



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

Pablo Brañas-Garza^{1,2}  · Diego Jorrat^{1,2} · Antonio M. Espín^{1,3} · Angel Sánchez^{4,5}

Received: 19 October 2020 / Revised: 27 August 2022 / Accepted: 26 September 2022
© The Author(s), under exclusive licence to Economic Science Association 2022

Abstract

The use of real decision-making incentives remains under debate after decades of economic experiments. In time preferences experiments involving future payments, real incentives are particularly problematic due to between-options differences in transaction costs, among other issues. What if hypothetical payments provide accurate data which, moreover, avoid transaction cost problems? In this paper, we test whether the use of hypothetical or one-out-of-ten-participants probabilistic—versus real—payments affects the elicitation of short-term and long-term discounting in a standard multiple price list task. We analyze data from a lab experiment in Spain and well-powered field and online experiments in Nigeria and the UK, respectively ($N=2,038$). Our results indicate that *the preferences elicited using the three payment methods are mostly the same*: we can reject that either hypothetical or one-out-of-ten payments change any of the four preference measures considered by more than 0.18 SD with respect to real payments.

Keywords Time preferences · Hypothetical vs. real payoffs · Lab · Field · Online experiments · BRIS

JEL Classification C91 · C93

✉ Pablo Brañas-Garza
branasgarza@gmail.com

¹ LoyolaBehLab, Universidad Loyola Andalucía, Córdoba, Spain

² Fundación ETEA, Universidad Loyola Andalucía, Córdoba, Spain

³ Department of Social Anthropology, Universidad de Granada, Granada, Spain

⁴ GISC & IBiDat, Universidad Carlos III de Madrid, Leganés, Spain

⁵ BIFI, Universidad de Zaragoza, Zaragoza, Spain

1 Introduction

Patience is becoming a major topic in economics.¹ Patience refers to the preference for larger rewards in the future over smaller rewards at a sooner date. Time discounting (TD) is the loss of utility associated to any reward that is deferred in time. Patient individuals, therefore, display lower TD than impatient individuals.

Time preferences are relevant in several domains. In the field of health behavior, there is evidence suggesting that experimental measures of individuals' patience negatively correlate with alcohol consumption, smoking behavior, and BMI (Borghans & Golsteyn, 2006; Chabris et al., 2008; Sutter et al., 2013). In education, the evidence suggests that subjects with a high level of patience have better education outcomes (Castillo et al., 2019; Duckworth & Seligman, 2005; Golsteyn et al., 2014; Kirby et al., 2005; Non & Tempelaar, 2016; Paola & Gioia, 2014) and are less likely to receive disciplinary referrals in school (Castillo et al., 2011) or drop out of high school or college (Cadena & Keys, 2015). As regards the field of finance, patience is correlated with income, savings, and credit card borrowing (negatively) and is a good predictor of real-life wealth distribution (Ashraf et al., 2006; Epper et al., 2020; Giné et al., 2017; Meier & Sprenger, 2010, 2013; Tanaka et al., 2010). In other domains, patience has also been correlated with cognitive ability (Bosch-Domènech et al., 2014; Dohmen et al., 2010; Frederick, 2005), criminal behavior (Åkerlund et al., 2016), the probability of divorce (Paola & Gioia, 2017; Schaner, 2015), and social behavior (Dewitte & Cremer, 2001; Espín et al., 2012, 2015; Rachlin, 2002).

TD is typically elicited using multiple price lists (MPL; Coller & Williams, 1999) where subjects decide whether to take (the “sooner” option) or save (the “later” option) a certain amount of money over a series of independent choices with increasing interest rates.²

Although for experimental economists the use of monetary incentives has been practically a hallmark, some of the most acclaimed papers on TD (e.g., Ashraf et al., 2006; Cadena & Keys, 2015; Golsteyn et al., 2014; Kirby et al., 2005; Sunde et al., 2022) do not pay participants real monetary incentives, that is, decisions are hypothetical. But *are choices over hypothetical rewards as equally informative as incentivized time preferences?*

Besides the obvious monetary and logistics costs, the use of monetary incentives in TD is challenging. First, transaction costs and payment reliability need to be constant across options regardless of the payment date (Cohen et al., 2020). For example, future payments must be just as reliable as immediate payments. This problem is more prominent in the field, especially in low-income populations where most

¹ According to the Scopus database, the number of papers containing “time preference” in the title was about four times greater in 2018 than in 2000.

² Another experimental method which is increasingly used to elicit time preferences is the so-called Convex Time Budgets (CTB) task designed by Andreoni and Sprenger (2012a, 2012b). Some recent experimental research in the field (Giné et al., 2017; Lührmann et al., 2018; Meier & Sprenger, 2013) implements this method. In contrast to standard MPLs, this method also allows for interior allocations.

people are frequently unbanked or change their cell-phone numbers. All these factors increase the uncertainty associated with future payments and hence the probability that a subject will prefer the sooner option for reasons other than time preferences.

Second, subjects' choices only reflect their time preference if they are liquidity constrained and cannot engage in intertemporal arbitrage (Frederick et al., 2002; Lührmann et al., 2018). Otherwise, subjects should compare the interest rate offered in the task (r) to the market interest rate (\bar{r}). For $r < \bar{r}$, subjects can just take the money today and deposit it in the bank. For $r > \bar{r}$, they should save all in the task.

