International Journal of Game Theory (2026) 55:27
https://doi.org/10.1007/s00182-026-00995-1

ORIGINAL PAPER

Paying £1(£5) or nothing in dictator games: unexpected differences

Pablo Brañas-Garza[1] · Antonio M. Espín[2] · Diego Jorrat[3]

Received: 11 November 2024 / Accepted: 17 March 2026 / Published online: 22 May 2026

Abstract
We conducted an online Dictator Game experiment (N=1195) to test three hypotheses about the role of monetary incentives in prosocial behavior. First, we examined whether real incentives of £1 reduce the dispersion of responses compared to hypothetical ones. Surprisingly, we found the opposite: hypothetical responses were less dispersed, with choices clustering around the egalitarian split. This pattern held in a replication (N=308) with higher stakes (£5), offering no support for the first hypothesis. Second, we tested whether real incentives—by involving actual monetary consequences—lead to more selfish decisions, as they are expected to reduce socially desirable responses. With £1 stakes, no significant differences emerged across conditions. However, when the stake was increased to £5, participants became more selfish under real incentives, supporting the second hypothesis only when the amount at stake is substantial. Third, we explored whether probabilistic payments trigger differential behavior. At low stakes, probabilistic incentives resembled real ones. But with higher stakes, real and probabilistic outcomes diverged, suggesting participants respond to expected value only when it is meaningful. Finally, in a separate study (N=299), we found that many participants facing standard hypothetical-payment instructions still expected real payments. Only explicit phrasing stating that "unfortunately, the money is not real" alleviated this confusion. This result underscores the importance of precise wording in experimental design and potentially explains why hypothetical treatments do not yield dramatically different results compared to real-money treatments.

Keywords Monetary incentives · Egalitarianism · Hyper-altruism · Selfishness · Dictator game

JEL Classification D64 · D91

Extended author information available on the last page of the article

1 Introduction

The debate on using real or hypothetical money to elicit experimental subjects' truthful responses is not new among economists and psychologists (see Camerer and Hogarth 1999). Experimental psychologists have a long tradition of not paying subjects (or not linking payments to their choices), under the argument that subjects are intrinsically motivated and engage in tasks with dedication and honesty (Camerer and Hogarth 1999). Conversely, experimental economists argue that without real incentives, subjects may be influenced by demand effects, social desirability, or lack of interest and attention, as financial rewards create a more realistic environment in the laboratory (Rosenboim and Shavit 2012; Zizzo 2010). As a result, non-incentivized choices may be biased or random (Carpenter et al. 2005).

The proliferation of online labor markets such as Amazon Mechanical Turk (MTurk; Paolacci and Chandler 2014) or Prolific Academic (PA; Palan and Schitter 2018) has introduced a new dimension to this debate, as researchers are increasingly conducting online economic experiments with low-stakes incentives. While these experiments typically involve the use of real money, the stakes are often set at small amounts such as \$1, £1, or even less. This raises the question of whether the use of low stakes in online settings is really incentivizing anything and, moreover, whether it leads to different behavioral outcomes as compared to the use of purely hypothetical rewards.

This paper investigates the case of social preferences: do hypothetical rewards elicit different behavior than real but small monetary rewards in online money allocation experiments? While several scholars have addressed this question previously, most have compared behavior in Dictator Games (DG) with no stakes and traditional stakes (ranging between \$5 and \$10) in laboratory settings. Amir et al. (2012) remains, to date, the only study comparing hypothetical and real low-stakes decisions in the context of online experiments.[1] The continued expansion of online experimental research—particularly with representative and non-'standard' samples—highlights the growing importance of this question.

Returning to the question of incentives themselves, the results of existing studies are somewhat mixed. Some suggest that real payoffs induce selfishness while hypothetical settings promote egalitarian choices (Sefton 1992; Forsythe et al. 1994; Dana et al. 2007; Amir et al. 2012; Clot et al. 2018), which is consistent with the existence of social desirability concerns or demand effects that are alleviated when giving has a real cost. Most closely related to our study, Amir et al. (2012) in an online experiment conducted on MTurk, find significant differences in both the average level and the dispersion of giving in the Dictator Game between hypothetical and real but small incentives (\$1), with real stakes leading to smaller and less dispersed donations. However, not all have found these results: Ben-Ner et al. (2008) found no significant differences in DG between real payments and purely hypothetical scenarios. Bühren

[1] Amir et al. (2012) study four games (Dictator Game, Ultimatum Game, Trust Game, and Public Goods Game) and two incentive schemes (Real and Hypothetical), with a total sample of $N=756$ in MTurk. Participants are randomly assigned to one of the eight experimental arms of the 4×2 design.

and Kundt (2015) reported similar results using three mini-dictator games, as did Locey et al. (2011) in the context of social discounting.[2]

To provide a more definitive answer on the causal effect of real vs. hypothetical low stakes incentives on social preferences we ran a well-powered DG online experiment. We recruited 1,195 subjects using PA. Each participant had to decide how to split £1 with another anonymous participant and was randomly assigned to one of three treatments with equal probability (1/3): the Real money treatment (R), where every participant received a payment; the Between-Subjects Random Incentivized System or BRIS treatment (B), where 1 out of every 10 subjects received a payment; and the Hypothetical treatment (H), where no payment was provided. Moreover, to gain deeper insight into the impact of incentives—more specifically, to assess whether participants perceive £1 as a small amount and thus behave as if the decision were hypothetical—we conducted a replication in which the monetary stakes were increased fivefold, to £5. This experiment was also run on PA ($N=308$), with the same payment conditions as in the original design: R, B or H.

What is the purpose of including the BRIS treatment in the study? It was introduced to test whether (the size of) expected earnings matter or it is the mere existence of monetary incentives that causes any shift in behavior with respect to a purely hypothetical condition. In addition to the above, experiments using probabilistic incentives have become increasingly common, both within and beyond the field—and in particular, the one-out-of-ten payment scheme is a frequently used protocol (e.g., Charness et al. 2016; Exadaktylos et al. 2013). According to a recent meta-analysis of DG lab experiments with student samples comparing Real and BRIS conditions, the two treatments with monetary incentives R and B should result in similar behavior (Umer 2023; see Clot et al. 2018; for a direct test). Yet this has not been tested in a DG online experiment with non-student participants.[3]

Based on previous literature, our main predictions were that (i) treatment H would result in more dispersed, noisier data than B and R, owing to more erratic or low-effort decision-making in the absence of real incentives (Camerer and Hogarth 1999); (ii) treatment H would lead to more generous donations than B and, especially, R, as the cost of socially desirable behavior is zero; and (iii) donations in treatments B and R would be rather similar, as both involve real (expected) payoffs (Umer 2023).

To analyze the data, we adopted a conservative and comprehensive approach. In addition to comparing the mean and SD of giving across treatments, as most of the literature has done, we also examined specific patterns of behavior: selfishness (giving = 0%), egalitarianism (giving = 50%), and hyper-altruism (giving = 100%). Furthermore, we introduced different levels of "trembling", i.e., small deviations from these three distribution rules, to account for potential errors in decision-making. This approach allows us to gain a more nuanced understanding of participants' behavior across treatments and to better capture any variation in social preferences that might

[2] There are other papers testing the effect of hypothetical vs. real payoffs on different measurements. For instance, Brañas-Garza et al. (2021a) tested incentives for risk taking in three countries, whereas Brañas-Garza et al. (2023) did similarly for time preferences in three settings: Lab, Field and Online.

[3] Interestingly, Ahles et al. (2024) find no differences in bidding between a fully incentivized condition (R) and either 1/10 or 1/100 BRIS in willingness-to-pay online experiments.

otherwise be masked by small decision-making errors, especially when stakes are low. Our results can be summarized as follows.

The first hypothesis posited that real incentives would reduce the dispersion of responses compared to hypothetical ones, as the presence of actual consequences is expected to induce more consistent, deliberate behavior. Surprisingly, we found the opposite: responses in the hypothetical condition were less dispersed, with choices clustering tightly around the egalitarian split.[4] This unexpected pattern held in the £5 replication, offering no support for Hypothesis 1. This result contradicts what Amir et al. (2012) observed for $1 stakes on MTurk.

The second hypothesis proposed that real incentives would lead to more selfish decisions, under the assumption that actual monetary consequences reduce socially desirable responding. With £1 at stake, we observed no significant differences in mean giving across treatments, thus contradicting Amir et al. (2012) as they found less giving with $1 stakes than in the hypothetical condition. However, when the stake was increased to £5, participants in the real incentive condition behaved more selfishly than those in the hypothetical (in line with Amir et al. 2012; yet for larger stakes) or probabilistic conditions. These results support Hypothesis 2, but only when the amount at stake is sufficiently large to be perceived as meaningful.

The third hypothesis focused on whether probabilistic payments differ behaviorally from certain real incentives. At low stakes, behavior under BRIS resembled that in the real condition, which in turn was similar to the Hypothetical one. This suggests two possible interpretations: either participants in the BRIS condition behave like those in the real one, or—more plausibly—participants in the real condition perceive the low incentive (£1) as too small to affect their behavior, effectively treating it as if it were hypothetical, regardless of whether incentives are certain or probabilistic. This second explanation gains support when stakes are increased to £5: participants under Real incentives begin to diverge from those in the BRIS condition, indicating that they respond to the higher incentive, while BRIS participants (facing a 1-in-10 chance of £5, i.e., an expected value of 50 pence) still behave as if the incentive were hypothetical. This supports Hypothesis 3 but only when stakes are small.

Independently of the above, another possible explanation is that participants may not fully understand the concept of “hypothetical”.[5] To explore this, we conducted an additional experiment (also in PA, N = 299) where participants were given one of three types of instructions: in one, the word “hypothetical” was explicitly used to refer to payments, as in our previous experiments; in another, the decision was framed as an imaginary situation; and in the third, it was clearly stated that no real

[4] Note, however, that one could reach the opposite prediction in the specific context of the dictator game if errors are disregarded and a sufficient proportion of individuals is not selfish but, for example, inequity averse. Under real incentives, selfish individuals would donate zero whereas inequity averse one would donate 50%. When monetary incentives are removed, the dominant response for both selfish and inequity averse individuals could be to donate exactly 50% (or even 100%, as socially desirable responses entail no cost), which might reduce dispersion depending on the proportions of each type. We thank Reviewer #2 for highlighting this alternative interpretation. Yet, dispersion would, in any case, increase if we consider errors associated with careless responding in the hypothetical treatment, which is the argument from Camerer and Hogarth (1999) that gives rise to our starting hypothesis.

[5] This idea was suggested by Reviewer #1, to whom we are very grateful.

money would be earned. The results suggest that participants are quite naïve about payment conditions. In the version where the word ‘hypothetical’ was used to refer to the money involved, only about 30% understood that no real payment was involved. The (apparent) misunderstanding was even more pronounced in the imaginary framing, where just 10% correctly recognized the absence of monetary incentives. When the absence of real incentives was explicitly emphasized, participants were more likely to understand that there was no actual money at stake, as about 70% of them reported zero probability of getting paid for real.

To summarize our contribution in a nutshell: First, we revisit the distinction between hypothetical and real (low stakes) incentives and introduce a treatment with more conventional laboratory stakes (£5) to explore the role of stake size for the results. We show that some of the null effects observed under £1 stakes disappear when the amount increases: specifically, selfish behavior rises and egalitarian responses decline in the real-money condition, while giving in the BRIS and hypothetical treatments remains largely unchanged. This suggests that low-stakes real treatments may resemble hypothetical ones not because monetary incentives are irrelevant, but because they are too small to be perceived as meaningful — a nuance often overlooked in existing meta-analyses and in studies that rely on a single stake level.

Second, we address for the first time a largely ignored methodological concern by conducting a survey experiment that examines how participants interpret hypothetical-payment instructions. A majority of respondents report some positive probability of being paid, even when the instructions explicitly state that payments are not real. Moreover, donation behavior in the Dictator Game is related to these beliefs, highlighting the role of perceived incentives in shaping behavior beyond the formal wording of instructions. These findings suggest that further research on how experimental instructions are interpreted may be valuable for improving experimental design.

In this sense, our findings complement and extend those of Amir et al. (2012) in two ways. First, by introducing a higher-stakes condition, we show that differences between hypothetical and real incentives seem to emerge once stakes are sufficiently salient, helping to reconcile mixed results obtained under very low incentives. We may speculate that if $1 was salient enough for Amir et al.‘s participants from MTurk but £1 is not for our PA participants a decade later, it is because either the two platforms are associated with different earnings expectations (i.e., thresholds for what is considered as meaningful/salient) or expectations have changed over time. Second, by explicitly eliciting participants’ beliefs about payment under hypothetical instructions in online experiments, we show that perceived incentives may differ from the intended design and that this misperception can attenuate differences in average giving between hypothetical and real conditions.

The rest of the paper is organized as follows. Section 2 describes the protocol, the sample, our working hypotheses, and the empirical strategy for the main experiment. Section 3 presents its results, Sect. 4 reports the replication using a £5 stake, and Sect. 5 presents the results of a new experiment that investigates the phrasing used in the instructions regarding hypothetical payments. Section 6 discusses the findings and concludes.

2 Materials and methods

2.1 Protocols and sample

The experiment was conducted using Prolific Academic (PA), a platform to recruit participants for online studies. There are certain advantages of running experiments using PA (and similar sites): it reduces costs and allows researchers to recruit a large and heterogeneous sample (as opposed to standard experimental subjects, see Exadaktylos et al. 2013). However, the downside is the lack of control, as we do not know what subjects are doing when they participate in the experiment, and participants may be professional subjects or "lab rats" (Guillen and Veszteg 2012). Yet, recent evidence suggests that data from online experiments using these platforms are reliable (Horton et al. 2011; Rand 2012; Arechar et al. 2018) and that the lack of control is not so problematic (Prissé and Jorrat 2022).

The experiment was published on PA on July 15th 2021 at 21:30 CET and ended four hours later, having gathered 1,195 participants. The experiment consisted of three parts: a discounting task to elicit time preferences following the design of Coller and Williams (1999), a task to elicit risk preferences based on Holt and Laury (2002), and a Dictator Game (DG) task.[6] See the experimental instructions in Appendix B for details.

We only invited UK residents to participate since this is the country with the largest number of potential participants in the platform. Additionally, we pre-screened the subjects based on having available data on education, gender, and different socio-economic questions to avoid losing observations with respect to the control variables. Table 1 provides summary statistics for these variables and subjects' choices in the experimental tasks preceding the DG.

In the DG, subjects were asked to divide £1 between themselves and another randomly selected anonymous participant, in £0.1 increments. We implemented a dual-role protocol with a known probability of being the recipient or the dictator of 50%.

All the participants received a fixed participation fee of £1.2, which was adjusted to a 10-minute experiment according to PA's recommendations. Those selected for real payments (including those randomly selected in the BRIS treatment) received a bonus payment based on their decision in the DG. Participants were fully informed of their payment scheme.

At the beginning of the experiment, participants were randomly assigned to treatments R, B, or H, each with a probability of *1/3*. The resulting sample sizes for each treatment were n_R=380, n_B=406, n_H=409. The average age was 32.2 years, 63.1% females. Regarding education, most participants had completed either secondary education (40.2%) or had an undergraduate degree (46.1%). Table 3 of the Appendix reveals that the treatments were homogenous in terms of sociodemographic charac-

[6] The randomization happened at the beginning of the experiment (time preferences) and treatment assignment remained the same along the entire session. The three tasks appeared in the same order being the DG always the last. More information on the time preferences experiment can be found in Brañas-Garza et al. (2023), studies III and IV. We also note that the experiment and the working hypotheses were not pre-registered, as this belonged to a broader study involving time and risk preference tasks.

Table 1 Summary statistics of participants' characteristics

	Obs	Mean	Std. dev.	Min	Max
Female	1195	0.631	0.483	0	1
Age	1195	32.238	11.924	18	77
Education	1194	2.905	1.514	0	6
SES	1195	5.147	1.569	1	10
Charity	1195	2.529	1.424	1	7
Risky choices	1195	5.433	2.355	0	10
Patient choices	1195	9.064	6.385	0	20

Education is a categorical variable (taking values from 0 to 6 for simplicity, from no formal education to doctorate degree) which refers to the highest education level. SES reflects the Socio-economic status using the position in the income ladder (scale from 1 to 10). Charity refers to a self-reported categorical variable that reflects different amounts of donations (in ascending order) made in the last year. Risky choices refers to the number of risky options chosen in the Holt-Laury task (in which participants had to choose between a safer and a riskier lottery), while patient choices refers to the number of later-larger allocations in the time discounting task (in which participants had to choose between a sooner smaller amount of money and a later but larger amount)

teristics and subjects' choices in the previous experimental tasks, using Westfall and Young's (1993) p-values correction for multiple testing.

