Online Appendix

for

Paid and hypothetical time preferences are the same: Lab, field and online evidence

A. Sample and balance across experiments

Table A.1 shows the total number of observations and the balance of the randomization across treatments in each experiment. In the lab experiment (Study I), we can see that all the individual characteristics gathered (age, gender, and CRT score) are balanced between the treatments, except for one marginally significant difference in age: in the BRIS treatment (B), individuals were on average 1.2 years younger than in the R treatment *($p = 0.064$*).

In the field experiment (Study II), we observe significant differences in age, as participants in the H treatment were 2.4 years older than those in the R treatment ($p = 0.024$). It should be noted that risk preferences were measured for only half of the sample ($n = 360$). Still, the treatments were balanced in risk preferences as well.

In the first experiment in Prolific (Study III), we observe a marginally significant difference in terms of age, as participants in the H treatment were 2.1 years younger than those in the R treatment ($p = 0.078$). Also, the average number of risky choices in the B treatment was 0.53 (12%) higher than in the R treatment ($p = 0.028$).

In the second experiment in Prolific (Study IV), we observe a significant difference in the average number of risky choices between the same two treatments ($p = 0.029$): the participants in H were slightly more educated and riskier than the participants in R.

However, none of the six significant or marginally significant differences between treatments would survive a Bonferroni-like correction for multiple testing. Thus, the results in Table A.1 suggest that the randomization worked properly. Having a balance between treatments is essential to isolate the impact of different payment mechanisms on TD choices. Since the results imply that our sub-samples are nearly identical, we can test the *causal effects of incentives* on decisions. Moreover, the regression analysis will allow us to control for potential confounds.

Table A.1: **Balance across treatments in studies I-IV**

	obs.	$mean_R$	$H - R$	*p*	$B - R$	*p*
Study I: Lab						
Age	119	21.846	-0.471	0.463	-1.196	0.064*
Female	119	0.385	0.065	0.563	0.115	0.308
CRT	118	1.282	-0.332	0.153	-0.308	0.188
Study II: Field						
Age	721	39.238	2.421	0.024**	-0.442	0.678
Female	721	0.527	0.049	0.283	0.030	0.512
Education	721	7.715	0.211	0.714	0.064	0.915
Sufficient	721	0.787	0.006	0.870	0.027	0.464
Risky choices	360	1.917	0.004	0.983	-0.226	0.252
Study III: Prolific 1						
Age	606	33.048	-2.076	0.078*	-1.411	0.253
Female	606	0.599	0.043	0.378	0.048	0.328
Education	605	16.011	-0.136	0.395	-0.080	0.624
Soc. economic status	606	5.069	0.037	0.813	0.225	0.156
Risky choices	606	5.203	0.141	0.555	0.537	0.028**
Study IV: Prolific 2						
Age	592	31.865	1.124	0.364	1.184	0.307
Female	592	0.637	-0.045	0.358	0.018	0.711
Education	592	15.812	0.213	0.264	0.114	0.547
Soc. economic status	596	5.145	0.049	0.748	-0.071	0.654
Risky choices	596	5.192	0.517	0.029**	-0.266	0.255

Note: Inference was made using OLS regression with robust standard errors. *R* refers to Real, *H* to Hypothetical, and *B* to BRIS. *CRT* refers to the number of reflective answers in the Cognitive Reflection Test (Frederick 2005) and takes values from 0 to 3. *Sufficient* refers to individuals self-reporting that they have enough money to feed the family (dummy). *Education* is a discrete variable referring to the highest education level reported by the participant (in years of schooling, from 0 to 19). *Risky choices* is the number of risky options chosen in a risk preferences task (in the field, it takes values from 0 to 5 because we used a trimmed version of the Holt-Laury (2002) task; while in Prolific it takes values from 0 to 10 because we used the original task). *Socio-economic status* is an income proxy (position in the income ladder; Likert scale from 1 to 10). $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

B. Additional analysis

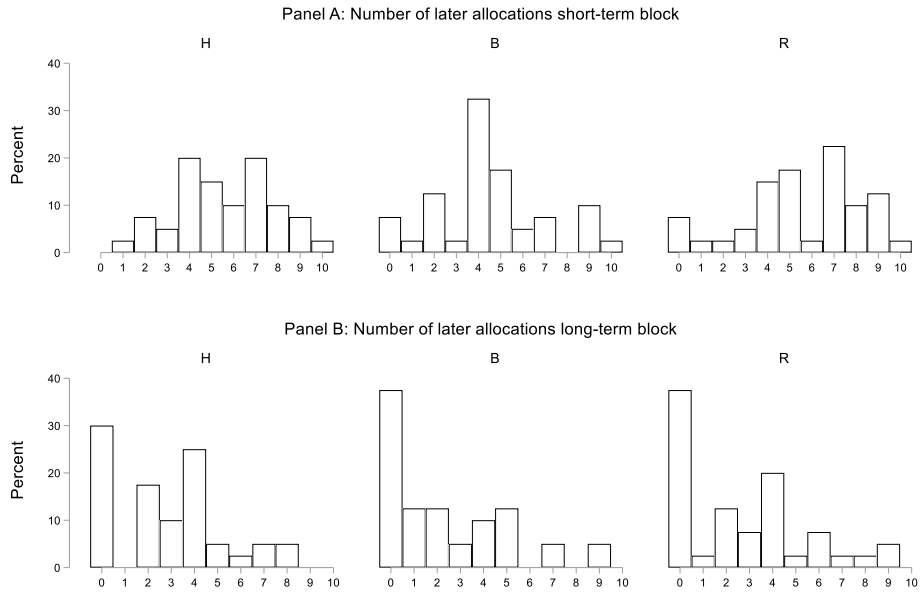
The results in the main text are rather clear: as compared to real rewards, hypothetical and BRIS payment methods do not result in different time preferences in terms of mean estimates, variances, or observed distributions. The effects are very small and non-significant and there is little variability between the studies (especially in the case of H vs. R). In this section, we summarize the results of a series of additional analyses. For the sake of conciseness, the results are presented in detail in sections B to F of the supplementary information.

Section B shows that all the aggregate results are not dramatically different across studies. This applies to the mean estimates, as we analyze through the meta-analytic heterogeneity statistics, as well as to the variances and observed distributions. The marginally smaller overall variance in treatment R as compared to treatment H (which does not reach significance when R is compared to B) can be attributed to Study II, in which the ratio test for both delta and number of later allocations (long) is significant at the 1% level when R is compared to both the H and B treatments. This analysis also shows that differences in terms of inconsistency of choices, which is a potentially interesting alternative measure of noise, are non-systematic and non-significant (expect for a marginally higher inconsistency in B compared to R in Study I; but note that inconsistencies were in general very infrequent).

B.1 Results from the lab experiment (Study I)

Figure B.1 shows the distribution of choices in the three treatments in the lab experiment. In Table B.1, we show the impact of hypothetical (H) and BRIS (B), as compared to real (R) incentives on individuals' patience using OLS regressions with different specifications. Columns 1 to 4 display the results when the dependent variable is β or δ from the beta-delta model. In columns 5 to 8 the dependent variable is the number of later allocations in each of the two blocks (short-term or long-term). The regressions in columns 3, 4, 7, and 8 control for age, gender, and CRT score. We use CRT as a proxy of participants' cognitive skills and takes values from 0 to 3. None of these variables are significant ($p > 0.20$).

Figure B.1: **Histogram of the number of later allocations in the short- and long-term blocks (Study I: Lab).**



B.1.1 Are hypothetical and real choices different (Q1)?

Columns 1 and 2 in Table B.1 show that the H dummy has no significant impact on beta or delta ($p > 0.89$), suggesting that hypothetical decisions do not differ from real incentivized decisions (R). After adding the control variables (columns 3 and 4), the coefficients remain non-significant ($p > 0.87$).

Regarding the number of later allocations (columns 5-8), H does not yield significant estimates on either the short-term or the long-term block ($p > 0.86$ without controls, $p > 0.78$ with controls). We also performed a *Kolmogorov-Smirnov test* for equality of distributions between H and R (see the distribution in Figure B.1). The results confirm that there are no statistical differences between both treatments in the distribution of number of later allocations in either the short-term or long-term block ($p > 0.98$ in both cases).

