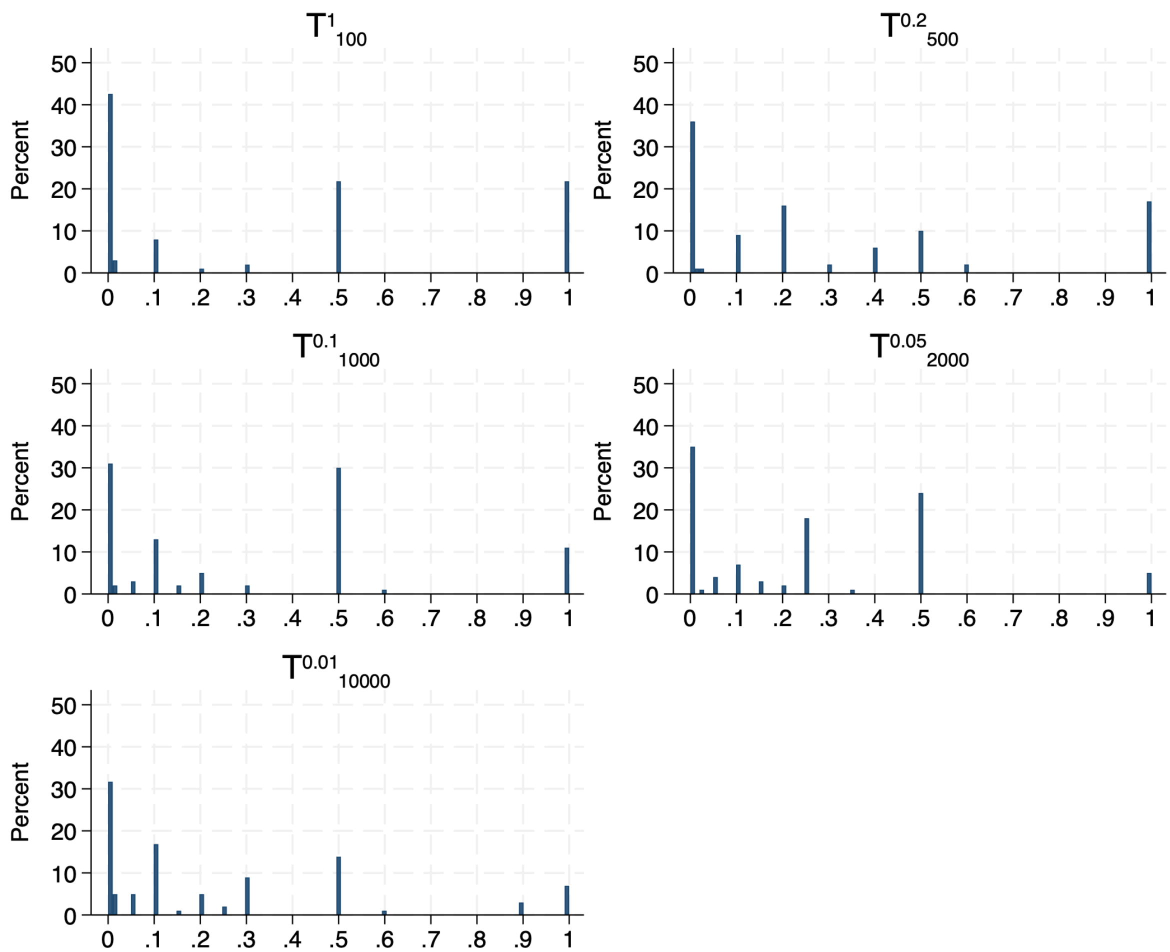

Figure 1: Histgrams of Donation Rates by Treatments

(2011) in that there are three peaks at zero, half, and full donations. A comparison of the five histograms indicates that decreasing the probability and increasing the stake size contribute to reducing the shares of both selfish and altruistic behaviors. This reduction in both extreme behaviors can explain the associated decrease in standard deviations of donation rates.

To confirm these descriptive observations, we regress the outcome variables on the four treatment dummies while using T^1_{100} as the base category.[21] Figure 2 visualizes the point estimates with 95% confidence intervals of the OLS model for *Rate* and the average marginal

[21] See Table A1 for full regression results.

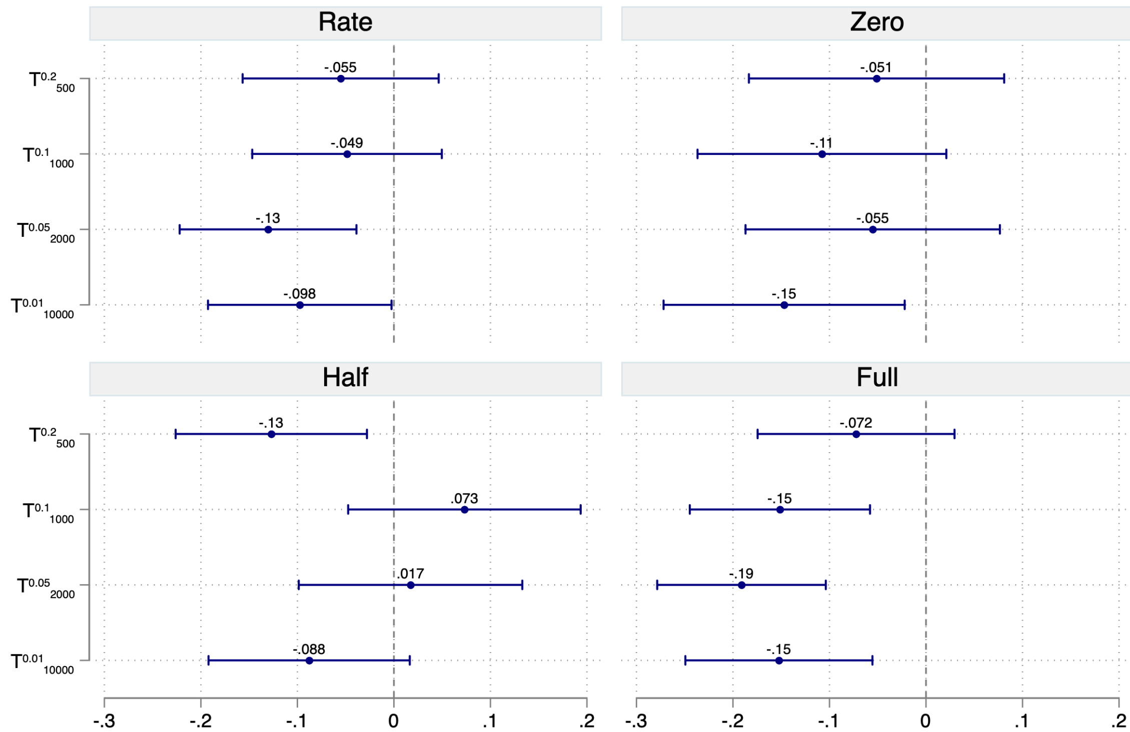

Figure 2: Differences in the observed prosocial behavior among five treatments with the same expected payment of 100 JPY. Point Estimates with 95% Confidence Intervals

Note: The base category is T^{1}_{100} in all model. OLS estimates are reported for Rate. Probit average marginal effects are reported for Zero, Half, and Full. Full estimation results including controls are presented in Table A1.

effects of the Probit models for $Zero$, $Half$, and $Full$.