2.2 Working hypotheses

The hypotheses to be tested arise from previous literature and are related to whether monetary incentives yield different giving behavior as compared to hypothetical incentives when low stakes are at play. First, we expected less noise and therefore less dispersion in the data when money is involved, as people may not take hypothetical scenarios seriously and may randomize their responses more (Camerer and Hogarth 1999; Amir et al. 2012). In addition, donations in the treatments involving actual money were expected to be smaller than in the hypothetical condition because being generous (or giving the impression of being generous as a socially desirable behavior) in the former involves a cost while in the latter is free (Amir et al. 2012). Finally, we expected similar donations in the two treatments involving actual money (Umer 2023), which would imply that some probability of being paid for real is enough to counteract social desirability or demand effects, as it is often implicitly thought given the number of studies using BRIS-like incentives.

In summary, we test the following three main hypotheses:

Hypothesis 1 *Monetary incentives (R&B) with low stakes cause subjects' donation decisions to be less dispersed than using hypothetical incentives (H).*

Hypothesis 2 *Monetary incentives (R&B) with low stakes lead subjects to donate less money than using hypothetical incentives (H).*

Hypothesis 3 *Donations in the two monetary conditions (R&B) do not differ.*

Table 2 Types of behavior by trembling

No trembling	Trembling	Large trembling
selfish ($g=0$)	t.selfish ($g\leq 0.1$)	2t.selfish ($g\leq 0.2$)
egalitarian ($g=0.5$)	t.egalitarian ($0.4\leq g\leq 0.6$)	2t.egalitarian ($0.3\leq g\leq 0.7$)
saint ($g=1$)	t.saint ($0.9\leq g\leq 1$)	2t.saint ($0.8\leq g\leq 1$)

2.3 Empirical strategy

In addition to studying differences in averages and SDs between the three treatments, we focus on specific types of donation behavior. We start by considering the extreme cases:

- Selfish ($giving=0$),
- Egalitarian ($giving=0.5$) and,
- Hyper-altruistic or "saint" ($giving=1$).

In line with the literature, we expect hypothetical incentives to reduce purely selfish choices and to increase egalitarian and hyper-altruistic behavior, as the cost of socially desirable actions is zero.

In addition, we introduce "trembling hand" cases, allowing first a decision-making error of $t=\pm 0.1$, and then a larger error of $2t=\pm 0.2$. From this, we derive nine measures which are described in Table 2.

3 Results

Before moving on to the main analyses, it is important to mention some descriptive results from our experiment. Interestingly, the mean donation in our sample is around 42% of the pie, which is comparatively rather high. Note that Engel's (2011) meta-analysis of lab experiments shows an average donation of 28.3% of the pie, while Brañas-Garza et al. (2018) found an average of 30.8%. Amir et al. (2012), on the other hand, found 43.8% and 33.2% mean giving in the hypothetical and real conditions, respectively. Our data show that the fraction of subjects giving half of the pie is substantial (>60%), but even more remarkable is the very low fraction, about 10%, of purely selfish choices (compared the 30% found in Brañas-Garza et al. 2018; using MTurk).

Even if other factors may play a role,[7] it seems reasonable to think that the size of the incentive helps explain why average giving is around 40% instead of 30%. All of this is consistent with studies showing that DG giving decreases with higher stakes (see Brañas-Garza et al. 2021b; Larney et al. 2019). While all of the above is true, we also should not rule out the possibility that PA participants are simply kinder than participants from other (online) platforms.

3.1 Testing predictions: dispersion of the data

Panel A of Fig. 1 shows the distribution of giving for each treatment. While the three distributions are statistically similar (we do not reject the null hypothesis that the three distributions are equal in a Kolmogorov–Smirnov test, $p > 0.50$), it seems that donations are more concentrated around the equal distribution (giving = 0.5) in the hypothetical treatment H than in the two monetary treatments R and B. Indeed, when comparing the dispersion of the data between H and R or between H and B, a Levene's test rejects the null hypothesis of equal variance in both cases, with H displaying lower SD (both $p < 0.01$).

Overall, our data fully contradict Hypothesis 1, as we observe that the dispersion is lower for H than for the treatments involving actual money. The comparison of H vs. R & B (i.e., the two incentivized treatments combined) yields the same conclusion (*p < 0.01*).

Result 1: Hypothetical donations are less dispersed than real and one-out-of-ten BRIS incentivized donations in the £1-DG.

3.2 Testing predictions: average donations and behavioral types

Before moving to the regression analysis, Panel B of Fig. 1 displays the average donations across treatments. Although we will test this below using regressions, average giving is nearly identical in the three cases. In fact, the t-tests confirms this result (*p > 0.75 for all the cases*).

Figure 2 presents the regression results for average giving and the nine behavioral types defined earlier (which will be analyzed in detail in the next subsection). All the regressions control for age, gender, education, risky and patient choices, SES status, and self-reported donations to charity. Each point represents the estimated coefficient

[7] For example, one might conjecture that the relatively high level of giving observed in our experiment could be driven by the dual-role protocol, in which participants face a 50% probability of being either the dictator or the recipient. Under a Rawlsian veil-of-ignorance argument, such uncertainty could, in principle, induce more egalitarian choices. However, if individuals expect their counterpart to behave selfishly, role uncertainty may instead lead them to behave more selfishly themselves. More generally, existing evidence suggests that dual-role protocols do not increase generosity or egalitarian behavior. Grossman et al. (2020) document even lower average transfers under dual-role in allocation games. Consistent with this pattern, Andreoni and Bernheim (2009) argue that role uncertainty can provide moral cover for selfish behavior, allowing individuals to justify self-interested choices while preserving a positive self-image. Moreover, in a previous study where we also implemented a dual-role protocol, average giving amounted to 32.4% of a €5 endowment (Brañas-Garza et al. 2013). Taken together, this evidence suggests that the high prevalence of equal splits observed in our experiment is unlikely to be driven by the dual-role protocol itself.

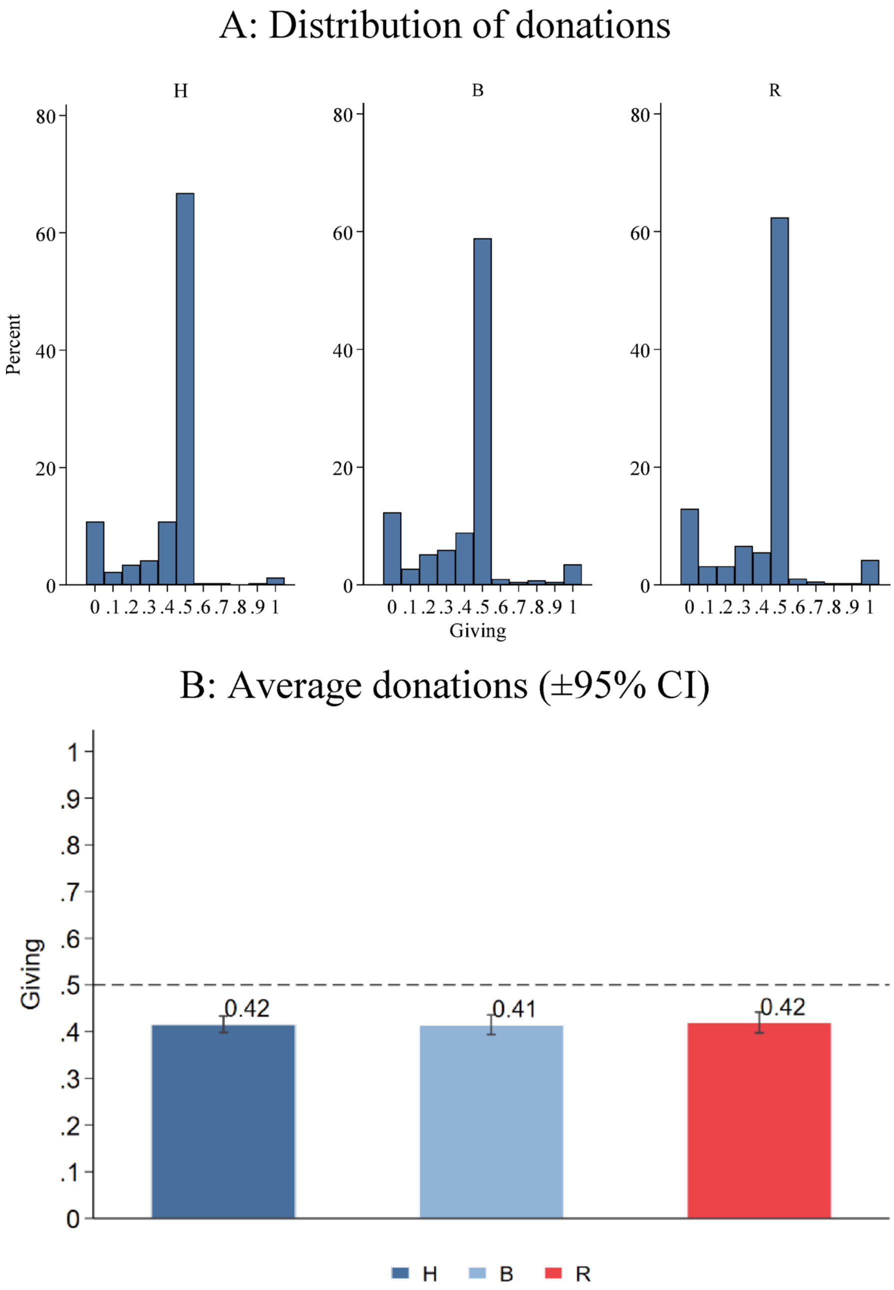


Fig. 1 £1-DG average donations by treatment. The dashed line represents the equal division and the numbers are the mean of giving for each group

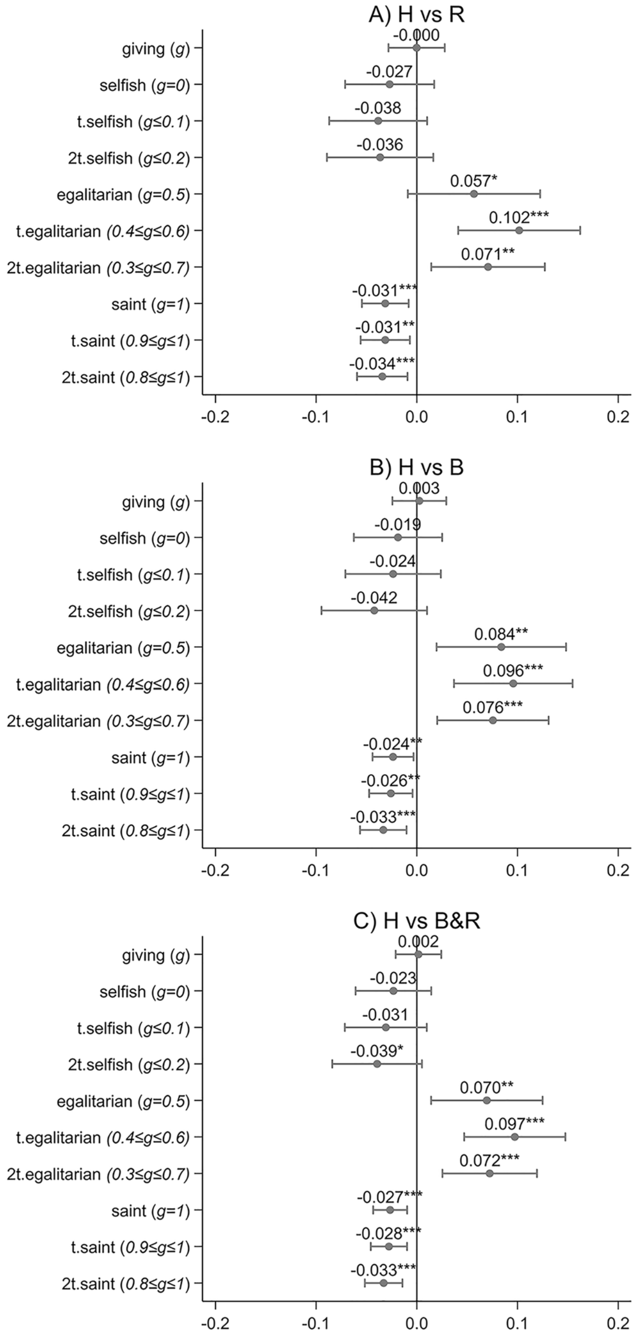


Fig. 2 Regression results for £1-DG. Point estimates denote the coefficient of H (±95% CI). All the regressions control for age, gender, education, risky and patient choices, SES status, and self-reported donations to charity. Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

of the dummy variable H (±95% CI) denoting the hypothetical treatment. Panels A, B, and C of Fig. 2 compare H against R, B, and R and B combined (R&B), respectively. Complete regression results are presented in Tables 4, 5, and 6 in the Appendix. In all cases, we further computed the p-value of the coefficient of H adjusted for multiple testing (hereafter, adj-p) using the free step-down resampling method of Westfall and Young (1993), following the Jones et al. (2019) procedure.

As could be inferred from panel B of Fig. 1, the regression analysis summarized in Fig. 2 indicates that there are no significant differences between treatments in average donations: *giving* (*g*) (all $p > 0.75$ and *adj-p* > 0.80; see Fig. 2 and Tables 4 to 6 for regression results).

These results therefore do not support Hypothesis 2, which predicts higher donations in the hypothetical treatment:

Result 2: Subjects with monetary incentives do not behave more selfishly than those with hypothetical incentives in the £1-DG.

In the following analysis, we want to see if there is more to the distributions beyond the mean, and we focus on the "types". Here we analyze the nine behavioral types arising from the three main categories (selfish, egalitarian, and saint; see Table 2) and their two trembling cases. As can be seen in Fig. 2, the coefficient of H is never significant at the 5% level for the selfish or the trembling selfish types, when compared against either R, B, or R&B (all nine comparisons yield negative effects between 2 and 4% points, all $p > 0.08$ and *adj-p* > 0.19).

For the egalitarian case, H is significant in eight out of nine comparisons (all $p < 0.05$ and *adj-p* < 0.06), with positive effects ranging from 6 to 10% points. The exception is the non-trembling definition when compared against R, which, although marginally significant at the 10% level ($p = 0.09$), is not significant after adjusting for multiple testing (*adj-p* $= 0.26$). Finally, for the saint category, H is always negative and significant, although the effects are very small (between 2 and 3% points, all $p < 0.05$).

The regressions results suggest that when payments are hypothetical, subjects' choices tend to cluster around the center, i.e. the egalitarian distribution. This result in not new but, in contrast to what others have shown in the past (e.g., Forsythe et al. 1994), the concentration of choices around the 50/50 split in the hypothetical condition is not driven by a statistically significant reduction in selfish behavior—although coefficients for selfish choices are negative—but rather by a robust decline of hyper-altruistic behavior ("saints").

Result 2b: Hypothetical donations do not yield a different proportion of selfish allocations but yield more egalitarian and less hyper-altruistic allocations than incentivized decisions in the £1-DG.

3.3 Testing predictions: real vs. BRIS incentives

Now we analyze Hypothesis 3, which states that there are no differences in behavior when we use real versus probabilistic incentives. As we have already seen in panel B of Fig. 1, R and B have similar average ($p > 0.75$, see also Fig. 3 and Table 7). Similarly happens with dispersion – SDs do not differ ($p = 0.32$). Hence, we can conclude that Hypothesis 3 is not rejected.