Table B.1: **Estimated differences between treatments (Study I: Lab)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	beta	delta	beta	delta	#later alloc. (short)	#later alloc. (long)	#later alloc. (short)	#later alloc. (long)
H	-0.003	0.000	0.001	0.001	-0.000	0.100	0.106	0.171
	(0.022)	(0.004)	(0.023)	(0.005)	(0.547)	(0.574)	(0.579)	(0.612)
	[0.899]	[0.943]	[0.968]	[0.872]	[1.000]	[0.862]	[0.855]	[0.780]
B	-0.037	-0.003	-0.030	-0.002	-1.100*	-0.350	-0.897	-0.274
	(0.023)	(0.005)	(0.025)	(0.005)	(0.579)	(0.601)	(0.637)	(0.646)
	[0.102]	[0.571]	[0.236]	[0.656]	[0.060]	[0.561]	[0.162]	[0.672]
Constant	0.840***	0.938***	0.835***	0.945***	5.600***	2.700***	5.389***	3.470*
	(0.016)	(0.003)	(0.060)	(0.013)	(0.422)	(0.433)	(1.471)	(1.800)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.056]
Observations	116	120	114	118	120	120	118	118
R-squared	0.030	0.005	0.063	0.015	0.043	0.006	0.066	0.015
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	0.839	0.937	0.839	0.937	5.601	2.701	5.601	2.701
H-B	0.034	0.003	0.031	0.003	1.100	0.450	1.003	0.445
p(H-B=0)	0.108	0.492	0.159	0.496	0.040**	0.425	0.063*	0.440

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, and CRT score. + MCG refers to the Mean for the Control Group (R treatment). Subjects making inconsistent choices are excluded from the analysis of beta-delta. *p<0.1, **p<0.05, ***p<0.01.

To study differences in variance, the results of a series of variance ratio tests is included in Table B.2. Since the hypothesis is that real incentives trigger less noisy decisions, we conduct one-tailed tests against this hypothesis. Panel i) shows the SD of the mean for each variable by treatment. It can be seen that, against our hypothesis, R yields the highest SD. Panel ii) confirms that the ratio of the standard deviation between R and H is not significantly lower than one for any of the outcome variables ($p > 0.80$). In addition, H does not increase inconsistency in any of the two blocks compared to R (see Table B.3, $p > 0.30$). Yet, as mentioned, the percentages of inconsistent individuals are extremely low.

Table B.2: **Variance ratio test for the outcome variables (Study I: Lab)**

	(1)	(2)	(3)	(4)
	Beta	Delta	#later alloc. (short)	#later alloc. (long)
i) Standard deviation by treatment				
SD(*R*)	0.103	0.021	2.668	2.738
SD(*H*)	0.089	0.018	2.204	2.388
SD(*B*)	0.098	0.020	2.511	2.636
ii) R vs H				
SD(*R*)/*SD*(*H*)	1.153	1.167	1.210	1.146
P (*ratio* < 1)	0.812	0.841	0.881	0.802
iii) R vs B				
SD(*R*)/*SD*(*B*)	1.051	1.050	1.063	1.039
P (*ratio* < 1)	0.622	0.552	0.6461	0.593

Note: The null hypothesis is that the ratio between the standard deviation of the variable in the R group and the standard deviation in the H (or B) group is smaller than 1. *p < 0.1, **p < 0.05, ***p < 0.01.

Table B.3: **OLS estimation of the effects on Inconsistency (Study I: Lab)**

	(1) Inconsistency Short	(2) Inconsistency Short
H	0.025	0.022
	(0.025)	(0.023)
	[0.319]	[0.355]
B	0.075*	0.066*
	(0.042)	(0.040)
	[0.078]	[0.097]
Constant	0.000	0.039
	(0.000)	(0.071)
	[1.000]	[0.578]
Observations	120	118
R-squared	0.030	0.073
Controls	No	Yes

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, and CRT score. There are no estimates for Inconsistency in the long-term block since there are no inconsistencies in this case. *p < 0.1, **p < 0.05, ***p < 0.01.

B.1.2 Are BRIS and real choices different (Q2)?

Now we focus on the comparison between BRIS and real payments. Columns 1 and 2 of Table B.1 show that the dummy for BRIS (B) payments does not have a significant impact on beta or delta ($p > 0.10$), suggesting that BRIS decisions do not differ from fully incentivized decisions (R). This result holds after adding controls (columns 3 and 4, $p > 0.23$).

On the other hand, column 5 shows that B yields a negative and marginally significant effect on the number of later allocations in the short-term block ($p = 0.06$). After adding the control variables (column 7), however, B is no longer significant ($p = 0.16$). No effect is found for the number of later allocations in the long-term block (columns 6 and 8, $p > 0.56$). A *Kolmogorov-Smirnov test* found no statistical differences between B and R in terms of the distribution of the number of later allocations in the long-term block ($p > 0.90$), but a close to significance difference in the short-term block ($p = 0.10$).

Panel iii) in Table B.2 shows that the ratio of the standard deviation between R and B is not significantly lower than one for any of the outcome variables ($p > 0.55$). However, Table B.3 shows that B marginally increases inconsistency rates by about 7 percentage

points compared to R (*p = 0.08 and p = 0.10 with controls*).

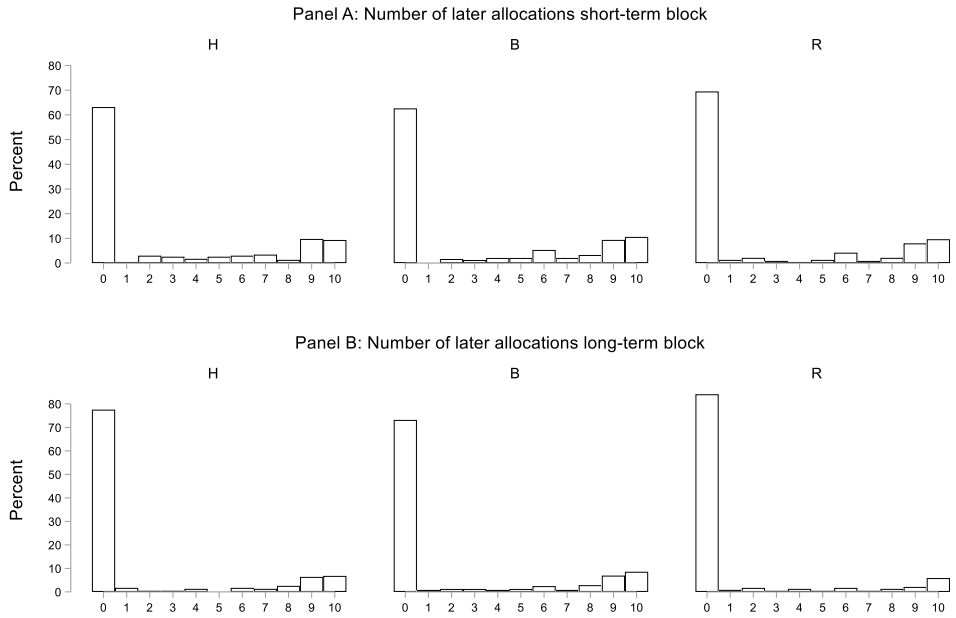
B.1.3 Hypothetical vs. BRIS payments

Finally, the last two rows of Table B.1 compare hypothetical and BRIS choices. Columns 1 and 2 show that there are no significant differences between both treatments in beta or delta ($p > 0.11$) and this holds after adding controls (columns 3 and 4, $p > 0.16$). On the other hand, column 5 shows that H increases the number of later allocations in the short-term block compared to B ($p = 0.04$), while there is no difference on the number of later allocations in the long-term block (*p > 0.40*). These results are robust to controls (*p = 0.06* and *p > 0.40*, respectively).

B.2 Results from the field experiment (Study II)

Figure B.2 shows the distribution of choices in the three treatments in the field experiment. Table B.4 provides the main results. We follow the same regression analysis as in Table B.1. All regressions control for enumerator fixed effects; and only columns 3, 4, 7, and 8 include controls for age, gender, education level (from 1 = "no education", to 19 = "postgraduate") and income (=1 if they report to have enough money to feed the family). None of these control variables are significant in the regressions ($p > 0.18$) except education ($p = 0.08$). Nonetheless, some enumerator dummies yield significance, implying that enumerators did have an influence on the outcomes and therefore regressions should control for this.

Figure B.2: **Histogram of the number of later allocations in the short- and long-term blocks (Study II: Field).**



B.2.1. Are hypothetical and real choices different (Q1)?