Looking at $Rate$ in Figure 2, the result confirms that the donation rates in $T^{0.05}_{2000}$ and $T^{0.01}_{10000}$ are significantly lower than T^{1}_{100} at the 5% level. Regarding $Zero$, probabilistic pay treatments tend to lower the share of zero donation, but only the difference between $T^{0.01}_{10000}$ and T^{1}_{100} is statistically significant at the 5% level. The result of $Half$ does not show a clear reduction or increase in the share of the half donation. Concerning $Full$, clearer trends are observed for the share of full donation with a significant decrease of around 15% from T^{1}_{100} except for $T^{0.2}_{500}$.

In short, this result indicates that a relatively large deviation from T^{1}_{100} towards $T^{0.01}_{10000}$

significantly reduces the observed generosity, although a relatively small deviation from T^{1}_{100} towards $T^{0.2}_{500}$ results in imprecise impacts on reducing generosity.

Finally, F and χ^2 tests reject that the coefficients on four treatment dummies are simultaneously zero for $Rate$, $Half$, and $Full$ at the 5% level (Table A1). We take this as evidence that the expected value of monetary incentives is not solely shaping prosocial behavior in the donation dictator game experiment.

Result 1: *Expected payoffs are not solely shaping donation decisions.*

This result suggests that prosocial behavior observed in a donation dictator game is altered by incentive mechanisms. Thus, a better understanding of how to set the probability of selection and the stake size requires further analyses.

To explore which combination of selection probability and stake size results in outcomes similar to those observed in the conventional laboratory setting where we pay all participants the standard stake of 1000 JPY (comparable to 10 USD), we compare the five treatments that have the same expected payoff of 100 JPY to this baseline (T^{1}_{1000}). The result is shown in Figure 3.

We find that, on one hand, low-stake, high-chance treatments (T^{1}_{100} and $T^{0.2}_{500}$) tend to generate significantly higher donation rates and a higher share of full donation than T^{1}_{1000}. On the other hand, high-stake, low-chance treatments ($T^{0.05}_{2000}$ and $T^{0.01}_{10000}$) tend to reduce the differences in the donation rate and the share of full donation from T^{1}_{1000}. However, we find no clear pattern for the shares of zero and half donation. Among these five alternative combinations, only $T^{0.05}_{2000}$ generates prosocial behavior that is statistically indistinguishable for all outcome variables from T^{1}_{1000}.

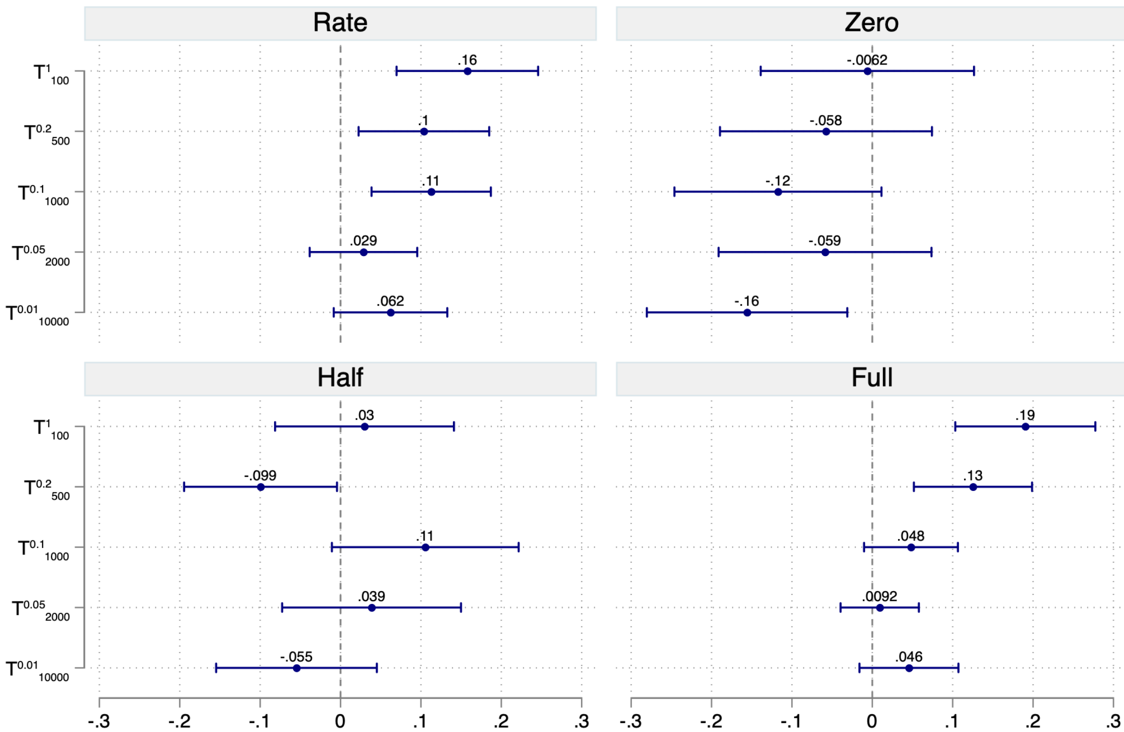

Figure 3: Difference in the observed behavior between the treatment guaranteed sure payment of 1000 JPY and five treatments with the expected payment of 100 JPY. Point Estimates with 95% Confidence Intervals

Note: The base category is T^1_{1000}. The number of observations is 602, consisting of T^1_{1000}, T^1_{100}, $T^{0.2}_{500}$, $T^{0.1}_{1000}$, $T^{0.05}_{2000}$, and $T^{0.01}_{10000}$. OLS estimates are reported for Rate. Probit average marginal effects are reported for Zero, Half, and Full. All control variables listed in Table 2 are included. 1000 JPY is comparable to 10 USD.

3.2 What drives behavioral changes?

We find that donation decisions are altered by selection probability and stake size even for the same expected payoffs (Result 1). If the expected payoff is not likely a determinant of prosocial behavior, then what drives behavioral changes? Do either or both probability and stake affect prosocial behavior? To identify determinants, we regress outcome variables on probability, stake size, and their interaction (i.e., expected payoff) using 701 observations from the seven incentivized treatments. Six of them are the full factorial combination of two stake sizes (100 and 1000) and three probabilities (0.01, 0.1, and 1): T^1_{1000}, $T^{0.1}_{1000}$, $T^{0.01}_{1000}$, T^1_{100},

Table 4: Effects of the selection probability and the stake size. Regression based on the seven incentivized treatments

	(1) Rate	(2) Zero	(3) Half	(4) Full
Payment Prob	-0.039	0.003	0.105*	-0.041
	(0.056)	(0.067)	(0.056)	(0.045)
Stake Size (1000)	-0.132***	-0.054	0.101**	-0.189***
	(0.042)	(0.056)	(0.046)	(0.040)
E[Payment] (1000)	-0.042	0.073	-0.131*	-0.090
	(0.067)	(0.096)	(0.076)	(0.086)
Controls	In	In	In	In
N	701	701	701	701

Note: Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$. The sample consists of T^{1}_{1000}, T^{1}_{100}, $T^{0.2}_{500}$, $T^{0.1}_{1000}$, $T^{0.01}_{1000}$, $T^{0.1}_{100}$, and $T^{0.01}_{100}$. Full estimation results including controls are presented in Table A2.