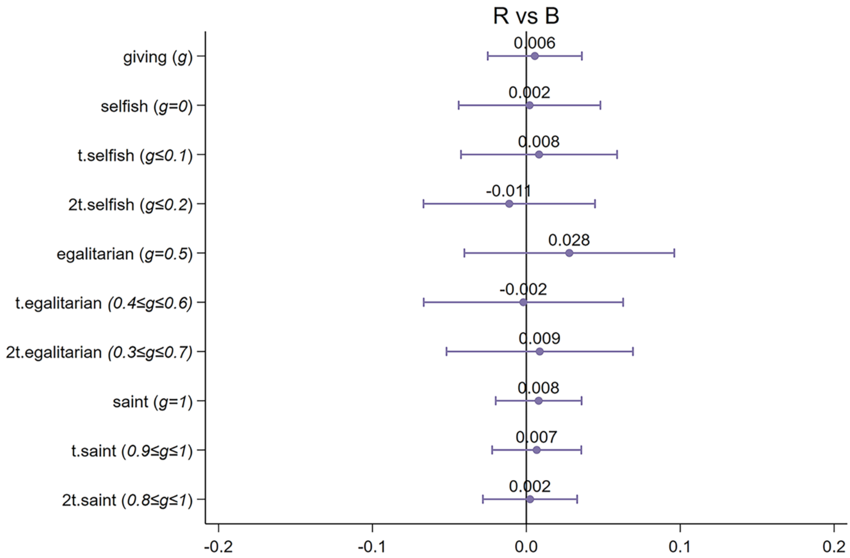


Fig. 3 Regression results for £1-DG. Point estimates denote the coefficient of R (±95% CI). All the regressions control for age, female, education, risky and patient choices, SES status and the self-reported donations to charity. Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Regarding behavioral types, Fig. 3 examines the estimates for the coefficient of R (vs. B) on the nine types considered. Full estimates can be found in Table 7 in the Appendix. Our results clearly indicate that paying one randomly selected participant out of every ten does not make any difference with respect to paying all participants. All nine comparisons are largely insignificant (all $p>0.40$ and adj-$p>0.90$). Therefore, we conclude:

Result 3: Real and one-out-of-ten BRIS incentives yield similar average donations, with similar dispersion and proportions of selfish, egalitarian, and hyper-altruistic choices in the in the £1-DG.

4 Replication for £5 pies

On June 26th 2025, we conducted a replication (REP1) of the original experiment with the only difference being the size of the pie to be divided: £5 instead of £1. The instructions were the same, the platform was the same (PA), and the sample selection criteria remained unchanged. Unlike the original £1 experiment, the £5 replication did not include time or risk preference tasks, and the Dictator Game was presented immediately after the initial instructions.[8] The sample consisted of 308 participants who were assigned to treatment H, B, or R with a probability of *1/3*. The resulting sample sizes for each treatment were n_H=103, n_B=103 and n_R=102. The average age of participants was 44.8 years, with 43.8% being female. Table 8 in Appendix C shows that all covariates are generally balanced across treatments.[9]

Figure 4 shows the distribution of responses for each of the treatments. Before discussing the differences between treatments, it is worth noting that in both conditions H and B, the average donation is approximately £2 out of £5 — that is, 40% of the pie. This average is quite like what we observed in Fig. 1. In other words, PA participants appear to be fairly "altruistic," at least compared to other experimental subjects (see Exadaktylos et al. 2013).

Using the data from this replication, we can test our original hypotheses. Our first hypothesis states that responses under real incentives are less dispersed than hypothetical ones. If we look at Panel A of Fig. 4, the distributions in the H and B conditions are statistically indistinguishable from one another (Kolmogorov–Smirnov test, $p = 1.00$), but both differ significantly from the distribution in the Real (R) condition ($p < 0.01$ in both cases). As before, we use Levene's tests to compare the dispersion of responses across conditions. When comparing H and R, the test rejects the null hypothesis of equal variances, with H displaying significantly lower SD ($p < 0.01$). A similar result holds for the comparison between B and R, with B also showing significantly lower variance ($p < 0.01$). In contrast, the variances in H and B are statistically indistinguishable ($p = 0.76$).

Therefore, consistent with Result 1, the replication confirms that variance is smaller in the hypothetical condition. Hypothesis 1 is thus rejected: incentivized responses are not less dispersed than hypothetical ones.

The second hypothesis refers to the idea that participants with monetary incentives (R & B) will donate less than those in the hypothetical condition (H). Panel B of Fig. 4 shows that dictators in the Real condition indeed donate less (£1.57, 31.4% of the endowment) than those in the Hypothetical condition (£2.05, 41.1%), and this difference is statistically significant (t-test, $p < 0.01$). Notably, these percentages are similar to those reported by Amir et al. (2012) in their online experiment, and to the average donation observed in the Hypothetical condition of our £1 main study, suggesting some stability in hypothetical giving across designs and stake lev-

[8] This change allows us to test whether the original findings might have been influenced by task order or by the broader context of the session. As shown below, the results for the BRIS and hypothetical treatments are consistent with the original experiment, while behavior in the real-money condition becomes more selfish — suggesting that stake size, rather than task order, is the key driver of the differences observed.

[9] Except for socioeconomic status: participants in the B treatment report a slightly lower status, and this difference is statistically significant at the 5% level. In addition, the average age in the £5 replication is higher than in the main £1 study (44 vs. 32 years). Both studies were conducted on Prolific using identical prescreening criteria, but they were run four years apart and differ substantially in sample size ($N \approx 300$ vs. $N \approx 1{,}200$). We interpret this age difference as a contextual feature of recruitment rather than as a design-driven difference. Importantly, participants did not know ex ante that stakes were higher in the £5 study, and posted hourly rates were comparable across experiments. The only procedural difference between the studies was the launch timing (evening vs. afternoon, UK time), which may also have contributed to the observed age gap.

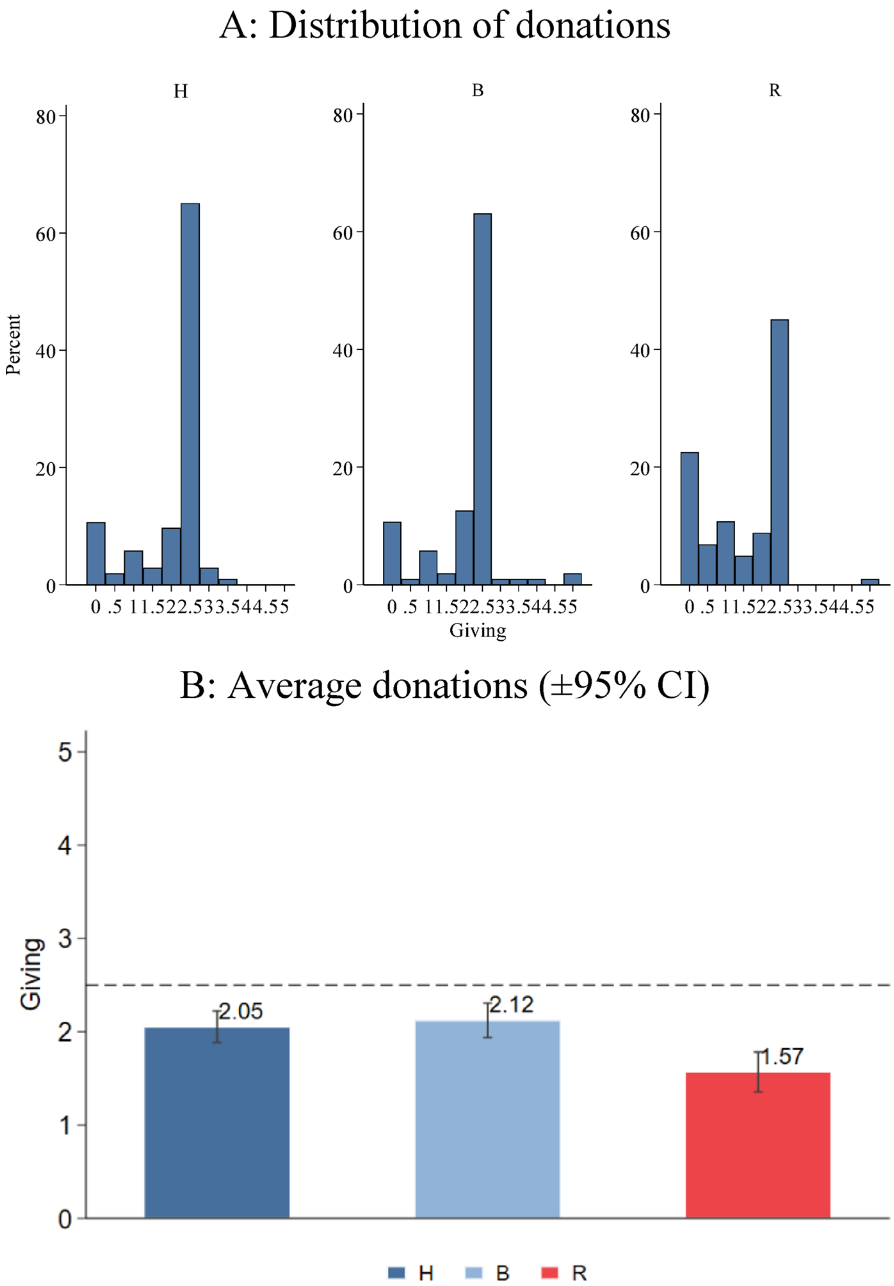


Fig. 4 £5-DG average donations by treatment. The dashed line represents the equal division and the numbers are the mean of giving in each group

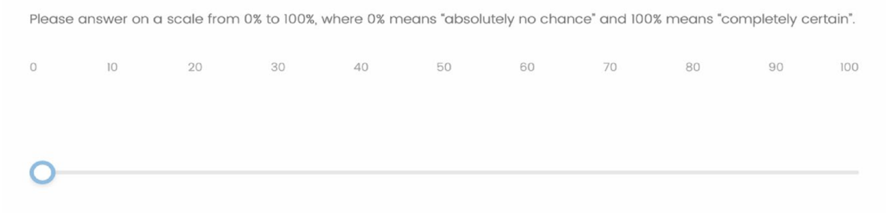


Fig. 5 Question about the probability of being paid (slider). The slider initially appeared at 0. However, participants were not allowed to proceed without moving it at least once

els. However, participants under the BRIS mechanism do not behave as predicted by H2; instead, they behave as if they were in the Hypothetical condition, donating on average £2.12 (42.4%; $p = 0.59$), which is significantly more than in R ($p < 0.01$). This mirrors the previous variance results. Tables 9 to 12 replicate the regression analyses and confirm these results.

In other words, the dictators under Real incentives group behave as expected – thus supporting H2 – but those under BRIS do not seem to perceive the monetary component of their decisions and end up donating the same as the hypothetical group – against H2. Therefore, Hypothesis 3 — which states that donations in the two monetary conditions (R & B) do not differ — is clearly rejected.

In Appendix C, we replicate the same comprehensive analysis conducted previously (Figs. 2 and 3) using data from the new £5-stakes sample. When we compare behavior across the full set of measures reported in Table 2, we see that observations from the R condition lean more toward selfish solutions to the problem, while the H participants behave in a more egalitarian way. In other words, all the metrics shown in Figure 8 are consistent with the results from Fig. 4. Besides, the comparison between B and R (Figure 9) provides compelling evidence against H3. When comparing across studies, the qualitative ordering of treatments is broadly similar, but the margins differ across stake levels: egalitarianism in H increases at the expense of saints in the £1 study but at the expense of selfishness in the £5 study. In addition, contrasts involving B change: H and B are harder to distinguish at £5, whereas R diverges more clearly from B at £5 than at £1.

5 Do subjects understand hypothetical payoffs?

In the final section of the paper, we ask whether *participants in H truly understand the hypothetical nature of the payment* — that is, whether the instructions lead them to believe that there is no real money involved. More generally, we examine how participants interpret hypothetical-payment instructions and how such interpretations shape behavior in the Dictator Game. If enough participants misinterpret that there is real money in hypothetical treatments, that could reduce differences between treatments.

Taking papers in the literature that employ hypothetical payments (Johnson and Bickel 2002; Ben-Ner et al. 2008, 2009; Locey et al. 2011; Bürhen & Kundt 2015; Bechler et al. 2015; Thielmann et al. 2016), we classified their instructions as follows:

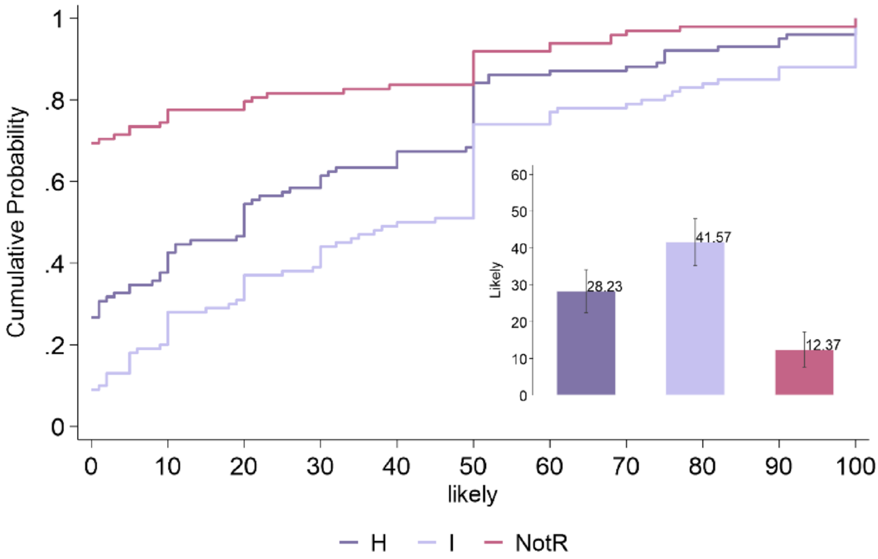


Fig. 6 Expected likelihood that payments are real: CDF by treatment. The small, embedded figure shows the mean (±95% CI) response in each treatment

- H: Participants are told that payments are hypothetical.
- I: An imaginative framing is used (e.g., "imagine you are making a decision where you divide a pie…").
- NotR: Participants are told something like "Unfortunately, the money you earn is not real…".

On July 9th 2025, we launched this survey experiment (REP2) on PA with 299 new participants (n_H=101, n_I=100, n_{NotR}=98). Participants faced a DG identical to the one used in the £5 condition, but the instructions followed one of the H, I, or NotR formats (see Appendix D for the exact wording used in each treatment). After reading the DG instructions and before making their decision, participants were asked:

> *How likely do you think it is that either you or the other person will actually receive real money from this allocation task?*

The Cumulative Distribution Function (CDF) for each of the treatments is shown in Fig. 6. Inside the figure (bottom right), we also display the average response along with its 95% confidence interval. Recall that if participants understood that the payment was not real, they should report a number close to zero — though we accept there may be some noise due to the mechanism (e.g., trembling-hand errors).

Figure 6 leaves little doubt on this matter. Participants in the "*imagine that you…*" (I) treatment appear not to grasp the hypothetical nature of the task. On average, they assign a 42% probability to being using real money, and only 9% of them expected no real payment. The statement "*payments are hypothetical*" (H) also seems insufficiently clear, as participants still report an average of 28% chance of being playing with real money. It is also concerning that less than 30% of participants in that condition expect no real payments. Only the wording "*Unfortunately*, *the money you earn is not real…*" (NotR) seems to achieve the intended effect: on average, these participants assign just a 12% probability to the existence of real payments, and 69% believe there is no actual money. In sum, the share of individuals displaying correct beliefs is 27% in H, 9% in I, and 69% in NotR.

And what about our own instructions for the £1 and £5 experiments? They fall under the category where participants are told that payments are hypothetical (H). Based on the result we just observed, less than 30% of participants in experiments using the term "hypothetical" assumed they were playing with no real money. This leads us into a somewhat murky territory, as in neither the £1 nor the £5 experiment do we truly know what participants expected in terms of payment (regardless of the hypothetical nature of the task).

Since for these 299 participants we also observe actual donations in the (supposedly) hypothetical Dictator Game, we next examine whether these differences across wording conditions are reflected in donation behavior. In the NotR condition, participants donate on average about 24% more than in the H condition, while there is no statistically significant difference between the H and I conditions (see Figure 10 in Appendix E).[10] These patterns are consistent with the evidence from Fig. 6, suggesting that the NotR wording is associated with a clearer perception of the hypothetical nature of the task and, consequently, with higher giving.

In addition, for these participants we observe both their perceived probability of receiving real money and their actual donation in the Dictator Game, allowing us to jointly analyze beliefs and behavior. Using participants' self-reported beliefs, regardless of what type of instructions they saw, we classified them into three groups: Hypothetical, Intermediate Beliefs, and Real. Participants who reported a 0% likelihood of being playing with real money were classified as "Hypothetical" ($L=0 \longrightarrow$ L:"H"), those who reported 100% as "Real" ($L=1 \longrightarrow$ L:"R"), and those in between as "Intermediate Belief" ($0<L<1 \longrightarrow$ L:"IB"). The subgroup sizes were quite unbalanced – L:"H" = 104 subjects (34.8%), L:"IB" = 177 (59.2%), L:"R" = 18 (6.0%). Similarly as we did before for giving, we applied a trembling-hand criterion allowing a 10% margin of error to L:"H", L:"IB" and L:"R", leading to L:*t*."H" = 147 (49.2%), L:*t*."IB" = 128 (42.8), L:*t*."R" = 24 (8.0%).[11]

[10] Average giving across wording conditions was: μ_H = £1.64, μ_I = £1.82, and μ_{NotR} = £2.04 (see Figure 10 In Appendix E). Welch pairwise *t*-tests indicate that the difference between the H and NotR conditions is statistically significant ($p=0.01$), while differences between H and I and between I and NotR are not ($p>0.20$).