Columns 1 to 2 in Table B.4 show that the use of hypothetical payments (H) does not have any significant impact on beta or delta ($p = 0.78$ and $p = 0.19$, respectively). This result holds after adding the control variables (columns 3 and 4).

Regarding the number of later allocations (columns 5-8), H does not yield significant estimates on either the short-term or the long-term block ($p > 0.19$ in both cases).

Also, the Kolmogorov-Smirnov test found no statistical differences between the distribution of *H* and *R* in the number of later allocations in the short-term and long-term block (*p > 0.50* in both cases; see Figure B.2).

Table B.4: **Estimated differences between treatments (Study II: Field)**

	(1)	(2)	(3)	(4)	(5) #later alloc.	(6) #later alloc.	(7) #later alloc.	(8) #later alloc.
	beta	delta	beta	delta	(short)	(long)	(short)	(long)
H	-0.003	0.003	-0.006	0.003	0.055	0.367	-0.005	0.339
	(0.013)	(0.002)	(0.013)	(0.002)	(0.340)	(0.278)	(0.335)	(0.275)
	[0.784]	[0.189]	[0.653]	[0.221]	[0.872]	[0.186]	[0.988]	[0.219]
B	-0.002	0.005**	-0.002	0.004**	0.070	0.550**	0.061	0.540**
	(0.013)	(0.002)	(0.012)	(0.002)	(0.333)	(0.269)	(0.330)	(0.268)
	[0.852]	[0.039]	[0.847]	[0.041]	[0.832]	[0.041]	[0.854]	[0.045]
Constant	0.719***	0.924***	0.774***	0.930***	1.705*	0.792	3.240***	1.553*
	(0.036)	(0.006)	(0.045)	(0.007)	(0.961)	(0.770)	(1.169)	(0.916)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.076]	[0.304]	[0.006]	[0.090]
Observations	717	716	717	716	721	721	721	721
R-squared	0.289	0.338	0.305	0.344	0.315	0.344	0.331	0.350
Enum. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	0.742	0.929	0.742	0.929	2.475	1.383	2.475	1.383
H-B	-0.001	-0.002	-0.003	-0.002	-0.016	-0.183	-0.066	-0.201
p(H-B=0)	0.925	0.495	0.789	0.457	0.962	0.531	0.839	0.493

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, sufficient income (equal to 1 if they have enough money to feed the family in last 7 days), and education. + MCG refers to the Mean for the Control Group (R treatment). Subjects making inconsistent choices are excluded from the analysis of beta-delta. *p < 0.1, **p < 0.05, ***p < 0.01.

Regarding the variance of responses, Table B.5 shows the variance ratio test for each outcome variable. It can be seen in panel *i*) that, except for beta, the *R* treatment displays the lowest SD, as hypothesized. Yet, panel *ii)* shows that the difference between *R* and *H* is not significant for either beta or the number of later allocations in the short-term block ($p > 0.41$), while it is significant for both delta and the number of later allocations in the long-term block ($p < 0.01$). This evidence suggests that, in the field, hypothetical (vs. real) incentives increase the variance of responses in TD tasks. However, this is true for long-term but not short-term discounting. The increase in long-term discounting SD is about 21%, which means that in order to obtain identical 95% confidence intervals for the estimations, the sample in H must be almost

50% larger than in R.

Regarding inconsistency, Table B.6 shows that H does not increase inconsistency rates ($p > 0.3$) compared to R.

Table B.5: **Variance ratio test for the outcome variables (Study II: Field)**

	(1) Beta	(2) Delta	(3) #later alloc. (short)	(4) #later alloc. (long)
i) Standard deviation by treatment				
SD(*R*)	0.145	0.023	3.818	2.888
SD(*H*)	0.139	0.028	3.873	3.480
SD(*B*)	0.143	0.029	3.999	3.689
ii) R vs H				
SD(*R*)/*SD*(*H*)	1.043	0.821	0.986	0.829
P (*ratio* < 1)	0.764	0.002***	0.413	0.002***
iii) R vs B				
SD(*R*)/*SD*(*B*)	1.014	0.793	0.956	0.783
P (*ratio* < 1)	0.589	0.000***	0.241	0.000***

Note: The null hypothesis is that the ratio between the standard deviation of the variable in the R group and the standard deviation in the H (or B) group is equal to 1. *p < 0.1, **p < 0.05, ***p < 0.01.

Table B.6: **OLS estimation of the effects on Inconsistency (Study II: Field)**

	(1) Inconsistency Short	(2) Inconsistency Long	(3) Inconsistency Short	(4) Inconsistency Long
H	-0.004		-0.004	
	(0.004)		(0.004)	
	[0.318]		[0.337]	
B	0.008	0.008	0.008	0.008
	(0.008)	(0.006)	(0.008)	(0.006)
	[0.327]	[0.157]	[0.313]	[0.158]
Constant	0.004	-0.000	0.019	0.010
	(0.004)	(0)	(0.014)	(0.007)
	[0.318]	[0]	[0.190]	[0.176]
Observations	721	721	721	721
R-squared	0.005	0.005	0.008	0.007
Controls	No	No	Yes	Yes

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, and education. Note that there are no inconsistencies in treatments R or H in the long-term block. *p < 0.1, **p < 0.05, ***p < 0.01.

B.2.2. Are BRIS and real choices different (Q2)?

Now we compare BRIS results with those from real payments. Columns 1-2 in Table B.4 show that the B dummy yields a positive and statistically significant effect on delta ($p = 0.04$), while it is not significant for beta ($p = 0.85$). Adding

controls does not change the picture (columns 3-4).

Similarly, B is significantly positive for later allocations in the long-term block (with and without controls, columns 6 and 8, $p < 0.05$), but non-significant for allocations in the short-term block. Kolmogorov-Smirnov tests confirm these results also for the distributions between B and R on these outcomes ($p = 0.08$ and $p > 0.30$, respectively).

Regarding the variance of responses, Panel *iii)* in Table B.5 shows that the ratio of the SD of the outcome variables between R and B is not significantly different from 1 for beta and short-term later allocations ($p > 0.24$). However, the variance in both delta and long-term later allocations is significantly higher in B compared to R ($p < 0.01$). The increase in SD is about 27%, meaning that to get identical 95% confidence intervals for the estimations, the sample in *B* must be about 60% larger than in *R*. Importantly, note that *B* does not yield smaller SD than *H* for any of the outcome variables, but even slightly larger.

Regarding inconsistency, Table B.6 shows that B does not affect inconsistency in either of the two blocks ($p > 0.15$).

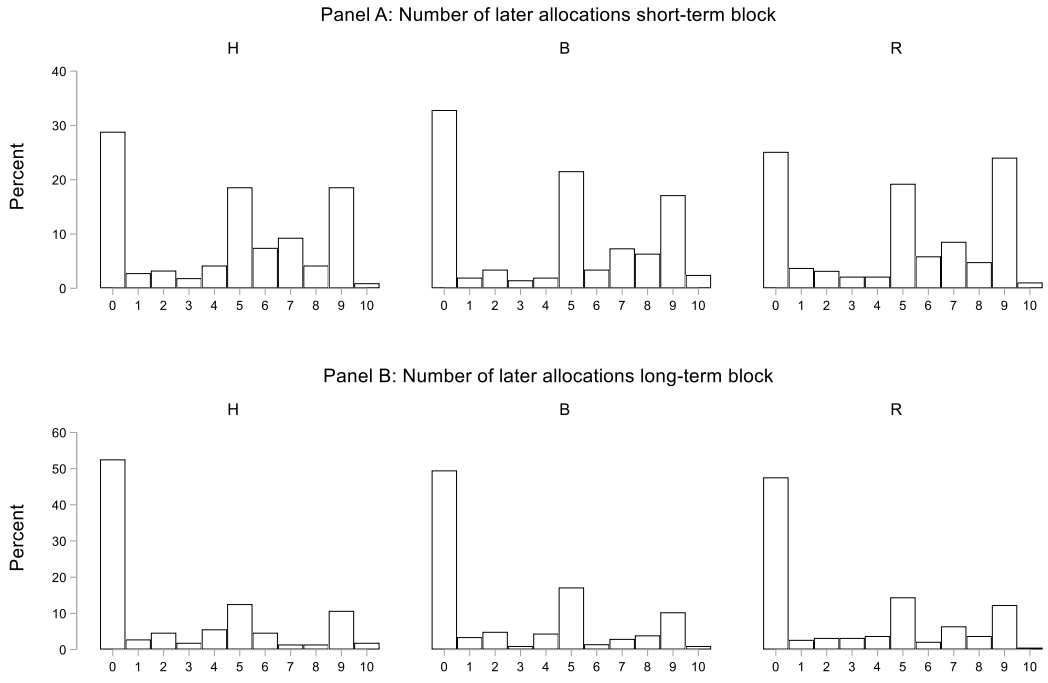
B.2.3 Hypothetical vs. BRIS payments

Finally, when comparing B and H in the last two rows of Table B.4 we can see that there are no significant differences between both treatments ($p > 0.40$).