$T^{0.1}_{100}$, and $T^{0.01}_{100}$. The latter $T^{0.2}_{500}$ is orthogonal to the others. Note that the analysis here is based on the limited range of stakes between 100 and 1000 JPY (which is comparable to the range of 1 and 10 USD).

The result shown in Table 4 suggests that stake size is the single significant factor affecting *Rate* and *Full*, after controlling for probability and expected payoff. A 1000 JPY (comparable to 10 USD) increase in stake size decreases donation rates by 13% (percentage point) and the share of full donation by 19% (percentage point), while the coefficients of the other two factors are small and imprecise. These impacts generated by a 10 dollar increase conditional on winning are economically significant given that the average donation rate of this sub-sample is 30% and the average share of full donation is 17%.

For *Half*, stake size significantly increases the share of half donation but may not be the only contributing factor since the coefficients of probability and expected payoff are large. This suggests that increasing probability (0 to 1) and stake size (0 to 1000) increases the likelihood by 8%, whereas increasing stake size while decreasing probability would be associated with 12% increase in the likelihood. For *Zero*, the result suggests that none of the three factors is significantly associated with the share of zero donation (which does not exclude the possibility of non-linear effects). In short, stake size decreases donation rates and the share of full donation, implying that increasing the potential stake size can diminish observed generosity in the range of stake size between 100 and 1000 JPY (comparable to 1 and 10 USD).

If donation decisions are responsive only to monetary incentives, any hypothetical treatments would produce a similar result on average. We first investigate whether four hypothetical pay treatments (i.e., T^0_{100}, T^0_{1000}, T^0_{2000}, and T^0_{10000}), where only hypothetical stake size varies (from 100 to 10000), generate the same behaviors. Kruskal-Wallis tests reject the hypothesis that donation decisions are equal across all four treatments for *Rate* ($p < 0.001$), *Half* ($p = 0.042$), and *Full* ($p < 0.001$), except for *Zero* ($p = 0.533$). Furthermore, the regression analysis reported in Table A3 shows that the hypothetical stake size influences donation decisions in the same way as in the incentivized treatments reported in Table 4.

Result 2: *Stake size is the primary determinant of donation decisions.*

3.3 Does a small incentive matter for donation decisions?

In this subsection, we explore whether and how giving a "small" (in terms of the expected value or in terms of probability of being selected for payment) but real incentive affects prosocial behavior by comparing outcome variables from treatment with small incentives to comparable treatment with hypothetical incentives.

First, we explore the impact of a "small" expected payment of 100 JPY (which is comparable to 1 USD) on donation decisions compared to the relevant hypothetical payment with the same stake size (RQ3-1). The only difference is that one of each pair is an actual stake conditional on winning (i.e., selection probability is positive), and the other is a hypothet-

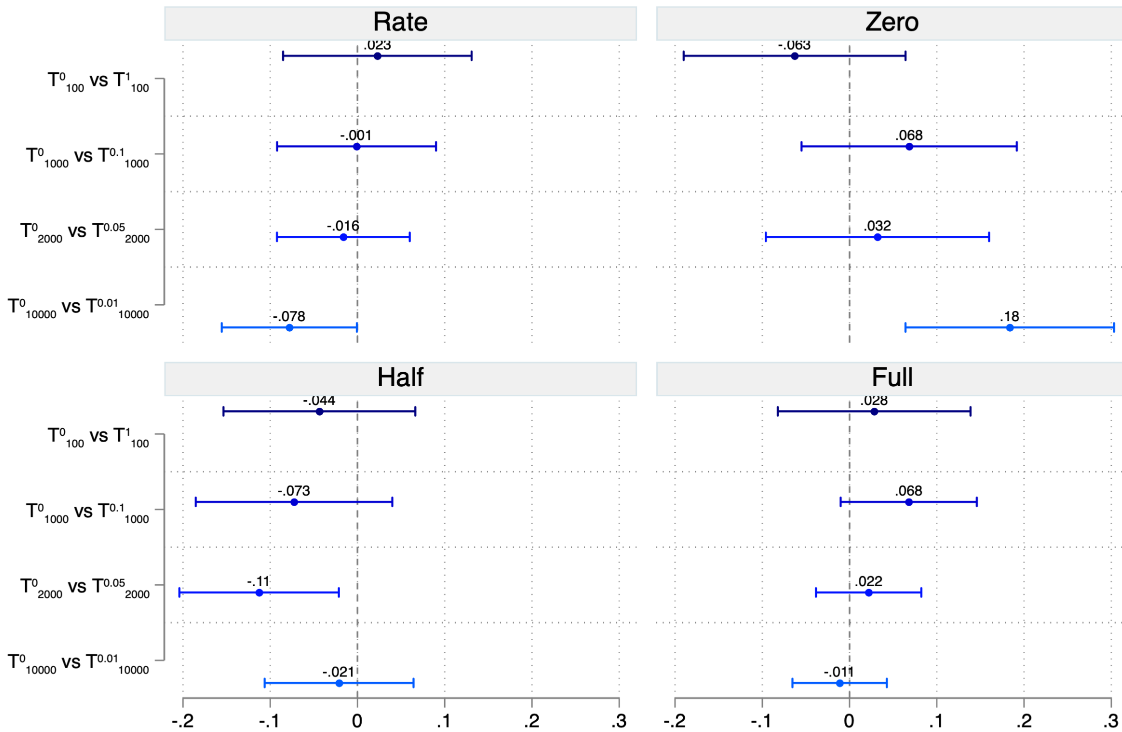

Figure 4: Effect of small expected payment compared to the hypothetical payment with the same stake size. Point Estimates with 95% Confidence Intervals (Based on four separate regressions)

Note: The base category is an incentivized treatment in all models. OLS estimates are reported for Rate. Probit average marginal effects are reported for Zero, Half, and Full. Detailed results are presented in Table A4.

ical stake (i.e., selection probability is zero).[22] Do these small incentive effects vary with selection probability and stake size conditional on winning while keeping expected payments constant? We run four separete regressions using two treatments each to conduct a pairwise comparison between hypothetical and incentivized decisions (Table A4). In these regressions, incentivized treatments are used as the base category.