[11] Here thus, $0\leq L\leq 0.1 \longrightarrow$ L:*t*."H"; $0.1<L<0.9 \longrightarrow$ L:*t*."IB"; $0.9\leq L\leq 1 \longrightarrow$ L:*t*."R". Figure 14 of Appendix E shows a locally weighted scatterplot smoothing (LOWESS) between giving and the perceived probability of receiving real money. It shows a negative relationship between payment beliefs and giving, with a marked negative slope at high belief levels ($L\geq 60$). In complementary regressions with belief intervals, individuals reporting very high certainty of payment ($L\geq 90$) give significantly less ($p=0.04$), while those reporting very low probabilities ($L\leq 10$) exhibit marginally different behavior ($p=0.06$). These patterns justify grouping extreme beliefs separately from intermediate ones.

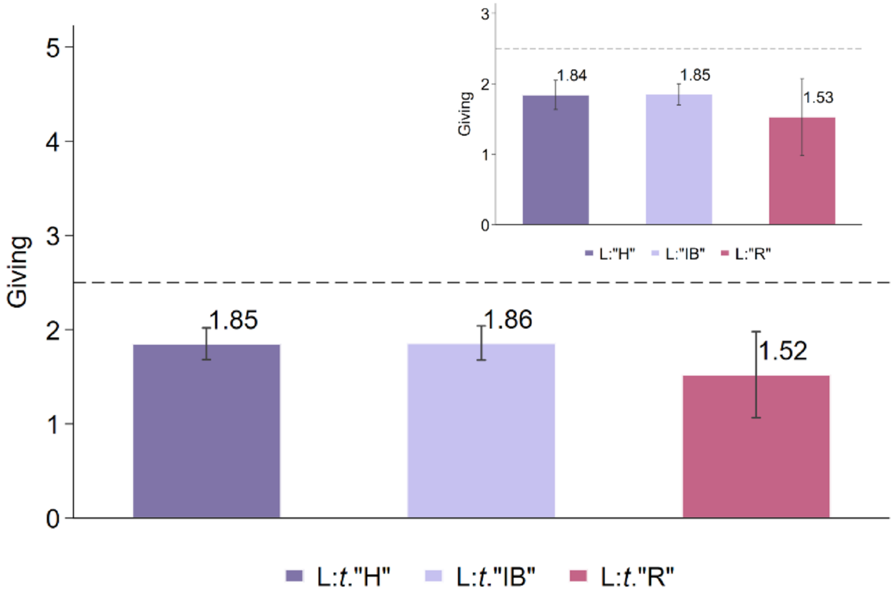


Fig. 7 £5-DG average (±95% CI) donations by types of expected real-payment likelihood. The dashed line represents the equal division. The small, embedded figure shows the non-trembling data

For the central part of Fig. 7, we use the trembling-hand classification (L:*t.*“H”, *t.*“IB”, *t.*“R”), as it includes more balanced groups. In the top-right corner, we replicate the same graph using the non-trembling data (L: “H”, “IB”, “R”). The resulting pattern mirrors that observed in Fig. 4B in qualitative terms: participants classified as L:*t.*“H” and L:*t.*“IB” donate similar amounts on average, while those classified as L:*t.*“R” donate less. However, differences across belief-based classifications are not statistically significant ($p > 0.16$), likely due to the small number of observations – particularly in the L:*t.*“R” group – and the resulting large standard errors. Given the small number of observations in the L:*t.*“R” group and the heterogeneity of beliefs among L:*t.*“IB” participants (ranging from 10% to 90%), these results should be interpreted with caution.

As an additional analysis, Appendix E reports further comparisons across replications, namely the £5 Dictator Game study (REP1) and the survey experiment (REP2). Figures 12 and 13 compare the distribution and average donation across three groups: Hypothetical participants in REP1 (H), participants classified as Hypothetical based on near-zero payment beliefs in REP2 (L:*t.*“H”), and Real-payment participants in REP1 (R). The results are twofold.

First, comparing H in REP1 with L:*t.*“H” in REP2, we find that the two groups display very similar behavioral patterns. Their distributions are not statistically different (Kolmogorov–Smirnov test, $p=0.55$), dispersion does not differ significantly (Levene test, $p=0.11$), and the proportion of participants choosing the egalitarian split is similar across groups (two-sample proportion test, $p=0.69$). Average giving is slightly lower in L:*t.*“H” (REP2) than in H (REP1), with the difference being marginally significant (Welch two-sample *t*-test, $p=0.09$). Overall, these results suggest that donation behavior in the belief-based Hypothetical group in REP2 closely resembles that observed in the Hypothetical condition of the £5 DG study (REP1).[12]

Second, we compare belief-based Hypothetical participants in REP2 (L:*t.*"H") with Real-payment participants in REP1 (R) as a descriptive robustness check of Results 1 and 2. Figures 12 and 13 show that the distribution of giving differs across these groups (Kolmogorov–Smirnov test, $p=0.04$). Regarding dispersion, L:*t.*"H" does not exhibit higher variability than R and is descriptively less dispersed, although the difference is not significant (Levene test, $p=0.14$). In addition, average donations in the R group are significantly lower than in the L:*t.*"H" group (Welch two-sample *t*-test, $p=0.04$). Taken together, these comparisons provide a descriptive robustness check of Results 1 and 2, as participants classified as Hypothetical based on beliefs donate more, and are not more dispersed, than those facing real monetary incentives, consistent with the patterns documented in the £5 Dictator Game.

The main result of this section is unexpected not because participants who believe they will be paid behave accordingly, but because a substantial share of participants hold such beliefs despite receiving hypothetical-payment instructions. In this sense, the novelty of the finding lies more in the beliefs themselves than in the resulting behavior. A natural interpretation is that participants differ in how they understand terms such as "hypothetical," or in how clearly they associate these instructions with the absence of real monetary consequences. While one might conjecture that such beliefs are driven by inattention, this explanation appears unlikely. An attention check included at the end of the study was passed by 295 out of 299 participants (98.66%), and failing the attention check is not correlated with stated payment beliefs ($r=0.06$, $p=0.25$; see Appendix for details). Accordingly, we attribute these beliefs to differences in interpretation rather than to lack of attention.

To further assess whether the patterns observed in Fig. 7 reflect a systematic phenomenon, we also elicited second-order (incentivized) beliefs[13] where participants guessed what others thought regarding whether the payments might be real, in particular, the modal response. Figure 11 replicates the same analysis as Fig. 7 but categorizes participants according to these second-order beliefs: (G: "H", "IB", "R") – and replicated for those with trembling (G: *t.*"H", *t.*"IB", *t.*"R"). The resulting pattern closely mirrors that obtained with first-order beliefs: participants classified as H and B donate similar amounts on average, while those classified as R donate less, albeit with large standard errors due to the small sample size. In addition, participants' own beliefs are strongly correlated with their second-order beliefs ($r=0.62, p<0.01$). Together, these results suggest that the belief-based patterns observed in Fig. 7 are not incidental.

[12] This comparison should be interpreted with caution. When comparing average giving in the Hypothetical condition of the £5 DG study (REP1) with the Hypothetical condition defined by the original instructions (H) in the survey experiment (REP2), average giving is higher in REP1 than in REP2 (Welch two-sample *t*-test, μ_{REP1} = £2.05 vs. μ_{REP2} = £1.64, $p<0.01$. Despite the two samples being highly comparable in terms of observable characteristics (see Table 13 in Appendix E), REP2 was not designed as a quantitative replication of the £5 DG study: the decision task is embedded in a survey context and is preceded by explicit questions eliciting payment beliefs, which may affect donation levels.

[13] We paid £1 for correct predictions and 50p if incorrect within a 10% margin error.

In summary, a substantial proportion of participants—except in the condition explicitly stating that there is no real money—believe that their decisions may involve real payments, despite the hypothetical nature of the instructions. While the presence of such beliefs is noteworthy in itself, participants' behavior in the Dictator Game is related to these beliefs: those who believe they will be paid with certainty behave similarly to participants in real-payment conditions and high-stakes, whereas those who believe payments are hypothetical behave similarly to participants in hypothetical conditions (or in low-stakes ones). Taken together, these findings underscore the importance of understanding how participants interpret hypothetical-payment instructions and how perceived incentives shape behavior in experimental settings.

6 Discussion and conclusions

We conducted an online experiment on Prolific with a large sample ($N = 1,195$) and subject-level randomization to test three well-established hypotheses in the discipline: The first hypothesis (H1) refers to the idea that outcomes in incentivized experiments exhibit lower dispersion (i.e., variance) compared to hypothetical ones. The second hypothesis (H2) suggests that monetary incentives lead participants to reveal their true preferences, whereas in the absence of incentives, individuals may present themselves as more prosocial without facing any real cost. In addition, we formulate a third hypothesis (H3), which compares certain (real) monetary incentives with probabilistic ones, expecting no differences.

We also conducted two additional studies. In the first one ($N = 303$), we examine the impact of using a £5 endowment instead of £1 to test for a stake-size effect. In the second study ($n = 299$), we investigate how different wordings or phrasings of the payment instructions in hypothetical treatments affect participants' estimated probability of being playing with real money.

First, we aimed to test whether donation decisions in a Dictator Game with low stakes, but real monetary incentives are less dispersed than those with the same low stakes but hypothetical payoffs, where participants are not motivated to tell the truth and can respond arbitrarily. Our experiment does not support this hypothesis. Instead, we found the opposite pattern: hypothetical responses in the DG are less dispersed, since they concentrate closer to the egalitarian distribution. This finding contrasts with Amir et al. (2012), who report lower dispersion under real incentives in an online Dictator Game with $1 stakes conducted on MTurk. Importantly, when we replicate our design using a £5 pie, we find exactly the same result as with the £1 pie. The hypothetical donations are less dispersed than the real (monetary) ones. Therefore, Hypothesis 1 is not supported by our data. An alternative explanation (i.e., not based on noise/errors like Hypothesis 1) for our result is that, if enough participants are inequity averse, then removing the cost of giving might reduce dispersion. For example, donations in the Hypothetical condition may concentrate on focal points such as 50% for both inequity averse and selfish individuals (who would donate zero with real incentives; see Footnote 4), as a socially desirable response.

The second hypothesis tested is that low-stakes real incentives lead to more honest decisions, as participants have their own money at stake. In other words, donating money in a Dictator Game with real incentives is not "cheap talk" but rather a true reflection of the participant's preferences. Thus, we should expect that when incentives have economic consequences, participants will be more selfish, as "appearing as a good person" comes at a cost, even with low stakes. However, our data do not support this hypothesis since we found no significant differences in average giving across treatments – similarly to other papers in the literature (see Ben-Ner and Kramer, 2008; Bühren and Kundt 2015). Furthermore, our results suggest that when participants face a hypothetical problem of distributing a small amount of money, they tend to use an egalitarian distribution rule more often. This finding aligns with previous literature (see Forsythe et al. 1994 and Dana et al. 2007). However, when real or probabilistic incentives are introduced, a small fraction of subjects become more hyper-altruistic —a result that contradicts other studies showing that real payments increase selfish behavior (see Amir et al. 2012; Clot et al. 2018).

Yet, the replication with the £5 pie changes the results dramatically and in the expected direction. When participants are faced with a stake five times larger, the proportion of selfish individuals increases substantially, and the percentage of egalitarians drops sharply — but this only occurs among those with real payments. We do not observe this effect in either the hypothetical payment condition or in the BRIS condition. In other words, making decisions involving £5 no longer seems to be cheap talk (as it might be with £1). This leads us to reconsider the results we had for Hypothesis 2 with the £1 endowment. It now appears that simply increasing the amount from £1 to £5 provides strong support for the hypothesis that real incentives lead to more honest decisions. In this sense, our £5 results reconcile our findings with Amir et al. (2012): while they report lower giving under real incentives even at $1 stakes, our evidence suggests that such differences emerge once stakes are sufficiently salient. The divergence between Amir et al.'s findings and ours might be rationalized if participants from different platforms or different moments in time (our experiments were conducted a decade later in PA instead of MTurk) have different reference levels to consider stakes to be salient enough.

We can thus summarize that the stake size helps explain both the null results observed in the £1 experiment and the change in behavior seen in the £5 condition. In the former case, the size of the payments (for both the dictator and the recipient) may have led participants to perceive the paid treatment as almost quasi-hypothetical. In the latter case, with five times more money at stake, participants appeared to think twice before donating. This interpretation is also highly consistent with previous literature showing that altruism tends to vanish as stakes increase (see Brañas-Garza et al. 2021b; Larney et al. 2019). In other words, if the payment is too low, participants may not perceive a meaningful difference between hypothetical and real incentives (Gneezy and Rustichini 2000).

Third, Hypothesis 3 asks whether paying subjects in probability makes any difference with real payments. Our results for £1 pie clearly indicate that paying one randomly selected participant out of every ten does not make any difference with respect to paying all participants. Here, the replication with the £5 pie changes what we observed previously. With the £1 pie, we found that the R and B payment conditions produced the same outcomes (and both were similar to H). With the £5 pie, we see that the B treatment still yields outcomes identical to H, but no longer to R. This is because the distribution in the R condition has shifted with the introduction of a stake five times larger. Perhaps the simplest explanation is that 10% (the expected payment in B) of £5 remains a very low incentive, and participants do not perceive it as meaningfully different.

In short, our results suggest the existence of a threshold in payment size for participants to consider it relevant. A £1 payment is perceived as minor, while £5 is apparently seen as substantial enough. Given that the expected payments in the BRIS treatments with £1 and £5 stakes are only 10 and 50 pence respectively—both below the threshold, which is placed between £1 and £5—it seems likely that these amounts are simply not large enough to be perceived as meaningful. Of course, this leaves an open question—and a broad avenue for future research—regarding how participants perceive and respond to different payment levels. This is an area that warrants much more detailed and careful investigation.

Beyond testing the three original hypotheses of this paper, we also set out to explore whether participants truly understand that payments are hypothetical. Using instruction styles from previous studies, we classify them into three types: (H) explicitly hypothetical, (I) using indirect “imagine” language, and (NotR) explicitly stating that the money is not real. We ran an experiment on Prolific with 299 subjects and found that participants often misunderstood the hypothetical nature of the task. In both the “imagine” and “payments are hypothetical” conditions, a large fraction of participants still believed they might earn real money. Only the explicit wording “the money you earn is not real” effectively corrected this, with most participants expecting to receive nothing.

What is noteworthy is not that participants who believe they will be paid behave accordingly, but rather that these beliefs arise under hypothetical-payment instructions. Consistent with this interpretation, donation behavior in the Dictator Game is related to perceived incentives: participants who believe payments are real tend to give less, whereas those who believe payments are hypothetical tend to give more, although average donation levels are not identical across the different studies considered in the paper. Overall, our findings highlight the sensitivity of experimental outcomes to how payment schemes are described and understood. They underscore the importance of ensuring that hypothetical-payment instructions are clearly interpreted by participants, and caution against assuming that such instructions are universally understood in the intended way. Future research could build on these findings by systematically testing alternative ways of conveying experimental instructions and by examining how participants’ interpretations of such instructions vary across contexts and populations. Importantly, the participants’ confusion reported here suggests that observed differences between hypothetical and real treatments may often represent lower-bound estimates of the true differences in previous studies.

One important limitation of our study concerns the generalizability of the findings, particularly across experimental platforms. The level of giving observed in our £1 real treatment is notably higher than what is typically reported in lab or MTurk studies (e.g., Engel 2011; Brañas-Garza et al. 2018), and the proportion of participants donating exactly half of the endowment is unusually large. While our £5 replication suggests that some of these patterns are driven by the low absolute and relative value of the stake, it remains possible that Prolific participants differ systematically from those in other samples, for example in terms of attentiveness or preference for socially desirable responses. Future work comparing the same design across platforms, using matched samples, would help assess the extent to which these results reflect platform-specific behavior or more generalizable effects.