B.3 Results from the online experiment Prolific 1 (Study III)

Figure B.3 shows the distribution of choices in the three treatments in the online experiment Prolific 1. Table B.7 provides the main results. Columns 1 to 4 display the results when the dependent variable is β or δ from the beta-delta model. In columns 5 to 8 the dependent variable is the number of later allocations in the two blocks (short-term or long-term). Columns 3, 4, 7, and 8 include controls for age, gender, education level (from 1 = "no education", to 19 = "postgraduate"), socio-economic status (Likert scale from 1 to 10) and the number or risky choices. Some of these control variables are significant in the regressions: education for all the outcomes ($p < 0.03$), gender and number of risky choices for beta and number of later allocations in the short-term block ($p = 0.02$ in both variables).

Figure B.3: **Histogram of the number of later allocations in the short and long-term blocks (Study III: Prolific 1).**



B.3.1 Are hypothetical and real choices different (Q1)?

Columns 1 and 2 of Table B.7 show that H has no significant impact on beta or delta ($p > 0.20$). After adding the control variables (columns 3 and 4), the coefficients remain non-significant ($p > 0.20$).

Regarding the number of later allocations, columns 5 to 8 show that H has no impact on either the short-term or the long-term block ($p > 0.20$ with and without controls). Also, the Kolmogorov-Smirnov test suggests that H and R do not have different distributions in either of the two outcomes ($p > 0.80$ and $p > 0.60$ respectively; see Figure B.3).

Regarding the variance of responses, Table B.8 shows the variance ratio test for each outcome variable. It can be seen in panel *i*) that, except for beta, the H treatment displays the lowest SD, as was found in Study I. However, panel *ii)* shows that the difference between R and H is not significant for either of the outcome variables ($p > 0.39$).

Table B.7: **Estimated differences between treatments (Study III: Prolific 1)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	beta	delta	beta	delta	#later alloc. (short)	#later alloc. (long)	#later alloc. (short)	#later alloc. (long)
H	-0.013	-0.003	-0.014	-0.003	-0.377	-0.431	-0.376	-0.432
	(0.013)	(0.003)	(0.012)	(0.003)	(0.347)	(0.340)	(0.345)	(0.342)
	[0.293]	[0.201]	[0.270]	[0.201]	[0.278]	[0.205]	[0.276]	[0.207]
B	-0.017	-0.002	-0.015	-0.002	-0.515	-0.268	-0.469	-0.239
	(0.013)	(0.003)	(0.013)	(0.003)	(0.358)	(0.344)	(0.357)	(0.345)
	[0.188]	[0.421]	[0.241]	[0.468]	[0.151]	[0.437]	[0.190]	[0.489]
Constant	0.811***	0.941***	0.624***	0.920***	4.824***	3.096***	0.604	0.305
	(0.009)	(0.002)	(0.054)	(0.011)	(0.256)	(0.253)	(1.631)	(1.369)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.711]	[0.824]
Observations	594	599	593	598	606	606	605	605
R-squared	0.003	0.003	0.044	0.017	0.004	0.003	0.037	0.018
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	0.811	0.941	0.811	0.941	4.823	3.096	4.823	3.096
H-B	0.004	-0.001	0.001	-0.001	0.138	-0.163	0.093	-0.193
p(H-B=0)	0.771	0.625	0.920	0.562	0.688	0.617	0.786	0.555

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, socioeconomic status, education and number of risky choices. + MCG refers to the Mean for the Control Group (R treatment). Subjects making inconsistent choices are excluded from the analysis of beta-delta. *p < 0.1, **p < 0.05, ***p < 0.01.

Table B.8: **Variance ratio test for the outcome variables (Study III: Prolific 1)**

	(1) Beta	(2) Delta	(3) #later alloc. (short)	(4) #later alloc. (long)
i) Standard deviation by treatment				
SD(*R*)	0.123	0.026	3.502	3.454
SD(*H*)	0.124	0.026	3.442	3.338
SD(*B*)	0.125	0.026	3.572	3.342
ii) R vs H				
SD(*R*)/*SD*(*H*)	0.998	1.031	1.017	1.035
P (*ratio* < 1)	0.507	0.400	0.402	0.312
iii) R vs B				
SD(*R*)/*SD*(*B*)	0.984	0.967	0.980	1.034
P (*ratio* < 1)	0.412	0.552	0.393	0.679

Note: The null hypothesis is that the ratio between the standard deviation of the variable in the R group and the standard deviation in the H (or B) group is smaller than 1. *p < 0.1, **p < 0.05, ***p < 0.01.

As Table B.9 shows, H does not increase inconsistency rates either in the short-term or the long-term block ($p > 0.10$ without controls and $p > 0.16$ with controls).

Table B.9: **OLS estimation of the effect on Inconsistency (Study III: Prolific 1)**

	(1) Inconsistency Short	(2) Inconsistency Long	(3) Inconsistency Short	(4) Inconsistency Long
H	0.022	0.008	0.020	0.007
	(0.014)	(0.012)	(0.014)	(0.012)
	[0.126]	[0.507]	[0.165]	[0.564]
B	0.004	-0.006	0.002	-0.006
	(0.011)	(0.009)	(0.012)	(0.009)
	[0.723]	[0.520]	[0.877]	[0.462]
Constant	0.011	0.011	0.090	0.085*
	(0.008)	(0.008)	(0.109)	(0.044)
	[0.157]	[0.157]	[0.406]	[0.053]
Observations	606	606	605	605
R-squared	0.005	0.003	0.013	0.011
Controls	No	No	Yes	Yes

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, education, number of risky choices in the Holt-Laury (2002) task and socio-economic status. *p < 0.1, **p < 0.05, ***p < 0.01.

B.3.2 Are BRIS and real choices different (Q2)?

Now we focus on the comparison between BRIS and real payments. Columns 1 and 2 of Table B.7 show that B has not a significant impact on beta or delta ($p > 0.18$),

suggesting that BRIS decisions do not differ from fully incentivized decisions (R). This result holds after adding controls (columns 3 and 4, $p > 0.23$).

Also, columns 5 to 8 show that B has no impact on the number of later allocations in either the short-term or the long-term block ($p > 0.15$ with and without controls). Additionally, the Kolmogorov-Smirnov test found no statistical differences between the distribution of B and R in either of the two outcomes ($p > 0.50$ and $p > 0.80$, respectively).

Regarding the variances of responses, Panel *iii)* in Table B.8 shows that the ratio of the standard deviation between R and B is not significantly lower than one for any of the outcome variables ($p > 0.40$). Also, Table B.9 shows that B does not affect inconsistency rates in either block compared to R ($p >0.40$ with controls).