Figure 4 visualizes the point estimates with 95% confidence intervals. Overall, the result suggests that a small incentive which has the expected payoff of 100 JPY ($\approx$ 1 USD) produces statistically indistinguishable differences in donation decisions for stake sizes of 1000 JPY or

[22]We have four pairwise comparisons: (T^0_{100}, T^1_{100}), $(T^0_{1000}, T^{0.1}_{1000})$, $(T^0_{2000}, T^{0.05}_{2000})$, and $(T^0_{10000}, T^{0.01}_{10000})$.

less. However, the same small incentive in terms of its expected value significantly reduces the share of zero donation by 18%, resulting in an 8% increase in the donation rate when the stake size is as large as 10000 JPY.[23]

This finding that hypothetical payments diminish generosity in dictator games contradicts the majority of previous studies that compare hypothetical payments with sure payments (Amir et al., 2012; Clot et al., 2018; Sefton, 1992; Dana et al., 2007) whereas it is consistent with the results of Brañas-Garza et al. (2024) and Bühren and Kundt (2015) who also use small incentives (1 GBP and 1 EUR, respectively). This suggests that the effect of incetivization on prosociality in dictator games can differ in its impact and even direction (making more or less generous) by the size of incentives.

Next, we explore the impact of monetary incentives with a "small" probability of winning (1% chance) on donation decisions compared to the relevant hypothetical payment that has the same stake size (RQ3-2).[24] Does this small incentive in terms of selection probability have different impact depending on the size of (conditional or hypothetical) stake? We run three separete regressions using two treatments each to conduct a pairwise comparison between hypothetical and incentivized decisions.

Figure 5 visualizes the point estimates with 95% confidence intervals.[25] The result is similar to the previous one shown in Figure 4 in the sense that no difference is precisely estimated except for the stake size of 10000 JPY. This evidence suggests that small incentive, in terms of either small expected payment (1 USD) or small selection probability (1% chance), generates significant differences in donation decisions from the hypothetical counter part only when stake size is as large as 100 USD.

[23] Note that the differences in the shares of half and full donations are very small and precisely estimated. This might suggest that the impact of small incentives can be different for selfish types (measured by $Zero$) and egatarian or altruistic types (measured by $Half$ or $Full$).

[24] We have three pairwise comparisons: (T^{0}_{100}, $T^{0.01}_{100}$), (T^{0}_{1000}, $T^{0.01}_{1000}$), and (T^{0}_{10000}, $T^{0.01}_{10000}$).

[25] Note that a pairwise comparison between T^{0}_{10000} and $T^{0.01}_{10000}$ is excatly same as one in the previous analysis shown in Figure 4.

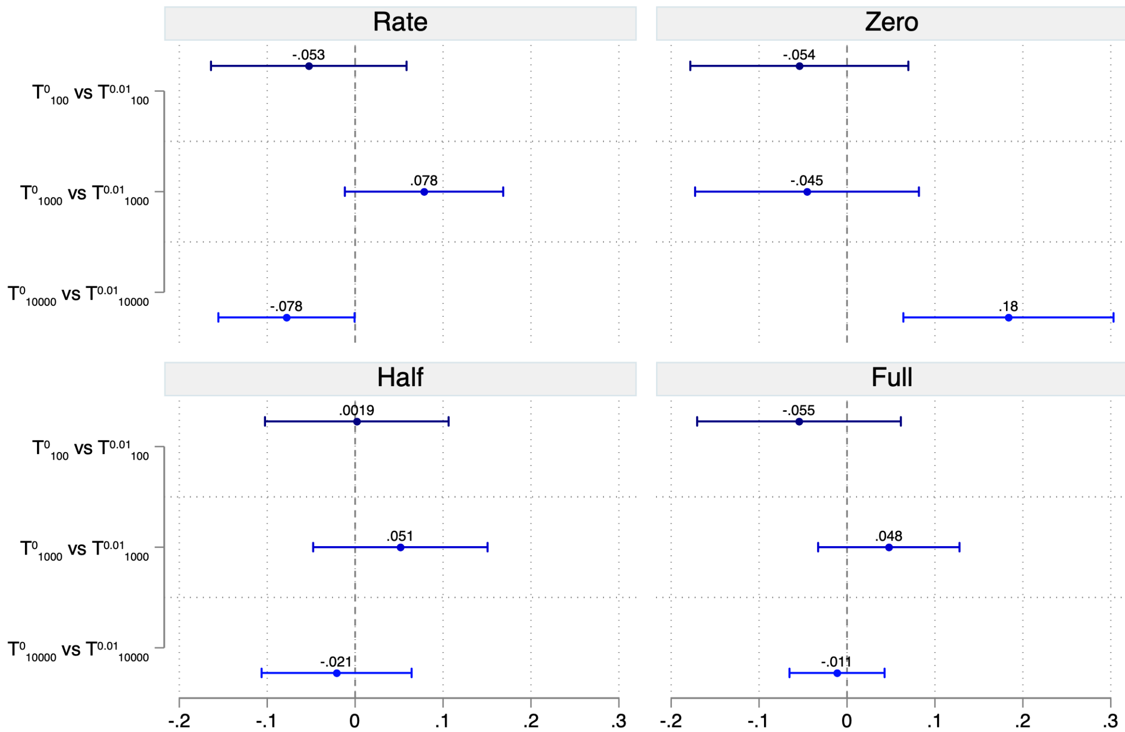

Figure 5: Effect of small probability of being selected for payment compared to hypothetical payment with the same stake size. Point Estimates with 95% Confidence Intervals (Based on three separate regressions)

Note: The base category is an incentivized treatment in all models. OLS estimates are reported for Rate. Probit average marginal effects are reported for Zero, Half, and Full. Detailed results are presented in Table A5.

Result 3: *The effect of small incentives relative to hypothetical ones on donation decisions emerges only for a large stake size.*

3.4 Testing incentive effects at different stake sizes

Accumulating evidence from this study suggests that stake size, conditional or hypothetical, shapes prosocial behavior and matters for incentive effects as well. Nonetheless, all previous studies testing incentive effects use a specific stake size for their comparison of transfers or