Appendix A: detailed statistical analyses

See Tables 3, 4, 5, 6, 7.

Table 3 Balance across treatments

	$Mean_H$	$p(R\text{-H})^ç$	$p(\text{B-H})^ç$
Female	0.621	0.986	0.946
Age	31.909	0.946	0.988
Education	2.941	0.808	0.988
SES	5.147	0.946	0.988
Risky choices	5.496	0.306	0.988
Patient choices	8.628	0.160	0.960
Charity	2.535	0.946	0.960

Baseline: Hypothetical (H). *Education* is a categorical variable (taking values from 0 to 6 for simplicity, from no formal education to doctorate degree) which refers to the highest education level. SES reflects the *Socio-economic status* using the position in the income ladder (scale from 1 to 10). *Charity* refers to a self-reported categorical variable that reflects different amounts of donations (in ascending order) made in the last year. *Risky choices* refers to the number of risky options chosen in the Holt-Laury task (in which participants had to choose between a safer and a riskier lottery), while *patient choices* refers to the number of later-larger allocations in the time discounting task (in which participants had to choose between a sooner smaller amount of money and a later but larger amount)

ç Inference was made regressing each control variable on H and using Westfall and Young adjusted p-values for multiple testing

Table 4 Regression analysis for the H vs. R comparison

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (g)	Selfish ($g=0$)	t.selfish ($g \leq 0.1$)	2t.selfish ($g \leq 0.2$)	Egalitarian ($g=0.5$)	t.egalitarian ($0.4 \leq g \leq 0.6$)	2t.egalitarian ($0.3 \leq g \leq 0.7$)	Saint ($g=1$)	t.saint ($0.9 \leq g \leq 1$)	2t.saint ($0.8 \leq g \leq 1$)
H	-0.000	-0.027	-0.038	-0.036	0.057*	0.102***	0.071**	-0.031***	-0.031**	-0.034***
	(0.014)	(0.023)	(0.025)	(0.027)	(0.033)	(0.031)	(0.029)	(0.012)	(0.012)	(0.013)
Westfall and Young p-value	0.994	0.374	0.262	0.324	0.256	0.008***	0.060*	0.034**	0.050*	0.028**
Age	0.002***	-0.002***	-0.003***	-0.004***	0.006***	0.005***	0.003***	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.042***	-0.090***	-0.085***	-0.069**	0.070**	0.066**	0.062**	0.006	0.005	0.007
	(0.015)	(0.025)	(0.027)	(0.029)	(0.035)	(0.032)	(0.030)	(0.011)	(0.012)	(0.012)
Education	-0.007	0.016**	0.023***	0.014	-0.011	-0.017	-0.017	0.003	0.003	0.003
	(0.005)	(0.008)	(0.009)	(0.009)	(0.012)	(0.011)	(0.010)	(0.004)	(0.005)	(0.005)
SES	-0.018***	0.024***	0.025***	0.035***	-0.040***	-0.043***	-0.031***	-0.004	-0.004	-0.003
	(0.005)	(0.008)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)	(0.004)	(0.004)	(0.004)
Risky choices	0.001	-0.003	-0.001	-0.002	-0.000	-0.000	0.002	0.000	-0.001	-0.000
	(0.003)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)
Patient choices	0.001	-0.003	-0.003	-0.004	0.006**	0.007***	0.005**	-0.001*	-0.002*	-0.002**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Charity	0.006	0.001	-0.002	-0.010	0.023*	0.015	0.009	0.002	0.002	0.001
	(0.006)	(0.009)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.005)	(0.005)	(0.005)
Constant	0.410***	0.133**	0.155**	0.205***	0.510***	0.655***	0.739***	0.055*	0.063**	0.056*
	(0.038)	(0.066)	(0.069)	(0.072)	(0.089)	(0.081)	(0.076)	(0.029)	(0.030)	(0.031)
Observations	789	789	789	789	789	789	789	789	789	789
R-squared	0.047	0.050	0.050	0.046	0.056	0.066	0.043	0.015	0.015	0.016

Linear regression estimates. Robust standard errors in parentheses

Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5 Regression analysis for the H vs. B comparison

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (g)	Selfish ($g=0$)	t.selfish ($g\leq 0.1$)	2t.selfish ($g\leq 0.2$)	egalitarian ($g=0.5$)	t.egalitarian ($0.4\leq g\leq 0.6$)	2t.egalitarian ($0.3\leq g\leq 0.7$)	Saint ($g=1$)	t.saint ($0.9\leq g\leq 1$)	2t.saint ($0.8\leq g\leq 1$)
H	0.003	-0.019	-0.024	-0.042	0.084**	0.096***	0.076***	-0.024**	-0.026**	-0.033***
	(0.014)	(0.022)	(0.024)	(0.027)	(0.033)	(0.030)	(0.028)	(0.010)	(0.011)	(0.012)
Westfall and Young p-value	0.878	0.574	0.532	0.252	0.050*	0.016**	0.038**	0.084*	0.084*	0.030**
Age	0.003***	-0.001*	-0.003***	-0.003***	0.005***	0.003**	0.002	0.001*	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.032**	-0.088***	-0.087***	-0.100***	0.110***	0.136***	0.128***	-0.028**	-0.028**	-0.028**
	(0.015)	(0.025)	(0.027)	(0.029)	(0.035)	(0.032)	(0.031)	(0.013)	(0.013)	(0.014)
Education	-0.008*	0.010	0.014	0.011	-0.019	-0.014	-0.010	-0.001	-0.002	-0.001
	(0.005)	(0.008)	(0.008)	(0.009)	(0.011)	(0.011)	(0.010)	(0.004)	(0.004)	(0.004)
SES	-0.018***	0.013	0.020**	0.031***	-0.045***	-0.040***	-0.026***	-0.006	-0.005	-0.005
	(0.005)	(0.008)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)	(0.004)	(0.004)	(0.004)
Risky choices	0.002	0.001	-0.003	-0.006	0.017**	0.012*	0.009	-0.002	-0.002	-0.003
	(0.003)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.002)	(0.003)	(0.003)
Patient choices	0.001	-0.002	-0.001	-0.003	0.004*	0.005**	0.005**	-0.001	-0.002*	-0.002*
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Charity	0.012**	-0.013	-0.019**	-0.025***	0.044***	0.038***	0.028***	-0.001	-0.002	-0.003
	(0.005)	(0.008)	(0.009)	(0.009)	(0.012)	(0.011)	(0.010)	(0.005)	(0.005)	(0.005)
Constant	0.381***	0.178***	0.220***	0.299***	0.420***	0.552***	0.616***	0.080**	0.064*	0.085**
	(0.041)	(0.066)	(0.068)	(0.072)	(0.088)	(0.082)	(0.078)	(0.034)	(0.038)	(0.040)
Observations	814	814	814	814	814	814	814	814	814	814
R-squared	0.056	0.029	0.037	0.053	0.073	0.075	0.055	0.026	0.037	0.034

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6 Regression analysis for the H vs. R&B comparison

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (g)	Selfish ($g=0$)	t.selfish ($g\leq0.1$)	2t.selfish ($g\leq0.2$)	Egalitarian ($g=0.5$)	t.egalitarian ($0.4\leq g\leq0.6$)	2t.egalitarian ($0.3\leq g\leq0.7$)	Saint ($g=1$)	t.saint ($0.9\leq g\leq1$)	2t.saint ($0.8\leq g\leq1$)
H	0.002	-0.023	-0.031	-0.039*	0.070**	0.097***	0.072***	-0.027***	-0.028***	-0.033***
	(0.012)	(0.019)	(0.021)	(0.023)	(0.028)	(0.026)	(0.024)	(0.009)	(0.009)	(0.010)
Westfall and Young p-value	0.902	0.322	0.250	0.192	0.056*	0.002***	0.010**	0.042**	0.04**2	0.010**
Age	0.002***	-0.002***	-0.003***	-0.004***	0.005***	0.003***	0.002**	0.001	0.001**	0.001**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Female	0.040***	-0.095***	-0.093***	-0.091***	0.080***	0.107***	0.099***	-0.011	-0.010	-0.008
	(0.013)	(0.021)	(0.022)	(0.024)	(0.029)	(0.027)	(0.025)	(0.011)	(0.011)	(0.011)
Education	-0.007*	0.011*	0.017**	0.011	-0.014	-0.015*	-0.013	0.001	0.001	0.002
	(0.004)	(0.007)	(0.007)	(0.008)	(0.010)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
SES	-0.014***	0.013**	0.017**	0.026***	-0.038***	-0.036***	-0.023***	-0.004	-0.004	-0.003
	(0.004)	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)
Risky choices	0.002	-0.003	-0.003	-0.003	0.005	0.004	0.004	-0.001	-0.001	-0.001
	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)
Patient choices	0.001	-0.003	-0.002	-0.003*	0.005**	0.006***	0.005**	-0.001	-0.001*	-0.001*
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Charity	0.006	-0.003	-0.006	-0.011	0.033***	0.023**	0.015*	-0.001	-0.002	-0.003
	(0.005)	(0.007)	(0.008)	(0.008)	(0.010)	(0.009)	(0.009)	(0.004)	(0.004)	(0.004)
Constant	0.373***	0.202***	0.225***	0.282***	0.471***	0.603***	0.667***	0.056**	0.043	0.052
	(0.033)	(0.054)	(0.057)	(0.060)	(0.074)	(0.069)	(0.065)	(0.028)	(0.031)	(0.032)
Observations	1194	1194	1194	1194	1194	1194	1194	1194	1194	1194
R-squared	0.043	0.034	0.038	0.041	0.052	0.057	0.039	0.012	0.016	0.017

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7 Regression analysis for the R vs. B comparison

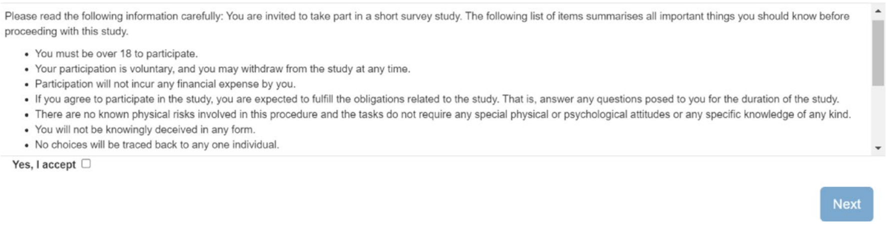
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (g)	Selfish ($g=0$)	t.selfish ($g \leq 0.1$)	2t.selfish ($g \leq 0.2$)	Egalitarian ($g=0.5$)	t.egalitarian ($0.4 \leq g \leq 0.6$)	2t.egalitarian ($0.3 \leq g \leq 0.7$)	Saint ($g=1$)	t.saint ($0.9 \leq g \leq 1$)	2t.saint ($0.8 \leq g \leq 1$)
R	0.006	0.002	0.008	-0.011	0.028	-0.002	0.009	0.008	0.007	0.002
	(0.016)	(0.023)	(0.026)	(0.028)	(0.035)	(0.033)	(0.031)	(0.014)	(0.015)	(0.016)
Westfall and Young p-value	0.992	0.998	0.996	0.988	0.912	0.998	0.996	0.974	0.978	0.978
Age	0.003***	-0.003***	-0.004***	-0.004***	0.004***	0.003**	0.002	0.001*	0.002**	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.045***	-0.105***	-0.104***	-0.105***	0.064*	0.119***	0.108***	-0.011	-0.005	-0.003
	(0.017)	(0.027)	(0.028)	(0.031)	(0.036)	(0.035)	(0.033)	(0.015)	(0.015)	(0.016)
Education	-0.005	0.006	0.014	0.009	-0.009	-0.015	-0.012	-0.000	0.002	0.003
	(0.006)	(0.008)	(0.009)	(0.010)	(0.012)	(0.012)	(0.011)	(0.005)	(0.005)	(0.006)
SES	-0.007	0.003	0.007	0.013	-0.030***	-0.026**	-0.013	-0.002	-0.001	0.000
	(0.005)	(0.007)	(0.008)	(0.009)	(0.011)	(0.011)	(0.010)	(0.005)	(0.005)	(0.005)
Risky choices	0.002	-0.006	-0.004	-0.003	0.001	0.000	0.003	0.000	0.001	0.000
	(0.003)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
Patient choices	0.002	-0.003	-0.003	-0.004	0.005*	0.005**	0.005*	-0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
Charity	-0.001	0.003	0.002	0.000	0.031**	0.017	0.007	-0.004	-0.005	-0.007
	(0.006)	(0.009)	(0.010)	(0.010)	(0.012)	(0.012)	(0.011)	(0.005)	(0.005)	(0.006)
Constant	0.320***	0.296***	0.300***	0.353***	0.458***	0.599***	0.637***	0.028	-0.003	0.010
	(0.045)	(0.069)	(0.074)	(0.080)	(0.097)	(0.093)	(0.088)	(0.042)	(0.046)	(0.049)
Observations	785	785	785	785	785	785	785	785	785	785
R-squared	0.036	0.036	0.035	0.033	0.035	0.035	0.026	0.008	0.012	0.011

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix B: instructions

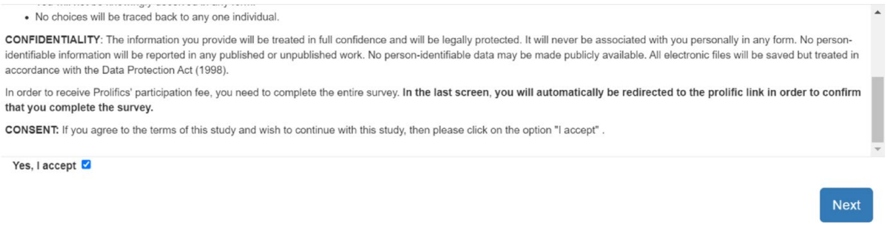
Consent


Please read the following information carefully: You are invited to take part in a short survey study. The following list of items summarises all important things you should know before proceeding with this study.

- You must be over 18 to participate.
- Your participation is voluntary, and you may withdraw from the study at any time.
- Participation will not incur any financial expense by you.
- If you agree to participate in the study, you are expected to fulfill the obligations related to the study. That is, answer any questions posed to you for the duration of the study.
- There are no known physical risks involved in this procedure and the tasks do not require any special physical or psychological attitudes or any specific knowledge of any kind.
- You will not be knowingly deceived in any form.
- No choices will be traced back to any one individual.

Yes, I accept

Next


- No choices will be traced back to any one individual.

CONFIDENTIALITY: The information you provide will be treated in full confidence and will be legally protected. It will never be associated with you personally in any form. No person-identifiable information will be reported in any published or unpublished work. No person-identifiable data may be made publicly available. All electronic files will be saved but treated in accordance with the Data Protection Act (1998).

In order to receive Prolifics' participation fee, you need to complete the entire survey. **In the last screen, you will automatically be redirected to the prolific link in order to confirm that you complete the survey.**

CONSENT: If you agree to the terms of this study and wish to continue with this study, then please click on the option "I accept" .

Yes, I accept

Next

Subjects were randomly assigned to one of the three treatments: Real (R), Hypothetical (H) and BRIS (B). The instructions of the time discounting, risk (Holt and Laury), and dictator game tasks were identical across treatments, except for the last sentence, where we introduced the specific payment condition. All participants remained in the assigned treatment across all the three tasks and completed them in the same order (time, risk, DG).

Real (R)

Time discounting task

Part I: present versus future

In this section, you will be asked to make a series of 20 decisions about whether you prefer to receive an amount of money sooner or later. The task consists of 2 blocks. In each block there are 10 identical decisions, but they differ in the time that you would receive the amount of money involved.

There are no right or wrong answers and the decisions are totally independent of each other, since only one of the 20 decisions will be randomly selected to calculate the payment. This will be done on the date established by the option you have chosen in that decision selected at random.

You will receive **REAL** money for this task.

Decision screens for the short and long run tasks:

Block SR

Please choose the option you prefer in each decision. Do you prefer….

For each row, choose between receiving an amount **today** or a larger amount **in one month**.

- Receive **£3 today** or **£3.0 in one month**?
- Receive **£3 today** or **£3.2 in one month**?
- Receive **£3 today** or **£3.4 in one month**?
- Receive **£3 today** or **£3.6 in one month**?
- Receive **£3 today** or **£3.8 in one month**?
- Receive **£3 today** or **£4.0 in one month**?
- Receive **£3 today** or **£4.2 in one month**?
- Receive **£3 today** or **£4.4 in one month**?
- Receive **£3 today** or **£4.6 in one month**?
- Receive **£3 today** or **£4.8 in one month**?