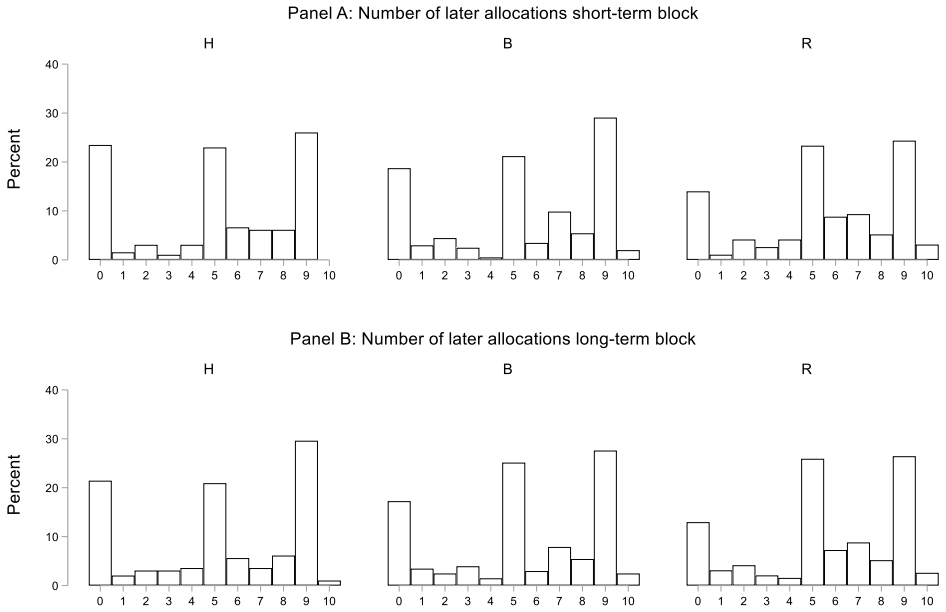
B.3.3 Hypothetical vs. BRIS payments

Finally, the last two rows of Table B.7 compare hypothetical and BRIS choices. No significant difference was found in any of the outcome variables ($p > 0.55$).

B.4 Results from the online experiment Prolific 2 (Study IV)

Figure B.4 shows the distribution of choices in the three treatments in the online experiment Prolific 2. Table B.10 provides the main results. As before, Columns 1 to 4 display the results for β or δ from the beta-delta model and columns 5 to 8 for the number of later allocations in the two blocks (short-term or long-term). Columns 3, 4, 7, and 8 include the same controls as in Study III. Some of these control variables are significant in the regressions: education for beta ($p = 0.01$); age for delta ($p = 0.06$) and the number of risky choices for delta and number of later allocations in the short- and long-term block ($p < 0.10$).

Figure B.4: **Histogram of the number of later allocations in the short and long-term blocks (Study IV: Prolific 2).**



B.4.1 Are hypothetical and real choices different (Q1)?

Columns 1 to 4 in Table B.10 show that H has no significant impact on beta or delta ($p > 0.30$). Regarding the number of later allocations, columns 5 show that H has a marginally significant effect on the number of latter allocations in the short-term ($p = 0.09$), but this effect is no longer significant after adding controls (see column 7, $p > 0.11$). Columns 6 and 8 show that H has no effect on the long-term block ($p > 0.18$). The distributions are also not different according to the Kolmogorov-Smirnov test for any of the two outcomes ($p > 0.40$ and $p > 0.50$ respectively, see Figure B.4).

Table B.10: **Estimated differences between treatments (Study IV: Prolific 2)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	beta	delta	beta	delta	#later alloc. (short)	#later alloc. (long)	#later alloc. (short)	#later alloc. (long)
H	-0.007	-0.013	-0.006	-0.011	-0.565*	-0.433	-0.514	-0.395
	(0.013)	(0.013)	(0.013)	(0.013)	(0.329)	(0.332)	(0.331)	(0.336)
	[0.599]	[0.323]	[0.643]	[0.385]	[0.086]	[0.193]	[0.121]	[0.239]
B	0.003	-0.007	0.003	-0.004	-0.187	-0.228	-0.141	-0.181
	(0.013)	(0.013)	(0.013)	(0.012)	(0.326)	(0.323)	(0.323)	(0.320)
	[0.803]	[0.605]	[0.795]	[0.730]	[0.566]	[0.479]	[0.662]	[0.571]
Constant	1.006***	0.794***	1.049***	0.756***	5.601***	5.637***	5.110***	4.373***
	(0.009)	(0.009)	(0.043)	(0.045)	(0.222)	(0.223)	(1.279)	(1.208)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Obs	583	584	583	584	592	592	592	592
R-squared	0.001	0.002	0.017	0.023	0.005	0.003	0.025	0.021
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	1.005	0.800	1.005	0.800	5.487	5.522	5.487	5.522
H-B	-0.010	-0.006	-0.010	-0.007	-0.378	-0.205	-0.373	-0.214
p(H-B=0)	0.435	0.629	0.465	0.589	0.267	0.546	0.272	0.529

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, socioeconomic status, education and number of risky choices. + MCG refers to the Mean for the Control Group (R treatment). Subjects making inconsistent choices are excluded from the analysis of beta-delta. *p < 0.1, **p < 0.05, ***p < 0.01

Regarding the variance of responses panel *ii*) in Table B.11 shows that the difference between R and H is not significant for any of the outcome variables ($p > 0.39$). Also, Table B.12 shows that H does not increase inconsistency rates in any of the two blocks ($p > 0.12$ without controls and $p > 0.16$ with controls).

Table B.11: **Variance ratio test for the outcome variables (Study IV: Prolific 2)**

	(1) Beta	(2) Delta	(3) #later alloc. (short)	(4) #later alloc. (long)
i) Standard deviation by treatment				
SD(*R*)	0.127	0.123	3.147	3.1714
SD(*H*)	0.125	0.131	3.397	3.446
SD(*B*)	0.132	0.129	3.406	3.317
ii) R vs H				
SD(*R*)/*SD*(*H*)	1.016	0.935	0.926	0.920
P (*ratio* < 1)	0.588	0.174	0.143	0.123
iii) R vs B				
SD(*R*)/*SD*(*B*)	0.961	0.960	0.924	0.956
P (*ratio* < 1)	0.289	0.283	0.133	0.263

Note: The null hypothesis is that the ratio between the standard deviation of the variable in the R group and the standard deviation in the H (or B) group is smaller than 1. *p < 0.1, **p < 0.05, ***p < 0.01.

Table B.12: **OLS estimation of the effect on Inconsistency (Study IV: Prolific 2)**

	(1) Inconsistency Short	(2) Inconsistency Long	(3) Inconsistency Short	(4) Inconsistency Long
H	-0.000	0.005	0.001	0.006
	(0.014)	(0.013)	(0.015)	(0.013)
	[0.982]	[0.719]	[0.925]	[0.675]
B	-0.016	-0.011	-0.015	-0.010
	(0.011)	(0.010)	(0.011)	(0.010)
	[0.166]	[0.298]	[0.191]	[0.345]
Constant	0.021**	0.016*	0.031	0.040
	(0.010)	(0.009)	(0.036)	(0.038)
	[0.044]	[0.082]	[0.398]	[0.292]
Observations	592	592	592	592
R-squared	0.004	0.003	0.012	0.007
Controls	No	No	Yes	Yes

Note: OLS estimates. Robust standard errors in parentheses and p-values in brackets. Controls are age, gender, education, number of risky choices in the Holt-Laury (2002) task and socio-economic status. *p < 0.1, **p < 0.05, ***p < 0.01.

B.4.2 Are BRIS and real choices different (Q2)?

Columns 1 and 2 of Table B.10 show that B has not a significant impact on beta or delta ($p > 0.60$). This result holds after adding controls (columns 3 and 4, $p > 0.70$). Also, columns 5 to 8 show that B has no impact on either the short-term or the long-term block in terms of number of later allocations ($p > 0.50$ with and without controls). Additionally, the Kolmogorov-Smirnov test found no statistical differences between the distribution of B and R in any of the two outcomes ($p > 0.90$ in both cases).

As before, neither the variance of responses ($p > 0.13$; Panel iii) in Table B.11) or the inconsistency rate ($p > 0.16$; Table B.12) differ significantly between B and R.

B.4.3 Hypothetical vs. BRIS payments

Finally, the last two rows of Table B.10 compare hypothetical and BRIS choices. No significant difference was found for any of the outcome variables ($p > 0.26$). These results suggest that both choices are the same.

C. Equivalence Test: Are *H* and *B* equivalent to *R*?

In this section we analyze for which ranges (in terms of Cohen's d or SD) the overall estimates of our dependent variables are equivalent in treatments H or B as compared to R. Note that the fact that p-values are larger than alpha (i.e., 0.05, or 0.10 for marginal significance) does not certify the absence of effect (Wagenmakers, 2007). They only tell us that we cannot reject the hypothesis that the effect is zero. To reject the hypothesis that the effect is different from zero, that is, to conclude that the true effect size is exactly zero, we would need a huge sample size (Lakens, 2018). One reasonable alternative is to ask whether the observed effect is large enough to be deemed worthwhile. This technique is called equivalence testing (Lakens, 2017; Wellek, 2010) and is based on testing whether the observed effect falls within or outside an equivalence interval.