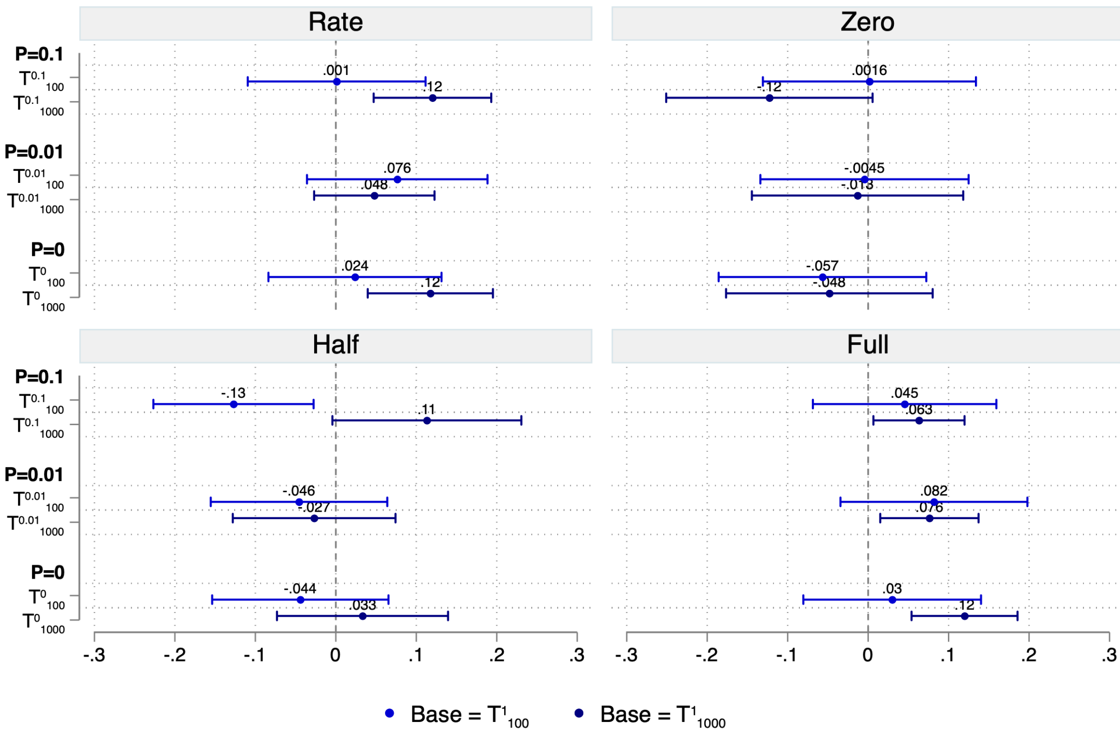

Figure 6: Comparing incentive effect under two stake sizes, 100 JPY and 1000 JPY. Point Estimates with 95% Confidence Intervals

Note: The base category is a All-Pay treatment in all models. OLS estimates are reported for Rate. Probit average marginal effects are reported for Zero, Half, and Full. Detailed results are presented in Table A6. 100 JPY is comparable to 1 USD.

donations between different payment methods.[26] This leads to the following question (RQ4). Can testing incentive effects at different stake sizes draw a different conclusion? We test incentive effects in the same way as previous studies, but at two different levels of stake size. We run two separete regressions using four treatments each to compare hypothetical (0% chance of being paid) and two probabilistic payments (1% and 10% chances) to sure payment (100%) at stake sizes of 100 JPY and 1000 JPY (comparable to 1 USD and 10 USD). In these regressions, All Pay treatment is used as the base category (see, Table A6).

[26]To our knowledge, about half a dozen studies have tested incentive effects on prosociality in dictator games so far. Earlier studies compare sure payment with hypothetical payment (Forsythe et al., 1994; Dana et al., 2007; Amir et al., 2012; Bühren and Kundt, 2015), while the rest of the studies add probabilistic payment to the comparison (Clot et al., 2018; Sefton, 1992; Brañas-Garza et al., 2024; Chisadza et al., 2023).

Figure 6 visualizes the point estimates with 95% confidence intervals. We find no clear incentive effects under the low stake size of 100 JPY ($\approx$ 1 USD). However, we confirm significant incentive effects on donation rates and the share of full donation under the standard stake size of 1000 JPY ($\approx$ 10 USD). Reducing the probability of being paid from 1 (T^{1}_{1000}) to 0.1 ($T^{0.1}_{1000}$) increases the donation rate by 12% and the share of the full donation by 6% when the stake size (certain or conditional) is 1000 JPY. Similarly, hypothetical payment (T^{0}_{1000}) increases the donation rate by 12% and the full donation by 12% relative to the guaranteed payment (T^{1}_{1000}).[27] This significant result under the 10 USD stake size is consistent with previous findings under similar stake sizes (Sefton, 1992; Clot et al., 2018; Chisadza et al., 2023), while our null result under the 1 USD stake size is also consistent with previous findings under similar stake sizes (Brañas-Garza et al., 2024; Bühren and Kundt, 2015).

Result 4: *Incentive effects on donation decisions are evident under the standard stake size (10 USD), although the effects become smaller and imprecise under the low stake size (1 USD).*

4 Concluding Remarks

This paper systematically investigates the relationship between stake size, the probability of being selected for payment, and the generosity observed in the online donation dictator game experiment. Compared to similar existing studies (Sefton, 1992; Clot et al., 2018; Walkowitz, 2021; Chisadza et al., 2023; Brañas-Garza et al., 2024), we examine a broader range of stake sizes and selection probabilities within one set of experiments.

The main takeaway from our experimental results is that *stake size matters for donation decisions regardless of its chance to be realized, including hypothetical scenarios.*

The incentive effects regarding whether reducing the selection probability from 1 towards

[27]In addition, reducing the probability of being paid from 1 to 0.01 increases the full donation by 7.6% while not having singificant effect on dnation rate.

0 increases the measured generosity are contingent on the stake size. For a small stake size (100 JPY $\approx$ 1 USD), the observed generosity remains consistent regardless of whether the experiment is hypothetical or not, and the selection probability is irrelevant. In contrast, for a larger stake size (1000 JPY $\approx$ 10 USD), the selection probability becomes important. That is, the donation rate and the fraction of full donation when all participants are surely paid are significantly lower than in the hypothetical scenario and the case where 10% of participants are paid, whereas there is no significant difference between these last two treatments.

What our results indicate is that even when the probability of being selected for payment is as low as 1%, if the stake size is as large as 100 USD, the observed prosocial behavior differs significantly from the hypothetical experiment with the same stake size. In contrast, the differences become statistically indistinguishable for small stake sizes of 10 USD or less. Thus, even small incentives, in terms of a low chance to be realized or small expected payments, can influence the observed generosity compared to the hypothetical experiment if the stake size is sufficiently large.

Furthermore, we find that increasing the stake size, while reducing the selection probability to keep the expected payment to be 1 USD, reduces the difference in observed generosity compared to the conventional laboratory setting, where all participants receive the standard stake of 10 USD. Therefore, our results suggest that, at least in large-scale donation dictator game experiments, paying a large amount to a few participants incentivizes donation decisions more effectively than paying a small amount to many.

References

Ahles, Amelia, Marco A Palma, and Andreas C Drichoutis, "Testing the effectiveness of lottery incentives in online experiments," *American Journal of Agricultural Economics*, 2024.

Amir, Ofra, David G Rand, and Ya'akov Kobi Gal, "Economic games on the internet: The effect of $1 stakes," *PloS one*, 2012, *7* (2), e31461.

Anderson, Lisa R, Beth A Freeborn, Patrick McAlvanah, and Andrew Turscak, "Pay every subject or pay only some?," *Journal of Risk and Uncertainty*, 2023, *66* (2), 161–188.

Arechar, Antonio A, Simon Gächter, and Lucas Molleman, "Conducting interactive experiments online," *Experimental economics*, 2018, *21*, 99–131.

Baltussen, Guido, G Thierry Post, Martijn J Van Den Assem, and Peter P Wakker, "Random incentive systems in a dynamic choice experiment," *Experimental Economics*, 2012, *15*, 418–443.