For each decision, participants selected **one option only** (Today vs. In one month).

Block LR

Please choose the option you prefer in each decision. Do you prefer….

For each row, choose between receiving an amount **in one month** or a larger amount **in seven months**.

- Receive **£3 in one month** or **£3 in seven months**?
- Receive **£3 in one month** or **£3.2 in seven months**?
- Receive **£3 in one month** or **£3.4 in seven months**?
- Receive **£3 in one month** or **£3.6 in seven months**?
- Receive **£3 in one month** or **£3.8 in seven months**?
- Receive **£3 in one month** or **£4.0 in seven months**?
- Receive **£3 in one month** or **£4.2 in seven months**?
- Receive **£3 in one month** or **£4.4 in seven months**?
- Receive **£3 in one month** or **£4.6 in seven months**?
- Receive **£3 in one month** or **£4.8 in seven months**?

For each decision, participants selected **one option only** (In one month vs. In seven months).

Holt and Laury task

Part II: decisions under uncertainty

For each of the following decisions, you have to choose the lottery you prefer (A or B). In each lottery you can win one amount with a certain probability *p* or a different amount with probability *(1–p)*.

In total, you are going to make 11 decisions. Only one of the 11 decisions will be randomly selected to calculate the payment and the prize that you can win will depend on the lottery you choose (A or B) and the result of that lottery.

Remember, you will receive **REAL** money for this task.

Decision screen:

Please select the lottery (A or B) you prefer in each decision.

Dictator game

Part III: allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £1 to split between you and the other person. One of the two (player A) is going to divide the £1 between both of you. The other (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be **either player A or player B; 50% chance).**

Remember, you will receive **REAL** money for this task.

Please mark the percentage of the £1 that you would want to send to the other person. *Choose one of the following answers*:

- *£0.*
- *£0.1.*
- *£0.2.*
- *£0.3.*
- *£0.4.*
- *£0.5.*
- *£0.6.*
- *£0.7.*
- *£0.8.*
- *£0.9.*
- *£1.*

BRIS (B)

Time discounting task

Part I: present versus future

In this section, you will be asked to make a series of 20 decisions about whether you prefer to receive an amount of money sooner or later. The task consists of 2 blocks. In each block there are 10 identical decisions, but they differ in the time that you would receive the amount of money involved.

There are no right or wrong answers and the decisions are totally independent of each other, since only one of the 20 decisions will be randomly selected to calculate the payment. This will be done on the date established by the option you have chosen in that decision selected at random.

ONE OUT OF EVERY TEN PARTICIPANTS, selected at random, will receive real money for this task.

[See the decision screen above – identical across treatments]

Holt and Laury task

Part II: decisions under uncertainty

For each of the following decisions, you have to choose the lottery you prefer (A or B). In each lottery you can win one amount with a certain probability p or a different amount with probability (1–p).

In total, you are going to make 11 decisions. Only one of the 11 decisions will be randomly selected to calculate the payment and the prize you can win will depend on the lottery you choose (A or B) and the result of that lottery.

ONE OUT OF EVERY TEN PARTICIPANTS, selected at random, will receive real money for this task.

[See the decision screen above – identical across treatments]

Dictator game

Part III: allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £1 to split between you and the other person. One of the two (player A) is going to divide the £1 between both of you. The other (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

Decision	Lottery A	Lottery B
1	0%: £0.50 100%: £0.40	0%: £1.00 100%: £0.01
2	10%: £0.50 90%: £0.40	10%: £1.00 90%: £0.01
3	20%: £0.50 80%: £0.40	20%: £1.00 80%: £0.01
4	30%: £0.50 70%: £0.40	30%: £1.00 70%: £0.01
5	40%: £0.50 60%: £0.40	40%: £1.00 60%: £0.01
6	50%: £0.50 50%: £0.40	50%: £1.00 50%: £0.01
7	60%: £0.50 40%: £0.40	60%: £1.00 40%: £0.01
8	70%: £0.50 30%: £0.40	70%: £1.00 30%: £0.01
9	80%: £0.50 20%: £0.40	80%: £1.00 20%: £0.01
10	90%: £0.50 10%: £0.40	90%: £1.00 10%: £0.01
11	100%: £0.50 0%: £0.40	100%: £1.00 0%: £0.01

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be either player A or player B; 50% chance).

ONE OUT OF EVERY TEN PARTICIPANTS, selected at random, will receive real money for this task.

- *Please mark the percentage of the £1 that you would want to send to the other person. Choose one of the following answers*:£0.
- £0.1.
- £0.2.
- £0.3.
- £0.4.
- £0.5.
- £0.6.
- £0.7.
- £0.8.
- £0.9.
- £1.

Hypothetical (H)

Time discounting task

Part I: present versus future

In this section, you will be asked to make a series of 20 decisions about whether you prefer to receive an amount of money sooner or later. The task consists of 2 blocks. In each block, there are 10 identical decisions, but they differ in the time that you would receive the amount of money involved.

There are no right or wrong answers and the decisions are totally independent of each other, since only one of the 20 decisions will be randomly selected to calculate the payment. This will be done on the date established by the option you have chosen in that decision selected at random.

Payments in this section are **HYPOTHETICAL**. Please make the decisions as if they were real.

[See the decision screen above – identical across treatments]

Holt and Laury task

Part II: decisions under uncertainty

For each of the following decisions, you have to choose the lottery you prefer (A or B). In each lottery you can win one amount with a certain probability p or a different amount with probability $(1-p)$.

In total, you are going to make 11 decisions. Only one of the 11 decisions will be randomly selected to calculate the payment and the prize that you can win will depend on the lottery you choose (A or B) and the result of that lottery.

Payments in this section are **HYPOTHETICAL**. Please make the decisions as if they were real.

[See the decision screen above – identical across treatments]

Dictator game

Part III: allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £1 to split between you and the other person. One of the two (player A) is going to divide the £1 between both of you. The other (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be either player A or player B; 50% chance).

Payments in this section are **HYPOTHETICAL**. Please make the decisions as if they were real.

- *Please mark the percentage of the £1 that you would want to send to the other person. Choose one of the following answers*: £0.
- £0.1.
- £0.2.
- £0.3.
- £0.4.
- £0.5.
- £0.6.
- £0.7.
- £0.8.
- £0.9.
- £1.

Appendix C: replication of the same analysis for the £5-DG replication (REP1)

See Tables 8, 9, 10, 11, 12 and Figs. 8, 9.

Table 8 Balance across treatments in the £5-DG replication

	$Mean_H$	$p(R\text{-H})^{ç}$	$p(\text{B-H})^{ç}$
Female	0.456	0.822	1.000
Age	44.213	0.940	0.574
Education+	3.252	0.992	0.724
SES	5.447	0.300	0.024
Charity	2.058	0.300	0.724

Baseline: Hypothetical (H). *Education* is a categorical variable (taking values from 0 to 6 for simplicity, from no formal education to doctorate degree) which refers to the highest education level. SES reflects the *Socio-economic status* using the position in the income ladder (scale from 1 to 10). *Charity* refers to a self-reported categorical variable that reflects different amounts of donations (in ascending order) made in the last year

ç Inference was made regressing each control variable on H and using Westfall and Young adjusted p-values for multiple testing

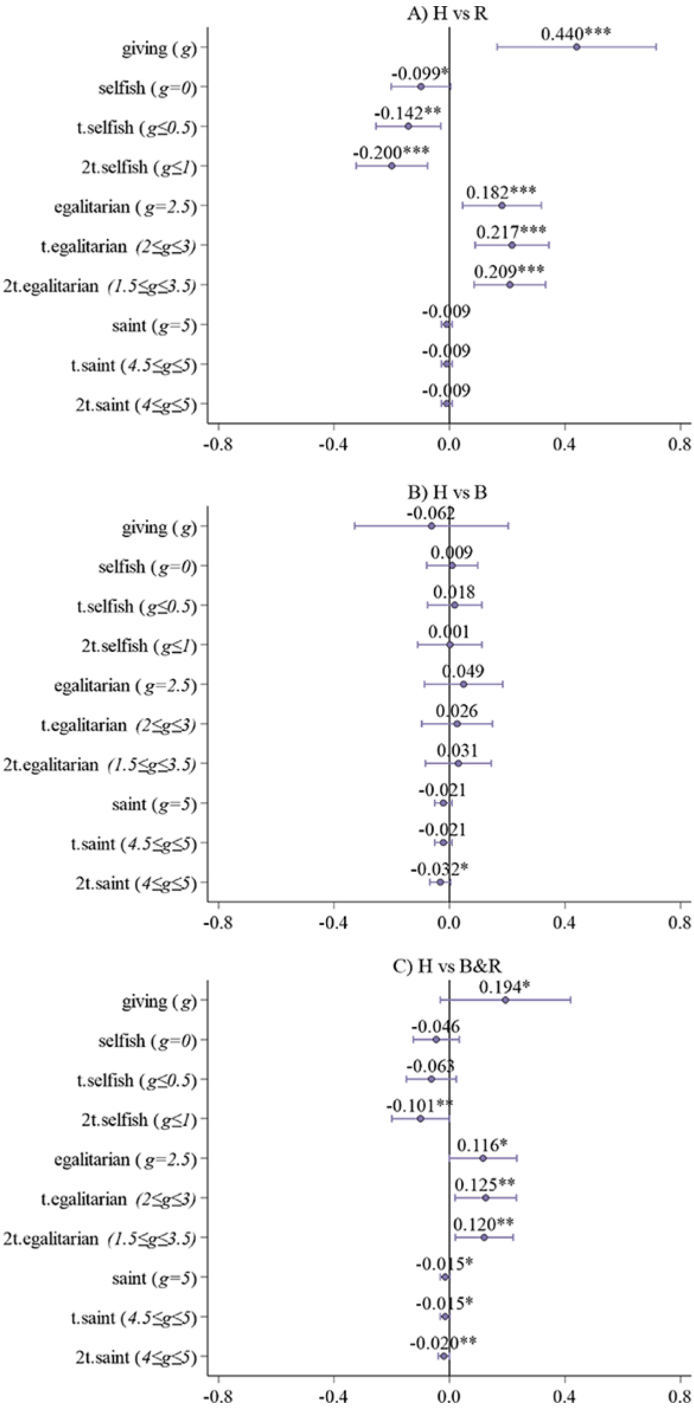


Fig. 8 Regression results - replication. Point estimates denote the coefficient of H (±95% CI). All the regressions control for age, gender, education, SES status, and self-reported donations to charity. Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

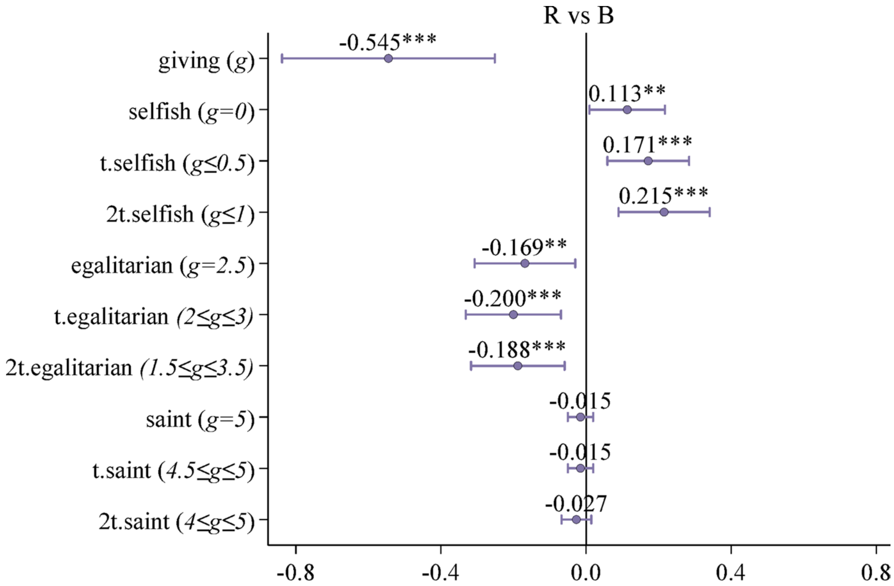


Fig. 9 Regression results - replication. Point estimates denote the coefficient of R (±95% CI). All the regressions control for age, gender, education, SES status, and self-reported donations to charity. Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9 Regression analysis for the H vs. R comparison (replication)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (g)	Selfish ($g=0$)	t.selfish ($g\leq 0.5$)	2t.selfish ($g\leq 1$)	Egalitarian ($g=2.5$)	t.egalitarian ($2\leq g\leq 3$)	2t.egalitarian ($1.5\leq g\leq 3.5$)	Saint ($g=5$)	t.saint ($4.5\leq g\leq 5$)	2t.saint ($4\leq g\leq 5$)
H	0.440***	-0.099*	-0.142**	-0.200***	0.182***	0.217***	0.209***	-0.009	-0.009	-0.009
	(0.139)	(0.052)	(0.057)	(0.063)	(0.069)	(0.065)	(0.063)	(0.009)	(0.009)	(0.009)
Westfall and Young p-value	0.010**	0.192	0.022**	0.010**	0.022**	0.004***	0.060*	0.080*	0.080*	0.080*
Age	0.016***	-0.004*	-0.004*	-0.007***	0.008***	0.009***	0.008***	-0.000	-0.000	-0.000
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Female	-0.086	0.085	0.029	0.004	0.033	0.059	0.004	-0.008	-0.008	-0.008
	(0.138)	(0.053)	(0.057)	(0.062)	(0.069)	(0.064)	(0.062)	(0.008)	(0.008)	(0.008)
Education	0.097*	-0.025	-0.037*	-0.040*	0.017	0.041*	0.039*	0.001	0.001	0.001
	(0.050)	(0.020)	(0.021)	(0.022)	(0.025)	(0.022)	(0.022)	(0.001)	(0.001)	(0.001)
SES	0.061	-0.021	-0.028	-0.023	0.018	0.013	0.020	0.002	0.002	0.002
	(0.060)	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)	(0.003)	(0.003)	(0.003)
Charity	0.041	-0.034**	-0.034**	-0.012	0.011	0.018	0.014	-0.002	-0.002	-0.002
	(0.039)	(0.014)	(0.016)	(0.019)	(0.020)	(0.019)	(0.019)	(0.002)	(0.002)	(0.002)
Constant	0.230	0.595***	0.758***	0.986***	-0.063	-0.087	0.000	0.014	0.014	0.014
	(0.357)	(0.145)	(0.157)	(0.167)	(0.177)	(0.170)	(0.167)	(0.014)	(0.014)	(0.014)
Observations	205	205	205	205	205	205	205	205	205	205
R-squared	0.139	0.114	0.127	0.127	0.089	0.140	0.132	0.016	0.016	0.016

Linear regression estimates. Robust standard errors in parentheses. Columns 8, 9 and 10 show the same results, since the dummy variables are the same in the three cases as there are no additional subjects in the trembling cases

Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 10 Regression analysis for the H vs. B comparison (replication)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (*g*)	Selfish ($g=0$)	t.selfish ($g \le 0.5$)	2t.selfish ($g \le 1$)	egalitarian ($g=2.5$)	t.egalitarian ($2 \le g \le 3$)	2t.egalitarian ($1.5 \le g \le 3.5$)	Saint ($g=5$)	t.saint ($4.5 \le g \le 5$)	2t.saint ($4 \le g \le 5$)
H	-0.062	0.009	0.018	0.001	0.049	0.026	0.031	-0.021	-0.021	-0.032*
	(0.135)	(0.045)	(0.048)	(0.056)	(0.069)	(0.062)	(0.058)	(0.015)	(0.015)	(0.018)
Westfall and Young p-value	0.940	0.968	0.940	0.988	0.884	0.940	0.920	0.386	0.386	0.316
Age	0.010**	-0.002	-0.003*	-0.004**	0.003	0.004*	0.005**	0.000	0.000	-0.000
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Female	-0.098	0.040	0.009	-0.054	0.036	0.085	0.082	-0.018	-0.018	-0.028*
	(0.134)	(0.045)	(0.048)	(0.057)	(0.071)	(0.061)	(0.058)	(0.013)	(0.013)	(0.016)
Education	-0.019	0.000	0.007	0.023	-0.041*	-0.029	-0.029	0.006	0.006	0.006
	(0.051)	(0.017)	(0.017)	(0.020)	(0.025)	(0.021)	(0.020)	(0.007)	(0.007)	(0.007)
SES	0.033	-0.016	-0.019	-0.009	-0.025	-0.006	0.004	0.005	0.005	0.005
	(0.054)	(0.019)	(0.020)	(0.022)	(0.026)	(0.024)	(0.022)	(0.004)	(0.004)	(0.004)
Charity	0.020	-0.030***	-0.033***	-0.009	0.005	0.011	0.017	-0.006	-0.006	-0.008
	(0.036)	(0.011)	(0.011)	(0.016)	(0.020)	(0.017)	(0.016)	(0.005)	(0.005)	(0.005)
Constant	1.550***	0.332***	0.383***	0.401***	0.701***	0.632***	0.586***	-0.011	-0.011	0.013
	(0.313)	(0.115)	(0.119)	(0.134)	(0.170)	(0.145)	(0.138)	(0.017)	(0.017)	(0.029)
Observations	206	206	206	206	206	206	206	206	206	206
R-squared	0.034	0.060	0.063	0.035	0.033	0.034	0.044	0.039	0.039	0.045