To test for equivalence, a two one-sided test (TOST) approach is applied in which two composite null hypotheses are tested: $H01 \rightarrow \gamma \leq -\gamma L$ and $H02 \rightarrow \gamma \geq \gamma U$, where γL and γU refer to the lower and upper bounds of the equivalence interval. When both null hypotheses are rejected, we can conclude that $-\gamma L < \gamma < \gamma U$ or, in other words, that the observed effect falls within the equivalence bounds and it is close enough to zero to be practically equivalent (Lakens, 2017).

The challenge of this procedure is to objectively define the lower and upper bounds of the equivalence interval. For the sake of conciseness, we will provide the range around the estimated value in treatment R for which both H and B would be deemed equivalent to R. These ranges are given in terms of SDs of the variable in R. The analysis is performed using the overall effects obtained from the random-effects meta-analysis of the main text.

Following Lakens et al. (2018), we use the 90% CI of the estimates because in this way two one-sided tests are performed with an $\alpha = 5\%$ each. There are four possible outcomes in the analysis. The observed effect can be:

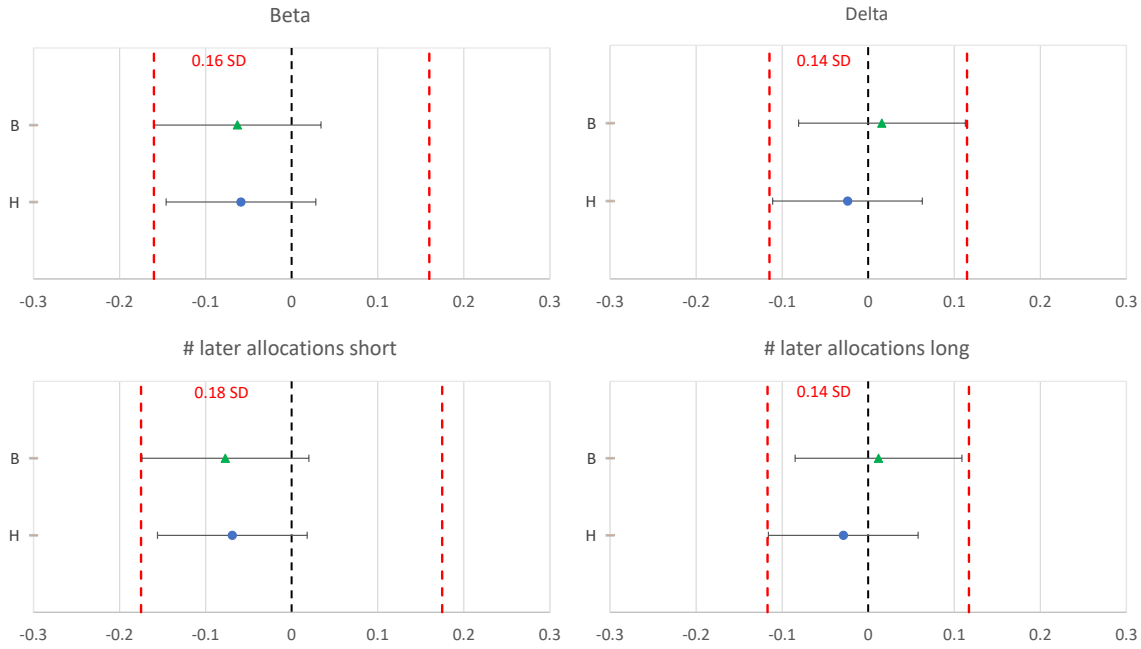
- both statistically indistinguishable from zero and statistically equivalent ($-\gamma L < \gamma < \gamma U$) – this is labeled as Equivalence;
- statistically different from zero and not statistically equivalent (Relevant Difference);
- statistically different from zero but statistically equivalent (Trivial Difference);

- neither statistically different from zero nor statistically equivalent (Undetermined).

Since the overall effects are never statistically different from zero, our analysis would never end up in either Relevant or Trivial Difference. We will therefore provide the range for which any effect goes from Undetermined to Equivalence. Increasing (reducing) the range used to determine equivalence would lead to a greater (smaller) probability of obtaining an Equivalence result.

Figure C.1 shows the overall coefficients, their 90% CI and the upper and lower bounds defining the interval beyond which both H and B would be equivalent to R.

Figure C.1: **Equivalence tests for the main results (overall)**



We find that for any equivalence interval larger than 0.16 SD around the beta estimated in R, both H and B are equivalent to R at 95% confidence level. That is, we can reject with 95% confidence that hypothetical or BRIS payments change beta by more than 0.16 SD with respect to real payments. For delta, the interval that deems H and B equivalent to R is 0.14 SD. Regarding the number of later allocations in the short- and long-term blocks, the values of the minimum interval for equivalence are 0.18 SD and 0.14 SD around the value in R, respectively. Thus, in aggregate terms, treatments H and B yield not only similar, but also fairly equivalent TD values as treatment R.

D. Alternative specifications: Interval and negative-binomial regressions

As robustness checks, we run interval regressions for beta and delta on the different treatment dummies. Both beta and delta are actually measured in intervals (although the main results use the upper bound of the intervals to define each variable), and thus all observations are right and/or left censored. Therefore, we re-estimate the regressions using interval regression techniques (e.g., Harrison et al., 2002). We also run a negative binomial model for the number of later allocations in the short- and long-term blocks because these can be considered count variables.

The results for the lab (Study I) are shown in Table D.1 (interval regressions in columns 1-4, negative binomial regressions in columns 5-8). The coefficients of the treatments variables are very similar to those estimated in Table B.1: H is never significant while B is marginally significant for beta and the number of later allocations in the short-term.

Table D.1: **Interval and negative binomial regressions (Study I: Lab)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					# later alloc.	# later alloc.	# later alloc.	# later alloc.
	beta	beta	delta	delta	(short)	(short)	(long)	(long)
H	-0.003	0.002	0.001	0.002	-0.000	0.095	0.104	0.160
	(0.022)	(0.006)	(0.022)	(0.006)	(0.507)	(0.543)	(0.534)	(0.573)
	[0.898]	[0.795]	[0.967]	[0.720]	[1.000]	[0.861]	[0.845]	[0.780]
B	-0.037*	-0.003	-0.030	-0.002	-1.144*	-0.363	-0.934	-0.228
	(0.022)	(0.006)	(0.025)	(0.007)	(0.602)	(0.620)	(0.649)	(0.655)
	[0.097]	[0.646]	[0.222]	[0.789]	[0.057]	[0.558]	[0.150]	[0.728]
Constant	0.843***	0.929***	0.839***	0.933***				
	(0.016)	(0.005)	(0.059)	(0.019)				
	[0.000]	[0.000]	[0.000]	[0.000]				
Observations	116	120	114	118	120	120	118	118
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	0.839	0.937	0.839	0.937	5.601	2.701	5.601	2.701

Note: Robust standard errors in parentheses and p-values in brackets. *p < 0.1, **p < 0.05, ***p < 0.01.

Table D.2 provides the results of the interval and negative binomial regressions for the field experiment (Study II). The coefficients of the treatment's variables are very similar than those estimated in Tables B.4: H is never significant while B is significant for delta and the number of later allocations in the long-term.

Tables D.3 and D.4 provide the results of the interval and negative binomial regressions

for the online experiments (Study III and IV). The coefficients are very similar than those estimated on Tables B.7 and B.10. These suggest that results are robust to different estimations models.

Table D.2: **Interval and negative binomial regressions (Study II: Field)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					# later alloc.	# later alloc.	# later alloc.	# later alloc.
	Beta	delta	beta	delta	(short)	(short)	(short)	(long)
H	-0.004	0.019	-0.006	0.015	0.272	0.673*	0.256	0.614
	(0.013)	(0.013)	(0.012)	(0.013)	(0.378)	(0.355)	(0.381)	(0.557)
	[0.741]	[0.147]	[0.607]	[0.252]	[0.472]	[0.058]	[0.502]	[0.270]
B	-0.003	0.028**	-0.003	0.025**	0.476	0.972***	0.522	1.011*
	(0.012)	(0.013)	(0.012)	(0.013)	(0.368)	(0.338)	(0.372)	(0.573)
	[0.819]	[0.028]	[0.813]	[0.048]	[0.197]	[0.004]	[0.161]	[0.078]
Constant	0.723***	0.813***	0.779***	0.850***				
	(0.036)	(0.035)	(0.045)	(0.041)				
	[0.000]	[0.000]	[0.000]	[0.000]				
Observations	717	716	717	716	721	721	721	721
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MCG+	No	No	Yes	Yes	No	No	Yes	Yes

Note: Robust standard errors in parentheses and p-values in brackets. *p < 0.1, **p < 0.05, ***p < 0.01.