Ben-Ner, Avner, Amit Kramer, and Ori Levy, "Economic and hypothetical dictator game experiments: Incentive effects at the individual level," *The Journal of Socio-Economics*, 2008, *37* (5), 1775–1784.

Berlin, Noemi, Emmanuel Kemel, Vincent Lenglin, and Antoine Nebout, "Paying none, some or all? between-subject random incentives and preferences towards risk and time," *Between-subject random incentives and preferences towards risk and time (August 01, 2024)*, 2024.

Brañas-Garza, Pablo, Antonio Espín, and Diego Jorrat, "Paying£ 1 or Nothing in Dictator Games: No Differences," *Available at SSRN 4723871*, 2024.

— , Diego Jorrat, Antonio M Espín, and Angel Sánchez, "Paid and hypothetical time preferences are the same: Lab, field and online evidence," *Experimental Economics*, 2023, *26* (2), 412–434.

Branas-Garza, Pablo, Diego Jorrat, Jaromír Kovářík, and María C López, "Hyper-altruistic behavior vanishes with high stakes," *PloS one*, 2021, *16* (8), e0255668.

Bühren, Christoph and Thorben C Kundt, "Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences," *Judgment and Decision Making*, 2015, *10* (2), 185–190.

Cartwright, Edward and Adam Thompson, "Using dictator game experiments to learn about charitable giving," *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 2023, *34* (1), 185–191.

Chapman, Jonathan, Erik Snowberg, Stephanie W Wang, and Colin Camerer, "Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters," Technical Report, National Bureau of Economic Research 2024.

— , — , — , and — , "Looming large or seeming small? Attitudes towards losses in a representative sample," *Review of Economic Studies*, 2024, p. rdae093.

Charness, Gary, Uri Gneezy, and Brianna Halladay, "Experimental methods: Pay one or pay all," *Journal of Economic Behavior & Organization*, 2016, *131*, 141–150.

— , — , and Michael A Kuhn, "Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics," *Journal of economic behavior & organization*, 2013, *91*, 93–100.

Chisadza, Carolyn, Nicky Nicholls, and Eleni Yitbarek, "The role of incentive structure in eliciting willingness to donate," *Economics Letters*, 2023, *224*, 111005.

Clot, Sophie, Gilles Grolleau, and Lisette Ibanez, "Shall we pay all? An experimental test of random incentivized systems," *Journal of behavioral and experimental economics*, 2018, *73*, 93–98.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 2007, *33*, 67–80.

Diederich, Johannes, Catherine C Eckel, Raphael Epperson, Timo Goeschl, and Philip J Grossman, "Subsidizing unit donations: Matches, rebates, and discounts compared," *Experimental Economics*, 2022, *25* (2), 734–758.

Engel, Christoph, "Dictator games: A meta study," *Experimental economics*, 2011, *14*, 583–610.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *The quarterly journal of economics*, 2018, *133* (4), 1645–1692.

_ , _ , _ , David Huffman, and Uwe Sunde, "The preference survey module: A validated instrument for measuring risk, time, and social preferences," *Management Science*, 2023, *69* (4), 1935–1950.

Forsythe, Robert, Joel L Horowitz, Nathan E Savin, and Martin Sefton, "Fairness in simple bargaining experiments," *Games and Economic behavior*, 1994, *6* (3), 347–369.

Frederick, Shane, "Cognitive reflection and decision making," *Journal of Economic perspectives*, 2005, *19* (4), 25–42.

Grech, Philip D and Heinrich H Nax, "Rational altruism? On preference estimation and dictator game experiments," *Games and Economic Behavior*, 2020, *119*, 309–338.

_ , _ , and Adrian Soos, "Incentivization matters: a meta-perspective on dictator games," *Journal of the Economic Science Association*, 2022, *8* (1), 34–44.

Hanaki, Nobuyuki, Keigo Inukai, Takehito Masuda, and Yuta Shimodaira, "Comparing behavior between a large sample of smart students and Japanese adults," *The Japanese Economic Review*, 2024, *75* (1), 29–67.

Hergueux, Jérôme and Nicolas Jacquemet, "Social preferences in the online laboratory: a randomized experiment," *Experimental Economics*, 2015, *18*, 251–283.

Larney, Andrea, Amanda Rotella, and Pat Barclay, "Stake size effects in ultimatum game and dictator game offers: A meta-analysis," *Organizational Behavior and Human Decision Processes*, 2019, *151*, 61–72.

Ozono, Hiroki and Daisuke Nakama, "Effects of experimental situation on group cooperation and individual performance: comparing laboratory and online experiments," *Plos One*, 2022, *17* (4), e0267251.

Prissé, Benjamin and Diego Jorrat, "Lab vs online experiments: No differences," *Journal of Behavioral and Experimental Economics*, 2022, *100*, 101910.

Schmelz, Katrin and Anthony Ziegelmeyer, "Reactions to (the absence of) control and workplace arrangements: experimental evidence from the internet and the laboratory," *Experimental Economics*, 2020, *23*, 933–960.

Sefton, Martin, "Incentives in simple bargaining games," *Journal of Economic Psychology*, 1992, *13* (2), 263–276.

Snowberg, Erik and Leeat Yariv, "Testing the Waters: Behavior across participant pools," *American Economic Review*, 2021, *111* (2), 687–719.

Umer, Hamza, "Effectiveness of random payment in Experiments: A meta-Analysis of dictator games," *Journal of Economic Psychology*, 2023, *96*, 102608.

_ **, Takashi Kurosaki, and Ichiro Iwasaki**, "Unearned endowment and charity recipient lead to higher donations: A meta-analysis of the dictator game lab experiments," *Journal of Behavioral and Experimental Economics*, 2022, *97*, 101827.

Walkowitz, Gari, "Dictator game variants with probabilistic (and cost-saving) payoffs: A systematic test," *Journal of Economic Psychology*, 2021, *85*, 102387.

A Additional Tables

Table A1: Regression Results: Comparison among five treatments with the expected payment of 100 JPY

	(1) Rate	(2) Zero	(3) Half	(4) Full
$T^{0.2}_{500}$	-0.055	-0.051	-0.127**	-0.072
	(0.052)	(0.067)	(0.051)	(0.052)
$T^{0.1}_{1000}$	-0.049	-0.108	0.073	-0.151***
	(0.050)	(0.066)	(0.061)	(0.048)
$T^{0.05}_{2000}$	-0.130***	-0.055	0.017	-0.191***
	(0.047)	(0.067)	(0.059)	(0.045)
$T^{0.01}_{10000}$	-0.098**	-0.147**	-0.088*	-0.152***
	(0.048)	(0.064)	(0.053)	(0.049)
CRT	0.014	-0.039**	0.048***	-0.010
	(0.012)	(0.019)	(0.015)	(0.012)
ρ (Risk)	0.021	-0.020	0.037	-0.011
	(0.044)	(0.065)	(0.052)	(0.039)
λ (Loss)	0.005	-0.006	0.015	-0.005
	(0.014)	(0.020)	(0.016)	(0.012)
δ (Time)	0.059	0.002	0.125	0.078
	(0.076)	(0.125)	(0.105)	(0.077)
β (Present)	-0.191	0.256	-0.220	-0.218*
	(0.164)	(0.225)	(0.169)	(0.114)
Age	0.005***	-0.005***	0.001	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.051*	0.093**	-0.104***	0.005
	(0.029)	(0.042)	(0.036)	(0.027)
R^2	0.09	0.05	0.07	0.15
N	502	502	502	502
p-value	0.043	0.190	0.002	0.000

[1] The base category is T^{1}_{100}. [2] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full. [3] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.
[4] p-value is from F or χ^2 tests of joint hypotheses that the coefficients on four treatment dummies are simultaneously zero.