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11 Regression analysis for the H vs. R&B comparison (replication)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (*g*)	Selfish ($g=0$)	t.selfish ($g\leq 0.5$)	2t.selfish ($g\leq 1$)	Egalitarian ($g=2.5$)	t.egalitarian ($2\leq g\leq 3$)	2t.egalitarian ($1.5\leq g\leq 3.5$)	Saint ($g=5$)	t.saint ($4.5\leq g\leq 5$)	2t.saint ($4\leq g\leq 5$)
H	0.194*	-0.046	-0.063	-0.101**	0.116*	0.125**	0.120**	-0.015*	-0.015*	-0.020**
	(0.115)	(0.040)	(0.044)	(0.050)	(0.059)	(0.054)	(0.051)	(0.009)	(0.009)	(0.010)
Westfall and Young p-value	0.232	0.288	0.286	0.196	0.196	0.068*	0.068*	0.288	0.288	0.286
Age	0.013***	-0.003**	-0.004**	-0.006***	0.005**	0.006***	0.006***	0.000	0.000	-0.000
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Female	-0.114	0.060	0.018	-0.013	0.012	0.064	0.035	-0.017*	-0.017*	-0.022**
	(0.116)	(0.042)	(0.045)	(0.051)	(0.058)	(0.053)	(0.051)	(0.010)	(0.010)	(0.011)
Education	0.032	-0.012	-0.017	-0.005	-0.011	-0.002	0.001	0.005	0.005	0.004
	(0.045)	(0.016)	(0.017)	(0.019)	(0.021)	(0.019)	(0.019)	(0.005)	(0.005)	(0.005)
SES	0.030	-0.014	-0.015	-0.005	-0.008	-0.009	0.000	0.004	0.004	0.005
	(0.048)	(0.019)	(0.019)	(0.020)	(0.021)	(0.020)	(0.020)	(0.003)	(0.003)	(0.003)
Charity	0.036	-0.033***	-0.036***	-0.016	0.007	0.020	0.022	-0.006	-0.006	-0.006*
	(0.033)	(0.011)	(0.012)	(0.015)	(0.017)	(0.016)	(0.015)	(0.004)	(0.004)	(0.004)
Constant	0.996***	0.446***	0.546***	0.618***	0.355**	0.352***	0.374***	-0.006	-0.006	0.008
	(0.289)	(0.108)	(0.115)	(0.128)	(0.145)	(0.133)	(0.129)	(0.018)	(0.018)	(0.023)
Observations	308	308	308	308	308	308	308	308	308	308
R-squared	0.053	0.067	0.068	0.047	0.033	0.058	0.055	0.027	0.027	0.029

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 12 Regression analysis for the R vs. B comparison (replication)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Giving (*g*)	Selfish ($g=0$)	t.selfish ($g \leq 0.5$)	2t.selfish ($g \leq 1$)	Egalitarian ($g=2.5$)	t.egalitarian ($2 \leq g \leq 3$)	2t.egalitarian ($1.5 \leq g \leq 3.5$)	Saint ($g=5$)	t.saint ($4.5 \leq g \leq 5$)	2t.saint ($4 \leq g \leq 5$)
R	-0.545***	0.113**	0.171***	0.215***	-0.169**	-0.200***	-0.188***	-0.015	-0.015	-0.027
	(0.149)	(0.053)	(0.057)	(0.064)	(0.070)	(0.067)	(0.065)	(0.018)	(0.018)	(0.021)
Westfall and Young p-value	0.002***	0.102	0.026**	0.012**	0.066*	0.026**	0.032**	0.354	0.358	0.314
Age	0.010*	-0.003	-0.003	-0.004*	0.005*	0.005**	0.004*	0.000	0.000	-0.000
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)
Female	-0.178	0.061	0.019	0.014	-0.030	0.047	0.021	-0.026*	-0.026*	-0.035**
	(0.147)	(0.055)	(0.058)	(0.064)	(0.071)	(0.066)	(0.065)	(0.015)	(0.015)	(0.017)
Education	0.044	-0.017	-0.030	-0.008	0.002	-0.007	0.001	0.007	0.007	0.007
	(0.058)	(0.022)	(0.023)	(0.024)	(0.026)	(0.025)	(0.025)	(0.007)	(0.007)	(0.007)
SES	0.029	-0.011	-0.007	0.002	-0.006	-0.018	-0.009	0.006	0.006	0.007
	(0.058)	(0.024)	(0.024)	(0.024)	(0.025)	(0.024)	(0.024)	(0.004)	(0.004)	(0.004)
Charity	0.010	-0.025	-0.027	-0.010	-0.005	0.016	0.020	-0.009	-0.009	-0.011*
	(0.045)	(0.015)	(0.017)	(0.020)	(0.022)	(0.021)	(0.020)	(0.006)	(0.006)	(0.006)
Constant	1.445***	0.357***	0.422***	0.387**	0.464***	0.571***	0.595***	-0.007	-0.007	0.018
	(0.350)	(0.134)	(0.140)	(0.152)	(0.176)	(0.156)	(0.154)	(0.021)	(0.021)	(0.032)
Observations	205	205	205	205	205	205	205	205	205	205
R-squared	0.095	0.064	0.088	0.077	0.048	0.088	0.070	0.036	0.036	0.043

Linear regression estimates. Robust standard errors in parentheses

Asterisk denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix D: instructions in the £5 DG experiment (REP1) and the survey experiment (REP2)

£5 DG experiment (REP1)

Subjects were randomly assigned to one of three treatments: Hypothetical (H), BRIS (B), or Real (R). The consent screen was the same as in the main study. The instructions were as follows:

Hypothetical (H)

Part I: Allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £5 to split between you and the other person. One of the two (player A) is going to divide the £5 between both of you. The other player (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be either player A or player B; 50% chance).

Payments in this section are **HYPOTHETICAL**. Please make the decisions as if they were real.

BRIS (B)

Part I: Allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £5 to split between you and the other person. One of the two (player A) is going to divide the £5 between both of you. The other player (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be either player A or player B; 50% chance).

ONE OUT OF EVERY TEN PARTIPANTS, selected at random, will receive real money for this task.

Real (R)

Part I: Allocations

In this task, you have to divide an amount of money between you and another person. Initially, we give you £5 to split between you and the other person. One of the two (player A) is going to divide the £5 between both of you. The other player (player B) will receive the amount that player A sends him/her but he or she does not have to make any decision (player B is passive).

The payment will correspond to the decision you have made about how to divide the money. You can also be the one who receives the money that someone else sends you instead of who makes the division (that is, you can be either player A or player B: 50% chance).

Remember, you will receive **REAL** money for this task.

Then, independently of the treatment, all subjects saw the same decision screen.

Please indicate how much of the £5 you would like to send to the other person.

- ○ £0
- ○ £0.5
- ○ £1.0
- ○ £1.5
- ○ £2.0
- ○ £2.5
- ○ £3.0
- ○ £3.5
- ○ £4.0
- ○ £4.5
- ○ £5.0

Survey experiment (REP2)

Subjects were randomly assigned to one of the three treatments: Hypothetical (H), Imagine (I) and Not Real (NotR). The instructions were as follows:

Hypothetical (H)

Part I: Read the following instructions

There are £5 (in 10 coins of 50 pence) provisionally allocated to you and another person. Your task is to divide this amount of money between the two of you, assigning one portion to yourself and the other portion to the other person, such that the total adds up to £5. Any division — even one where one person receives nothing — is acceptable.

There are no right or wrong answers, and the money is **hypothetical.** Please make the decisions as if they were real.

→

Imagine (I)

Part I: Read the following instructions

Imagine a situation where £5 (in 10 coins of 50 pence) are provisionally allocated to you and another person. Your task would be to divide this money between the two of you, assigning one portion to yourself and the other portion to the other person, so that the total adds up to £5. Any division — even one in which one person receives nothing — would be acceptable.

There are no right or wrong answers.

→

Not real (NotR)

Part I: Read the following instructions.

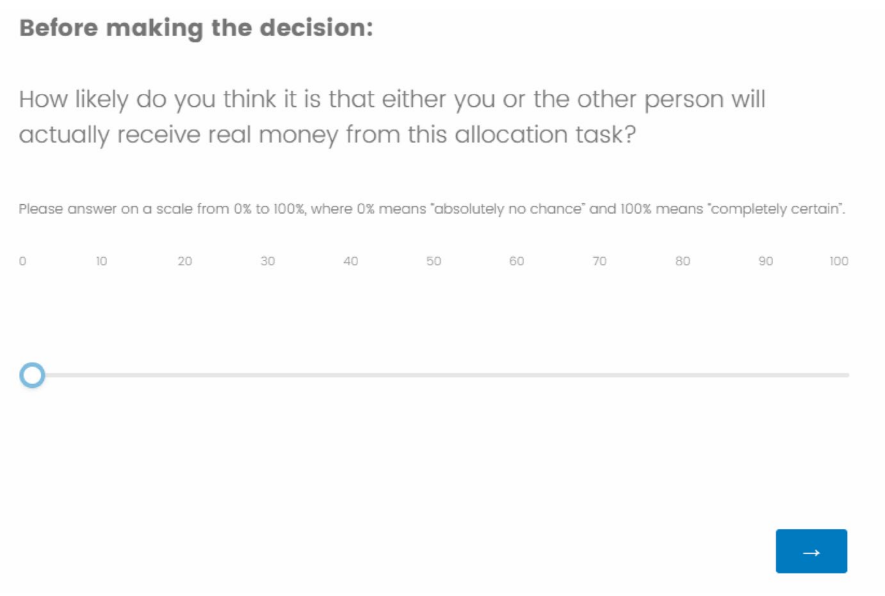
There are £5 (in 10 coins of 50 pence) provisionally allocated to you and another person. Your task is to divide this amount of money between the two of you, assigning one portion to yourself and the other portion to the other person, such that the total adds up to £5. Any division — even one where one person receives nothing — is acceptable.

There are no right or wrong answers. Unfortunately, the money you earn is **not real**, but please make your choices as you would if the money were real.

→

After this screen, follow the different questions made to subjects.

Likely



Before making the decision:

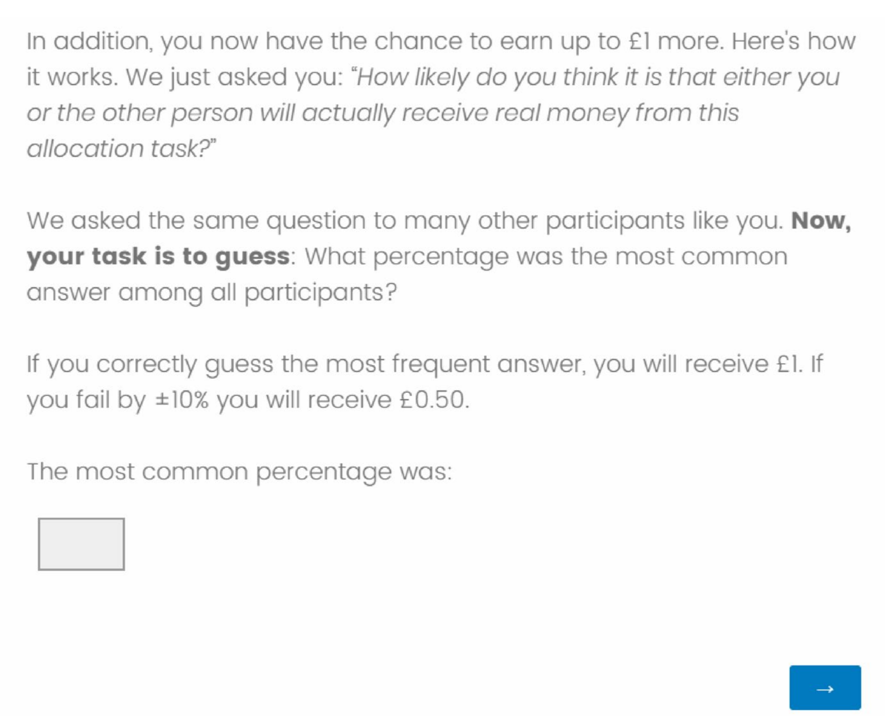
How likely do you think it is that either you or the other person will actually receive real money from this allocation task?

Please answer on a scale from 0% to 100%, where 0% means "absolutely no chance" and 100% means "completely certain".

0 10 20 30 40 50 60 70 80 90 100

→

Guess



In addition, you now have the chance to earn up to £1 more. Here's how it works. We just asked you: "*How likely do you think it is that either you or the other person will actually receive real money from this allocation task?*"

We asked the same question to many other participants like you. **Now, your task is to guess**: What percentage was the most common answer among all participants?

If you correctly guess the most frequent answer, you will receive £1. If you fail by ±10% you will receive £0.50.

The most common percentage was:

→

Decision: giving

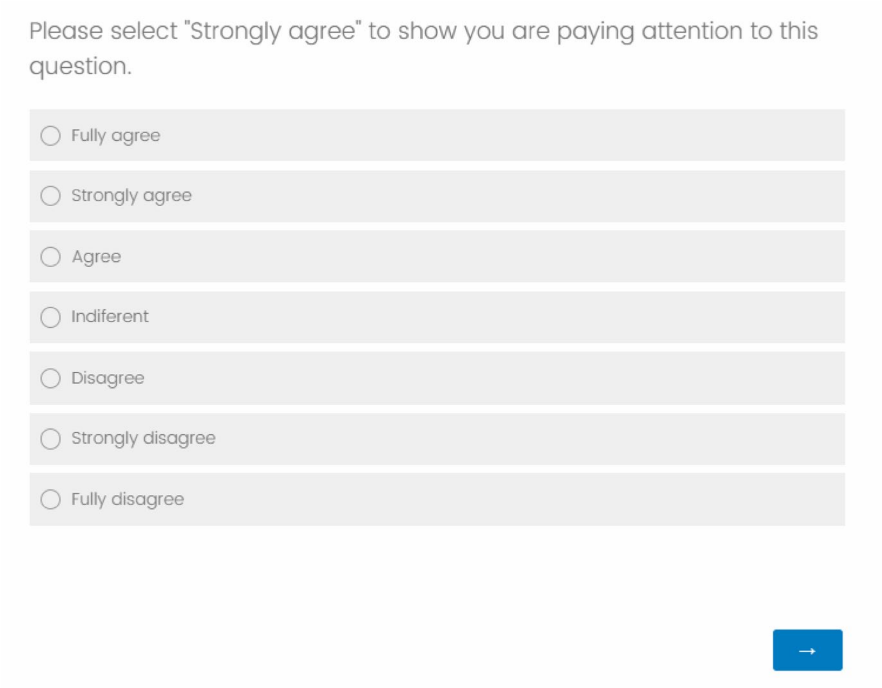
Now make your decision.

Please indicate how much of the £5 you would like to send to the other person:

- £0
- £0.5
- £1.0
- £1.5
- £2.0
- £2.5
- £3.0
- £3.5
- £4.0
- £4.5
- £5.0

→

Attention question


Please select "Strongly agree" to show you are paying attention to this question.

- Fully agree
- Strongly agree
- Agree
- Indiferent
- Disagree
- Strongly disagree
- Fully disagree

→

Appendix E: additional analysis and figures

Average giving by treatment condition

See Fig. 10.