Table D.3: **Interval and negative binomial regressions (Study III: Prolific 1)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					# later alloc.	# later alloc.	# later alloc.	# later alloc.
	beta	beta	delta	delta	(short)	(short)	(long)	(long)
H	-0.013	-0.006	-0.014	-0.006	-0.368	-0.428	-0.381	-0.425
	(0.013)	(0.006)	(0.012)	(0.006)	(0.338)	(0.336)	(0.344)	(0.343)
	[0.305]	[0.345]	[0.277]	[0.343]	[0.276]	[0.203]	[0.268]	[0.216]
B	-0.017	-0.006	-0.015	-0.006	-0.510	-0.258	-0.490	-0.276
	(0.013)	(0.006)	(0.013)	(0.006)	(0.354)	(0.331)	(0.360)	(0.340)
	[0.187]	[0.329]	[0.238]	[0.347]	[0.150]	[0.436]	[0.173]	[0.417]
Constant	0.814***	0.909***	0.626***	0.879***				
	(0.009)	(0.005)	(0.054)	(0.027)				
	[0.000]	[0.000]	[0.000]	[0.000]				
Observations	594	599	593	598	606	606	605	605
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG+	0.811	0.941	0.811	0.941	4.823	3.096	4.823	3.096

Note: Robust standard errors in parentheses and p-values in brackets. *p < 0.1, **p < 0.05, ***p < 0.01.

Table D.4: **Interval and negative binomial regressions (Study IV: Prolific 2)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					# later alloc.	# later alloc.	# later alloc.	# later alloc.
	beta	beta	delta	delta	(short)	(short)	(long)	(long)
H	-0.008	-0.019	-0.007	-0.017	-0.569*	-0.433	-0.517	-0.370
	(0.013)	(0.015)	(0.013)	(0.015)	(0.332)	(0.332)	(0.338)	(0.338)
	[0.555]	[0.200]	[0.606]	[0.250]	[0.086]	[0.193]	[0.126]	[0.275]
B	0.003	-0.009	0.003	-0.007	-0.182	-0.224	-0.152	-0.167
	(0.013)	(0.014)	(0.013)	(0.014)	(0.316)	(0.316)	(0.316)	(0.315)
	[0.808]	[0.518]	[0.792]	[0.626]	[0.565]	[0.478]	[0.631]	[0.597]
Constant	1.031***	0.767***	1.075***	0.718***				
	(0.009)	(0.010)	(0.043)	(0.055)				
	[0.000]	[0.000]	[0.000]	[0.000]				
Observations	583	584	583	584	592	592	592	592
Controls	No	No	Yes	Yes	No	No	Yes	Yes
MCG	0.811	0.941	0.811	0.941	4.823	3.096	4.823	3.096

Note: Robust standard errors in parentheses and p-values in brackets. *p < 0.1, **p < 0.05, ***p < 0.01.

E. Study V: Methodological issues in hypothetical TD (Online HB)

We have shown that hypothetical payoffs provide basically the same information as real incentives in TD tasks. However, hypothetical TD might be affected by methodological issues related to the design of the whole survey or experiment. Study V (labeled as "Online HB") allows us to examine in detail how certain design features, which are common in large-scale experiments and surveys, affect discounting using hypothetical payments. Thus, we want to know more about the performance of hypothetical incentives, given that they seem to be a valid alternative to real ones in TD elicitation. In particular, we study the impact of within-task order (i.e., either the short-term or the long-term block first), and the possible contamination arising from the existence of previous (paid) tasks.

E.1 Implementation and sample

Study V has a different design than Studies I to IV. To answer the aforementioned questions, we implemented a 2x2x2x2 between-subjects design. Subjects were randomly assigned to each condition.

The first arm refers to the use of BRIS vs. hypothetical payments. The other three arms refer to the within-task order, the position of the task, and the use of other paid (vs. hypothetical) tasks before the TD task. The entire sample consists of 637 subjects and 23 made inconsistent choices. The distribution by treatments is as follows:

Hypothetical vs BRIS: The first two treatments refer to the use of Hypothetical (H, n = 315) or BRIS (B, n = 315) payment schemes.

Within-task order: Here we explore whether deciding first either for the short- or the long-term block makes any difference in hypothetical TD. Particularly, we randomly assigned the order of the two blocks: short → long, or long → short (with 332 and 305 observations, respectively).

Position of the task: This arm refers to the order of the task within the entire experiment. We combined experiments with strategic interaction (games) with TD. While in studies I to IV the TD task was set to be always in the same place, either first or third, in Study V we used two sequences: TD → games, or games → TD (with 357 and 280 observations respectively).

Previous paid tasks: Finally, we test the effect of having other tasks which involve real money within the same experimental setup on the elicitation of hypothetical time

preferences. Particularly, we randomly assigned subjects to play all other tasks (strategic games) with either hypothetical or BRIS incentives. Hence the two arms are: the other tasks within the experiment are BRIS vs. hypothetical (with 314 and 323 observations respectively). Actually, since having other (paid) tasks after TD elicitation should not affect the latter because subjects did not learn the payment method before facing each specific block of tasks (i.e. either the games or the TD), we specifically test the interaction between the variables "other tasks are BRIS vs. hypothetical" and "other tasks are before vs. after TD elicitation".

To conduct the experiment, we designed an online platform. The experiment was run between July and August of 2014. Ibercivis Foundation, based in Zaragoza, helped us to disseminate the experiment through its network of collaborators to recruit participants. They used Twitter and other social media to invite people to participate. No other restriction than having an email address and being at least 18 years old was imposed.

As in previous studies, we followed a number of procedures to ensure trust and reduce issues related to payment-uncertainty and transaction costs. These procedures were clearly explained in the instructions. Participants selected for real payments (1 out of 10 among those under BRIS) were notified the same day by email. As in studies I-IV, we randomly selected one out of the 20 MPL decisions to compute final payoffs. We used Amazon gift cards – with specified dates – to pay winners.

Participants faced the same MPL task as in studies I-III with monetary amounts equivalent to a one-day minimum wage (initial amount = 30 euros and final amount = 48 euros). Participants who were selected to be paid earned 32.5 euros on average. We also elicited self-reported risk aversion based on three hypothetical questions.

Participants were on average 39 years old, 26.7% females, 49% had completed university education, 23% were unemployed, and had an average monthly income of 1,031 euros.

All participants gave their informed consent, and the data were anonymized in accordance with the Spanish Law on Personal Data Protection 15/1999.

E.2 Results from Study V (Online HB)

Table E.1 shows the results of the stress test to hypothetical TD (we only consider this treatment). Models 1, 3, 5, and 7 test the main effects of the three dummies that represent

the three treatments (i.e. Games first, Long first, and Paid games) on β, δ, short- and long-term later allocations, respectively. On the other hand, models 2, 4, 6, and 8 add the interactions between the three treatment variables. All the models control for age, gender, education level and household income. Education and female have a significant impact on long-term and beta ($p < 0.01$).

The elicitation of both β and the number of short-term later allocations is sensitive to Games first ($p < 0.01$; columns 1, 5). If other games are played before the TD task, subjects show higher level of short-term patience, according to both measures. Since the interaction Games first*Paid games is not significant (indeed, none of the interactions tested is ever significant; $p > 0.50$), the positive effect of Games first on short-term patience holds regardless of whether the games are paid or not (see columns 2 and 6). In addition, the non-significant interaction between Games first and Long first suggests that within-task order does not moderate the effect of Games first. Also, the sequence long→short (vs. short→long), captured by Long first is marginally associated to a lower beta ($p = 0.08$; column 1). Nevertheless, the remaining regressions suggest that this is not a robust effect.