Table A2: Regression Results: Effect of the stake size and the selection probability in incentivized treatments

	(1) Rate (OLS)	(2) Zero (Probit)	(3) Half (Probit)	(4) Full (Probit)
Payment Prob	-0.039	0.003	0.105*	-0.041
	(0.056)	(0.067)	(0.056)	(0.045)
Stake Size (1000)	-0.132***	-0.054	0.101**	-0.189***
	(0.042)	(0.056)	(0.046)	(0.040)
E[Payment] (1000)	-0.042	0.073	-0.131*	-0.090
	(0.067)	(0.096)	(0.076)	(0.086)
CRT	0.014	-0.023	0.025*	-0.007
	(0.012)	(0.017)	(0.013)	(0.012)
ρ (Risk)	0.126***	-0.124**	0.078*	0.098**
	(0.043)	(0.060)	(0.044)	(0.040)
λ (Loss)	-0.002	-0.001	-0.007	-0.003
	(0.013)	(0.017)	(0.013)	(0.012)
δ (Time)	0.026	0.037	-0.068	0.089
	(0.078)	(0.109)	(0.083)	(0.085)
β (Present)	-0.224*	0.500**	-0.269**	-0.118
	(0.131)	(0.206)	(0.127)	(0.132)
Age	0.006***	-0.006***	0.002**	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.053**	0.078**	-0.085***	-0.001
	(0.027)	(0.037)	(0.029)	(0.027)
R^2	0.12	0.05	0.04	0.13
N	701	701	701	701

[1] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full.
[2] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.
[3] The sample consists of T^{1}_{1000}, T^{1}_{100}, $T^{0.2}_{500}$, $T^{0.1}_{1000}$, $T^{0.01}_{1000}$, $T^{0.1}_{100}$, and $T^{0.01}_{100}$.

Table A3: Regression Results: Effect of stake size in hypothetical treatments

	(1) Rate (OLS)	(2) Zero (Probit)	(3) Half (Probit)	(4) Full (Probit)
Stake Size (1000)	-0.016***	0.010*	-0.008*	-0.016***
	(0.004)	(0.006)	(0.005)	(0.005)
CRT	-0.010	-0.011	0.001	-0.019
	(0.015)	(0.022)	(0.017)	(0.013)
ρ (Risk)	0.142**	-0.162**	-0.017	0.123**
	(0.055)	(0.078)	(0.060)	(0.049)
λ (Loss)	-0.024	0.031	0.009	-0.026*
	(0.015)	(0.022)	(0.018)	(0.014)
δ (Time)	0.033	-0.005	0.063	0.026
	(0.100)	(0.141)	(0.108)	(0.087)
β (Present)	0.071	-0.259	-0.076	0.035
	(0.144)	(0.216)	(0.167)	(0.140)
Age	0.005***	-0.006***	0.002	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.043	0.134***	-0.069*	-0.012
	(0.033)	(0.045)	(0.037)	(0.032)
R^2	0.11	0.06	0.03	0.12
N	439	439	439	439

[1] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full.
[2] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.
[3] The sample consists of T^0_{100}, T^0_{1000}, T^0_{2000}, and T^0_{10000}.

Table A4: Regression Results: Comparisons between hypothetical treatments and incentivized treatments with a small incentive while controlling for the expected payoff

	(1) T^0_{100} vs T^1_{100}	(2) T^0_{1000} vs $T^{0.1}_{1000}$	(3) T^0_{2000} vs $T^{0.05}_{2000}$	(4) T^0_{10000} vs $T^{0.01}_{10000}$
DV = Rate				
Hypo	0.023	-0.001	-0.016	-0.078**
	(0.055)	(0.046)	(0.039)	(0.039)
Controls	In	In	In	In
N	208	212	209	212
DV = Zero				
Hypo	-0.063	0.068	0.032	0.184***
	(0.065)	(0.063)	(0.065)	(0.061)
Controls	In	In	In	In
N	208	212	209	212
DV = Half				
Hypo	-0.044	-0.073	-0.113**	-0.021
	(0.056)	(0.057)	(0.047)	(0.044)
Controls	In	In	In	In
N	208	212	209	212
DV = Full				
Hypo	0.028	0.068*	0.022	-0.011
	(0.056)	(0.040)	(0.031)	(0.028)
Controls	In	In	In	In
N	208	212	209	212

[1] The base category is an incentivized treatment in all regressions. [2] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full. [3] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table A5: Regression Results: Comparison between hypothetical and incentivized treatments with 1% selection probability

	(1)	(2)	(3)
	T^{0}_{100} vs $T^{0.01}_{100}$	T^{0}_{1000} vs $T^{0.01}_{1000}$	T^{0}_{10000} vs $T^{0.01}_{10000}$
DV = Rate			
Hypo	-0.053	0.078*	-0.078**
	(0.056)	(0.046)	(0.039)
Controls	In	In	In
N	207	212	212
DV = Zero			
Hypo	-0.054	-0.045	0.184***
	(0.063)	(0.065)	(0.061)
Controls	In	In	In
N	207	212	212
DV = Half			
Hypo	0.002	0.051	-0.021
	(0.053)	(0.051)	(0.044)
Controls	In	In	In
N	207	212	212
DV = Full			
Hypo	-0.055	0.048	-0.011
	(0.059)	(0.041)	(0.028)
Controls	In	In	In
N	207	212	212

[1] The base category is an incentivized treatment in all regressions.
[2] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full. [3] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table A6: Regression Results: Comparison with Sure Payment Treatments

	(1) Rate	(2) Zero	(3) Half	(4) Full
Base: T^1_{100}				
T^0_{100}	0.024	-0.057	-0.044	0.030
	(0.055)	(0.066)	(0.056)	(0.056)
$T^{0.1}_{100}$	0.001	0.002	-0.127**	0.045
	(0.056)	(0.068)	(0.051)	(0.058)
$T^{0.01}_{100}$	0.076	-0.005	-0.046	0.082
	(0.057)	(0.066)	(0.056)	(0.059)
Controls	In	In	In	In
N	408	408	408	408
p-value	0.519	0.778	0.106	0.579
Base: T^1_{1000}				
T^0_{1000}	0.117***	-0.048	0.033	0.120***
	(0.040)	(0.066)	(0.054)	(0.034)
$T^{0.1}_{1000}$	0.120***	-0.123*	0.113*	0.063**
	(0.037)	(0.065)	(0.060)	(0.029)
$T^{0.01}_{1000}$	0.048	-0.013	-0.027	0.076**
	(0.038)	(0.067)	(0.052)	(0.031)
Controls	In	In	In	In
N	412	412	412	412
p-value	0.002	0.248	0.085	0.008

[1] The base category is an sure payment treatment in all regressions. [2] OLS estimates are reported for Rate. Probit average marginal effect estimates are reported for Zero, Half, and Full. [3] Heteroscedasticity-robust standard errors are in parentheses. Asterisks indicate significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$. [4] p-value is from F or χ^2 tests of joint hypotheses that the coefficients on three treatment dummies are simultaneously zero.