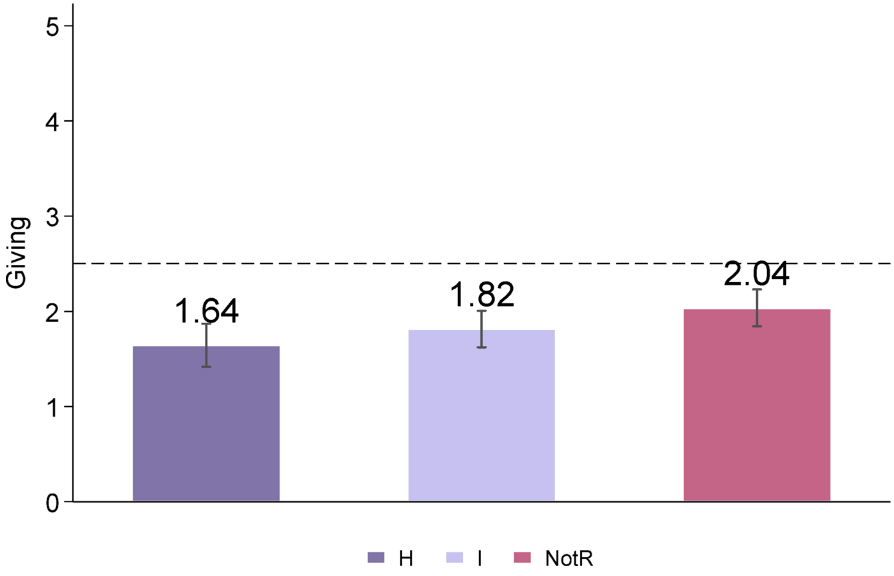


Fig. 10 Average (±95% CI) giving by treatment conditions in the survey experiment

Second-order beliefs

See Fig. 11.

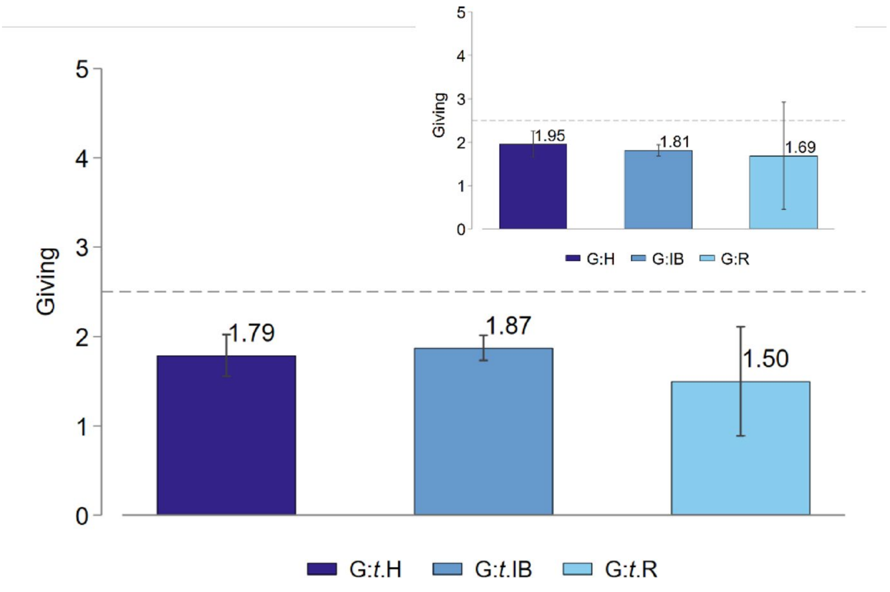


Fig. 11 Average (±95% CI) giving by types of guessed modal real-payment likelihood (second-order beliefs)

Additional analysis – comparisons across studies

See Table 13 and Figs. 12, 13, 14.

Table 13 Comparison of participant characteristics in the £5 DG study (REP1) and the survey experiment (REP2)

Variable	(1) REP1	(2) REP2	(3) Difference
Age	44.815	46.846	2.031*
	(12.982)	(13.466)	(1.074)
Female	0.438	0.446	0.008
	(0.497)	(0.498)	(0.040)
Education	3.185	3.191	0.006
	(1.456)	(1.468)	(0.119)
SES	5.140	5.238	0.099
	(1.505)	(1.412)	(0.119)
Charitable giving	1.854	2.000	0.146
	(1.686)	(1.729)	(0.139)
Observations	308	299	607

The table reports means and standard deviations (in parentheses) for observable participant characteristics in each study. Column (3) reports differences in means (REP2 − REP1)

Asterisks denote significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

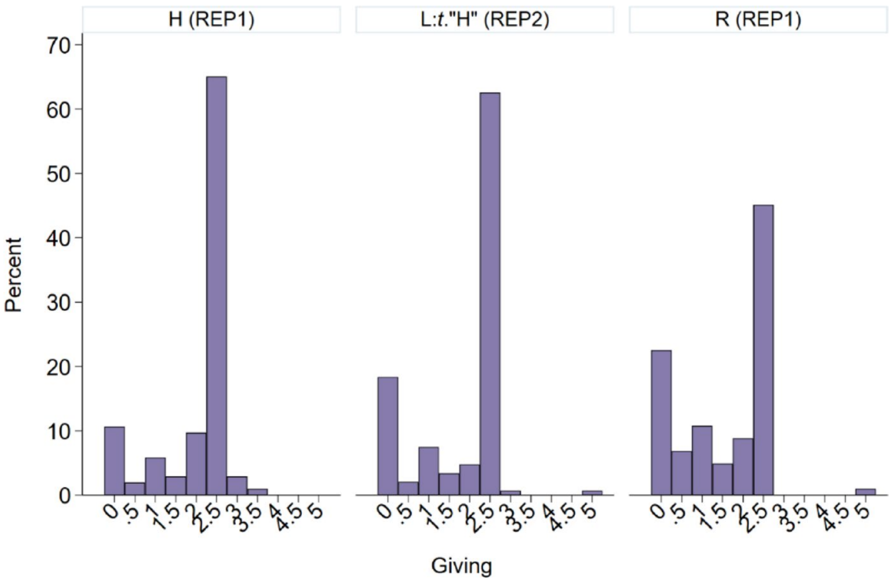


Fig. 12 Distribution of giving in the Hypothetical group (H) for the £5 DG study (REP1), in the belief-based Hypothetical group (L:*t*.“H”) for the survey experiment (REP2) and in the Real-incentivized group (R) for REP1

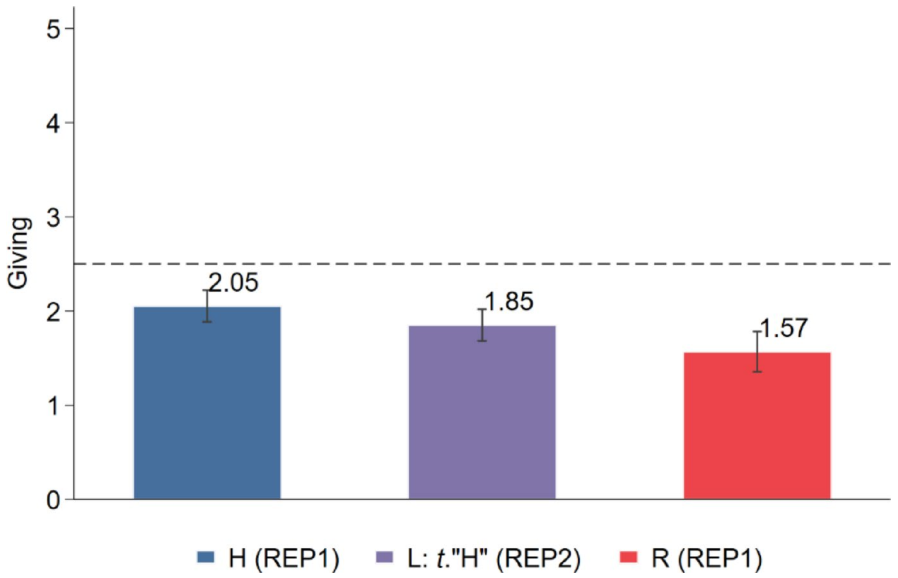


Fig. 13 Average (±95% CI) giving in the belief-based Hypothetical group (L:*t*."H") in the survey experiment (REP2), in the Real-payment (R) and Hypothetical (H) conditions in the £5 DG study (REP1)

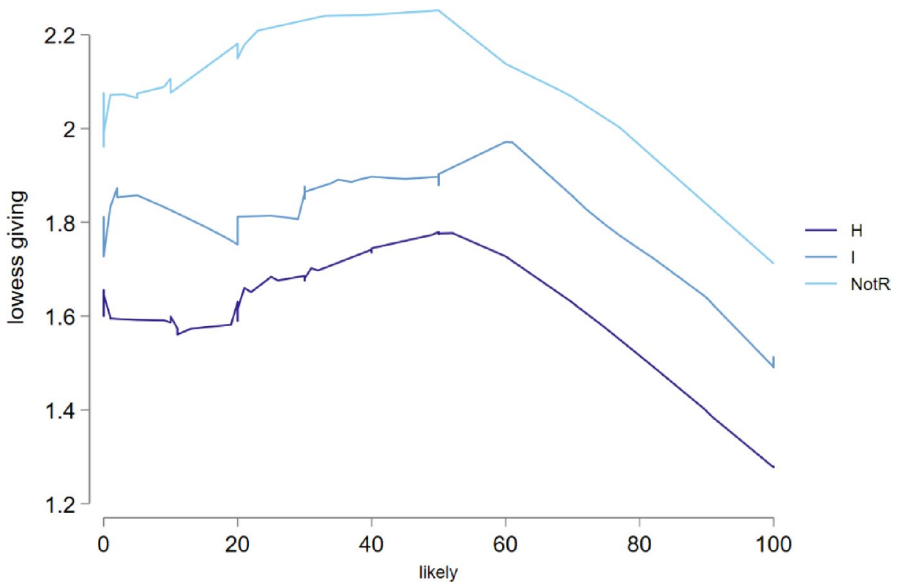


Fig. 14 LOWESS smoothed relationship between giving and perceived probability of real payment by treatment conditions in the survey experiment

Acknowledgements We are grateful to the Editors – Jerome Renault and Bernhard von Stengel – and to the two insightful and creative referees for their excellent comments. We also appreciate the comments of participants at seminars in Alicante, Barcelona UB, Derby, GATE-Lyon, Granada, KTI Budapest and Loyola. Financial support from MINECO-FEDER (PID2024-156629NB-I00), Consejería de Universidad, Investigación e Innovación, Andalusia Government, and ERDF (EMERGIA EMC21_00331; FEDER C-SEJ-371-UGR23). This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement number 101095175 and the UK Government.

Data Availability The data and code that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.19228647

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Ahles A, Palma MA, Drichoutis AC (2024) Testing the effectiveness of lottery incentives in online experiments. Am J Agric Econ 106(4):1435–1453

Amir O, Rand DG, Gal YK (2012) Economic games on the internet: the effect of $1 stakes. PLoS ONE 7(2):e31461

Andreoni J, Bernheim BD (2009) Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects. Econometrica 77(5):1607–1636

Arechar A, Gächter S, Molleman L (2018) Conducting interactive experiments online. Exp Econ 21(1):99–131

Bechler C, Green L, Myerson J (2015) Proportion offered in the dictator and ultimatum games decreases with amount and social distance. Behav Process 115:149–155

Ben-Ner A, Kramer A, Levy O (2008) Economic and hypothetical dictator game experiments: Incentive effects at the individual level. J Socio Econ 37(5):1775–1784

Ben-Ner A, McCall BP, Stephane M, Wang H (2009) Identity and in-group/out-group differentiation in work and giving behaviors: experimental evidence. J Econ Behav Organ 72(1):153–170

Brañas-Garza P, Kovárík J, Neyse L (2013) Second-to-fourth digit ratio has a non-monotonic impact on altruism. PLoS ONE 8(4):e60419

Brañas-Garza P, Capraro V, Rascón-Ramírez (2018) Gender differences in altruism on mechanical Turk: expectations and actual behaviour. Econ Lett 170:19–23

Brañas-Garza P, Estepa-Mohedano L, Jorrat D, Orozco V, Rascón-Ramírez E (2021a) To pay or not to pay: measuring risk preferences in lab and field. Judgm Decis Mak 16(5):1290–1313

Brañas-Garza P, Jorrat D, Kovarik J, López MC (2021b) Hyper-altruistic behavior vanishes with high stakes. PLoS ONE 16(8):e0255668

Brañas-Garza P, Jorrat D, Espín AM, Sánchez A (2023) Paid and hypothetical time preferences are the same: lab, field and online evidence. Exp Econ 26:412–434

Bühren C, Kundt T (2015) Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. Judgm Decis Mak 10(2):185–190

Camerer CF, Hogarth RM (1999) The effects of financial incentives in experiments: a review and capital-labor-production framework. J Risk Uncertain 19(1):7–42
Carpenter J, Verhoogen E, Burks S (2005) The effect of stakes in distribution experiments. Econ Lett 86(3):393–398
Charness G, Gneezy U, Halladay B (2016) Experimental methods: pay one or pay all. J Econ Behav Organ 131:141–150
Clot S, Grolleau G, Ibanez L (2018) Shall we pay all? An experimental test of random incentivized systems. J Behav Exp Econ 73:93–98
Coller M, Williams MB (1999) Eliciting individual discount rates. Exp Econ 2:107–127
Dana J, Weber RA, Kuang JX (2007) Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ Theor 33(1):67–80
Engel C (2011) Dictator games: a meta study. Exp Econ 14(4):583–610
Exadaktylos F, Espín AM, Brañas-Garza P (2013) Experimental subjects are not different. Sci Rep 3(1):1–6
Forsythe R, Horowitz JL, Savin NE, Sefton M (1994) Fairness in simple bargaining experiments. Games Econ Behav 6:347–369
Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. Q J Econ 115(3):791–810
Grossman PJ, Zhan W, Eckel CC (2020) Does how we measure altruism matter? Playing both roles in dictator games. Mimeo
Guillén P, Veszteg RF (2012) On lab rats. J Socio Econ 41(5):714–720
Holt CA, Laury SK (2002) Risk aversion and incentive effects. Am Econ Rev 92(5):1644–1655
Horton JJ, Rand DG, Zeckhauser RJ (2011) The online laboratory: conducting experiments in a real labor market. Exp Econ 14(3):399–425
Johnson MW, Bickel WK (2002) Within-subject comparison of real and hypothetical money rewards in delay discounting. J Anal Behav 77(2):129–146
Jones D, Molitor D, Reif J (2019) What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. Q J Econ 134(4):1747–1791
Larney A, Rotella A, Barclay P (2019) Stake size effects in ultimatum game and dictator game offers: a meta-analysis. Organ Behav Hum Decis Process 151:61–72
Locey ML, Jones BA, Rachlin H (2011) Real and hypothetical rewards in self-control and social discounting. Judgm Decis Mak 6(6):552–564
Palan S, Schitter C (2018) Prolific. ac—a subject pool for online experiments. J Behav Exp Finance 17:22–27
Paolacci G, Chandler J (2014) Inside the Turk: understanding mechanical Turk as a participant pool. Curr Dir Psychol Sci 23(3):184–188
Prissé B, Jorrat D (2022) Lab vs online experiments: no differences. J Behav Exp Econ 100:101910
Rand DG (2012) The promise of mechanical Turk: how online labor markets can help theorists run behavioral experiments. J Theor Biol 299:172–179
Rosenboim M, Shavit T (2012) Whose money is it anyway? Using prepaid incentives in experimental economics to create a natural environment. Exp Econ 15(1):145–157
Sefton M (1992) Incentives in simple bargaining games. J Econ Psychol 13(2):263–276
Thielmann I, Heck DW, Hilbig BE (2016) Anonymity and incentives: an investigation of techniques to reduce socially desirable responding in the Trust Game. Judgm Decis Mak 11(5):527–536
Umer H (2023) Effectiveness of random payment in experiments: a meta-analysis of dictator games. J Econ Psychol 96:102608
Westfall PH, Young SS (1993) Resampling-based multiple testing: examples and methods for p-value adjustment. John Wiley & Sons Inc, New York
Zizzo DJ (2010) Experimenter demand effects in economic experiments. Exp Econ 13(1):75–98

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Pablo Brañas-Garza[1] · **Antonio M. Espín**[2] · **Diego Jorrat**[3]

✉ Pablo Brañas-Garza
branasgarza@gmail.com

1 LoyolaBehLab, Department of Finance, Universidad Loyola Andalucía, C. Escritor Castilla Aguayo, 4, 14004 Córdoba, Spain

2 Department of Applied Economics, Universidad de Granada, Granada, Spain

3 Department of Applied Economics I, Universidad de Sevilla, Seville, Spain