Table E.1: **Results from Online HB (Study V): Stress test to H**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	beta	beta	delta	delta	# later alloc. (short)	# later alloc. (short)	# later alloc. (long)	# later alloc. (long)
Games first (GF)	0.036***	0.037*	0.005	0.003	1.096***	1.068**	0.695*	0.449
	(0.012)	(0.021)	(0.003)	(0.005)	(0.322)	(0.522)	(0.388)	(0.636)
	[0.004]	[0.071]	[0.123]	[0.564]	[0.001]	[0.042]	[0.074]	[0.481]
Long first (LF)	-0.022*	-0.025	0.003	0.003	-0.428	-0.720	0.345	0.231
	(0.013)	(0.023)	(0.003)	(0.005)	(0.326)	(0.596)	(0.384)	(0.652)
	[0.082]	[0.275]	[0.354]	[0.614]	[0.191]	[0.228]	[0.370]	[0.723]
Paid games (PG)	-0.003	0.000	0.006**	0.008	-0.023	0.043	0.741*	0.918
	(0.012)	(0.023)	(0.003)	(0.005)	(0.324)	(0.587)	(0.386)	(0.668)
	[0.827]	[0.996]	[0.048]	[0.149]	[0.945]	[0.942]	[0.056]	[0.170]
GF*LF		0.005		0.004		0.407		0.558
		(0.025)		(0.006)		(0.645)		(0.771)
		[0.851]		[0.541]		[0.529]		[0.470]
GF*PG		-0.008		0.000		-0.339		-0.052
		(0.025)		(0.006)		(0.640)		(0.763)
		[0.762]		[0.981]		[0.597]		[0.945]
LF*PG		0.002		-0.003		0.209		-0.300
		(0.026)		(0.006)		(0.673)		(0.792)
		[0.948]		[0.615]		[0.756]		[0.705]
Constant	0.884***	0.884***	0.936***	0.936***	6.344***	6.350***	2.509**	2.547**
	(0.038)	(0.039)	(0.009)	(0.009)	(0.966)	(0.982)	(1.104)	(1.119)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.024]	[0.024]
Observations	307	307	314	314	318	318	319	319
R-squared	0.046	0.047	0.055	0.057	0.057	0.059	0.056	0.058
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MCG+	0.881	0.881	0.952	0.952	6.757	6.757	4.553	4.553

Notes: OLS estimates. Robust standard errors in parentheses and p-values in brackets. + MCG refers to the Mean for the Control Group (i.e. the three treatment dummies = 0). *p < 0.1, **p < 0.05, ***p < 0.01.

The elicitation of δ, on the other hand, is robust both to other games being played before and to different within-task orders, while is apparently sensitive to the use of monetary incentives in other tasks: Paid games yields a positive and significant effect (*p = 0.05*; column 3). The effect is similar but marginally significant for long-term later allocations (*p = 0.06*; column 7). Yet, given that the interaction Games first*Paid games is never significant (see columns 4 and 8), this should be considered a spurious result. Since subjects could not know ex-ante whether the games would be paid (BRIS) or hypothetical, we should expect the interaction to be positive and significant, indicating that the observed positive effect of Paid games only exists when the games are played first but not when TD is first. We instead find a similar effect in both conditions. Finally, as for beta and short-term later allocations, playing other games first (paid or not) marginally increases the

number of long-term later allocations ($p = 0.07$; column 7).

Therefore, we can conclude that:

Hypothetical time preferences are robust to different within-task orders (long/short) and to whether other tasks are incentivized. However, patience, especially in the short-term, is larger if the TD task comes after other experimental tasks.

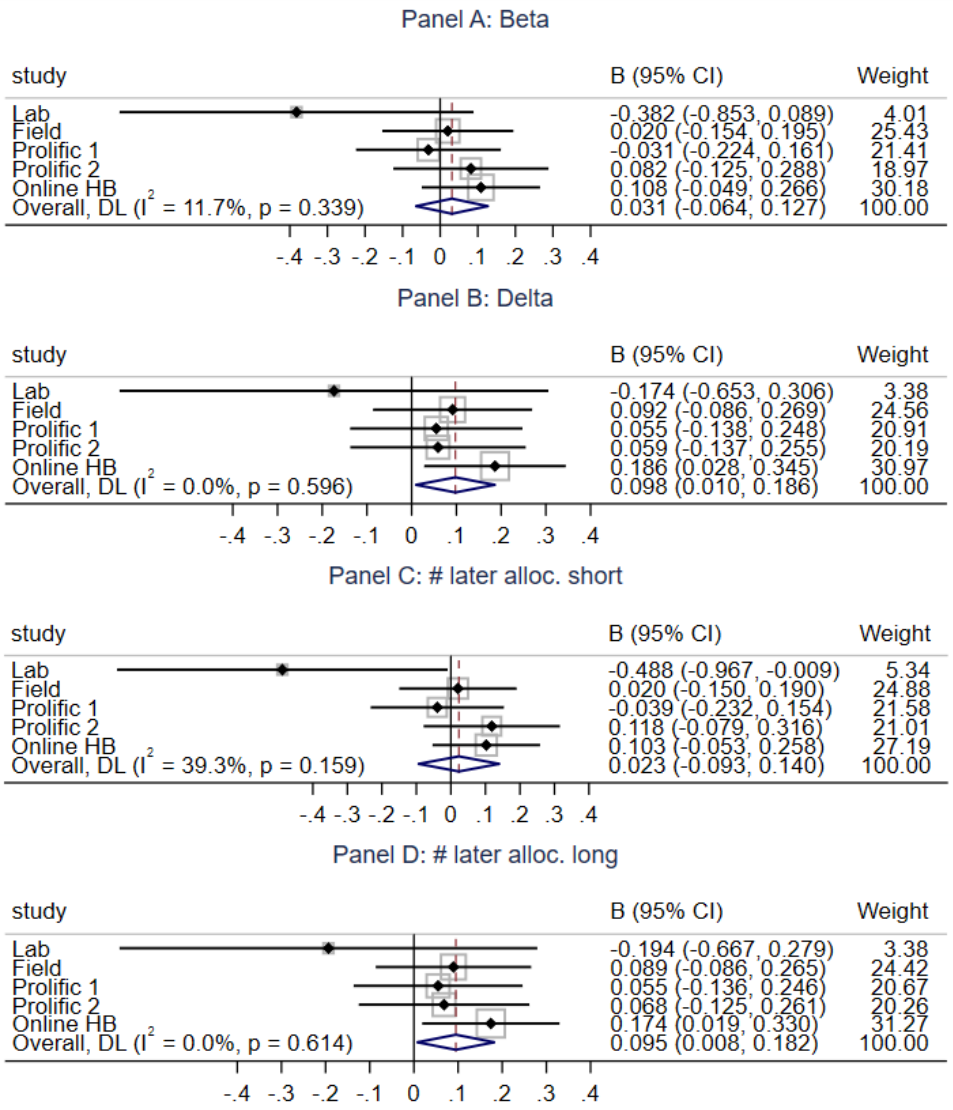
F. Comparing BRIS and hypothetical payments

Section F reports the meta-analytic comparison between treatments B and H for the sake of completeness (adding the aforementioned Study V). We find that treatment B is significantly associated with higher long-term patience than treatment H, but the result is nearly uniquely driven by (the newly added) Study V.

Figure F.1 displays the meta-analytic results for the comparison between BRIS and Hypothetical decisions (i.e., treatment B vs. treatment H) using the same protocol as for treatments H vs. R and B vs. R (sections 5.1 and 5.2 of the main text). The overall effect of B on beta is +0.031 SD (95% CI = [-0.064, 0.127], $p = 0.520$), while on delta the estimation yields a significant overall effect of +0.098 SD (95% CI = [0.010, 0.186], $p = 0.030$). Regarding the number of later allocations in the short-term block, for treatment B we find an overall effect of ±0.023 SD (95% CI = [-0.093, 0.140], $p = 0.696$); while for the long-term block we find a significant overall effect of +0.095 SD (95% CI = [0.008, 0.182], $p = 0.032$).

Therefore, the results suggest that B and H are similar in beta and the number of later allocations in the short-term block (the overall effects are very small and non-significant). However, we find significant differences between B and H in delta and the number of later allocations in the long-term block, which are still very small (less than 0.10 SD). It can be seen that only the estimates for Study V reach significance in these two variables, though. If we exclude Study V, the overall effects on the four outcome variables are not significant ($p>0.26$).

Figure F.1: **Meta-analytic results for the B vs. H treatments**



References

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *The American Economic Review*, *92*(5), 1606–1617.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644–1655.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and PersonalityScience*, *8*(4), 355–362.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin and Review*, *14*(5), 779–804.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. Florida: CRC press.