B Additional Figures

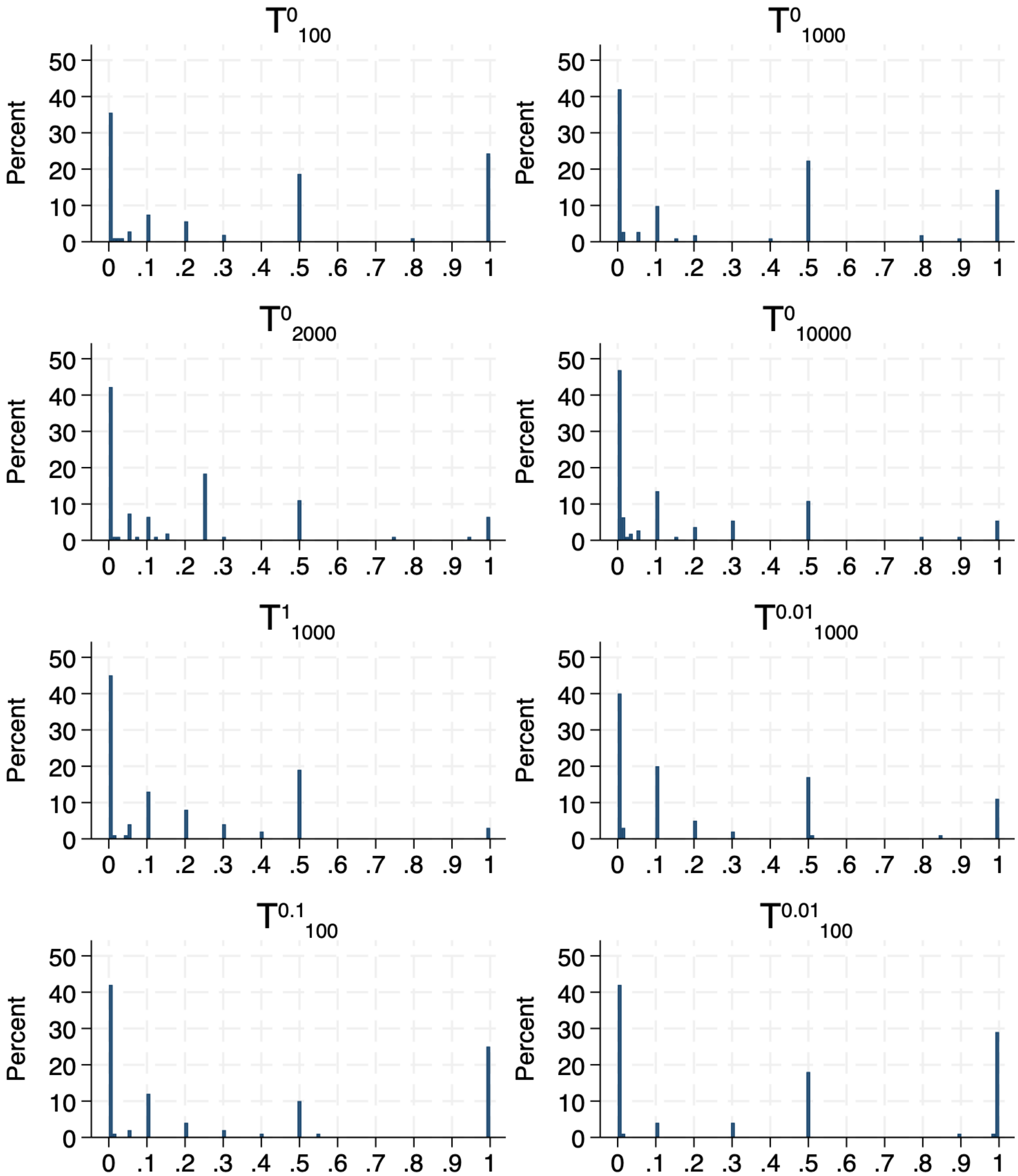

Figure B1: Histgrams of Donation Rates (Other Treatments)

C Experiment Instruction (English translation)

Consent Form: (3) Payment of points

- Additional points may be paid by lottery. However, the additional points that the winner will receive will increase or decrease depending on the content of the questions and the answers given. Please read the questions carefully and answer them. Those eligible for additional points will be paid with the above reward points (230 points).

Instruction

Section 5

In this section, you will be asked about how you will receive additional points when they are paid. Press the "Next" button to start.

Question (Instruction)

$[0 < X < 100]$[**Among the 100 people participating in this question,** X **people will be selected** by lottery to receive Y **yen (points) as additional points**.] $[X = 100]$[**All participants in this question will be paid** Y **yen (points) as additional points**.] $[X = 0]$[This is a question about a hypothetical setting, but please answer it as if it were described. Let's assume **you have been paid** Y **yen as additional points**.] Here, you will be asked to decide how much of this Y yen you want to donate to the Japan UNICEF Association's "UNICEF Fundraising."

About "UNICEF Fundraising" (from the official website of UNICEF Japan)

- This supports UNICEF's overall activities to protect children.
- This is an important donation campaign that supports UNICEF's overall activities in more than 150 countries and regions, including health, nutrition, water and sanitation, education, and child protection.

Payment of additional points

– $[0 < X < 100]$[X people will be selected by lottery from 100 people participating in this question, and additional points will be paid according to their answers.] $[X = 100]$[All participants in this question will be paid additional points.] $[X = 0]$[Assume that everyone will be paind additional points.]

– $[0 < X < 100]$[The X winners will actually receive Y yen.] $[X = 100]$[You will actually receive Y yen.] $[X = 0]$[Assume that you actually reveive Y yen.]

  * The portion of Y yen that is not donated will be paid as additional points.
  * The portion of the Y yen donated will be donated to the "UNICEF Fundraising" by MyVoice.com Co., Ltd. on your behalf.

– The announcement of the winner will be made with the payment of reward points.

Question

How much of the Y yen worth of additional points would you like to donate to the "UNICEF Fundraising"? Please enter the amount you want to donate in units of 1 yen between "0" yen and "Y" yen in **single-byte numbers** below.

c_i	yen