Third, a higher expectation of inflation may lead an individual to prefer sooner smaller rewards without the influence of time preference, simply because the money is worthless in the future (Frederick et al., 2002; Martín et al., 2019).³

Fourth, there is a severe problem with data privacy. If subjects need to be paid by bank transfer, then they must disclose private information. Relatedly, there is growing interest in the development, malleability, and stability of time preferences (Alan & Ertac, 2018; Kim et al., 2018; Lührmann et al., 2018; Perez-Arce, 2017). These studies employ adolescent (or children) samples for which the use of real money requires special parental consent and additional requirements from ethics committees.

Thus, if the use of real incentives in TD elicitation is expensive, may induce biased estimates, and is, in general, so problematic, why do we pay decisions at all? Previous evidence directly comparing both mechanisms (i.e., real vs. hypothetical) is typically based on lab experiments with small samples (6 to 60 subjects), and low statistical power. Using a within-subject design, Madden et al. (2003) compared hyperbolic discount rates estimated from students' choices over incentivized vs. hypothetical rewards and found no differences. Johnson and Bickel (2002), also using a within-subject design, found no significant differences between incentivized and hypothetical choices. Madden et al. (2004) replicated the (null) results using a between-subject design. Bickel et al. (2009) compared TD choices with real and hypothetical money using neuroimaging and found no significant behavioral or neurobiological differences. Similar results were reported in Lawyer et al. (2011) and Matusiewicz et al. (2013). In contrast, Coller and Williams (1999) found a weak significant difference between both mechanisms, but assignment of subjects to treatments was not random.

Evidence from field experiments is even scarcer. Harrison et al. (2002) examined in Denmark whether variations in the Between-Subjects Random Incentive System (BRIS) had an impact on TD. Keeping the number of paid subjects per session fixed, they found no significant effect of session size (i.e., of paying probability). Ubfal (2016) estimated discount rates for six different goods among rural Ugandan households and found no significant differences between hypothetical and incentivized choices. In both studies, however, participants were not randomly assigned to treatments. Besides, Ubfal (2016) does not control for order effects.

³ Macchia et al. (2018) found that experimental subjects from a high inflation country (Argentina) discounted future rewards more steeply than people from the UK.

Apart from these papers, Falk et al. (2016) validated a hypothetical TD task using a real incentives task. They found that both measures were highly correlated and predicted real-life behaviors similarly. In addition, Matousek et al. (2022) meta-analyzed 56 studies (lab and field experiments) and showed that discount rates do not differ between experiments using real and hypothetical rewards.

This paper aims to shed light on this knowledge gap and quantify the impact of hypothetical vs. real rewards on TD elicitation. We study this in a comprehensive manner across different populations and settings. We analyze data from a lab experiment in Spain, a lab-in-the-field experiment in rural Nigeria, and two online experiments with participants from the UK. In addition, we examine the impact of another commonly used payment method consisting of paying only a fraction of subjects (BRIS). In all cases, subjects were randomly assigned to one of the three different payment mechanisms and the data allow us to test for effects on both short- and long-term discounting decisions as well as on the individual parameters obtained assuming quasi-hyperbolic preferences (i.e., the beta and delta discount factors).

Our results contribute to the debate and suggest that choices over hypothetical rewards do not differ from choices over real rewards in TD tasks. In fact, we believe that the accumulated evidence now allows us to conclude that hypothetical rewards do not induce relevant estimation bias. Furthermore, while our results for BRIS payment schemes are slightly more ambiguous, they also suggest little if any difference with respect to real rewards. However, further studies should analyze BRIS in more detail, for example, considering different payment probabilities (we only test for 1/10 probability).

A notable feature of MPLs (and other experimental methods) is that they may yield biased discount rates as a consequence of the linear utility assumption. This assumption leads to upward-biased discount rate estimates if utility is concave (Andersen et al., 2008; Andreoni & Sprenger, 2012a). A solution to this issue is to use a double MPL method, to jointly estimate risk and time preferences.⁴ While our data do not allow us to accurately estimate time and risk preferences, in the Online Appendix (OA hereafter) we provide estimates of TD while controlling for risk preferences in some of our datasets.⁵ The main results are not qualitatively affected.

These results are more than a purely methodological contribution. They have important implications for the budgets of research groups running TD experiments, especially in the field in developing and volatile economies, as well as for the design of large-scale representative surveys which increasingly include behavioral tasks to estimate economic preferences in the population.

⁴ However, the most important critique to (double) MPL methods is that risk preferences are not time preferences, that is, different utility functions might apply for time and risk preferences (Andreoni & Sprenger, 2012b).

⁵ Because time preferences should not be elicited apart from risk preferences (Andersen et al., 2008), MPL TD tasks require eliciting risk preferences in a different task (e.g., Holt & Laury, 2002). This implies a significant increase in time response and costs. However, it also leads us to wonder if paying or not for the measurement of risk preferences produces the same findings. Brañas-Garza et al. (2021a) have shown that hypothetical and real rewards yield similar results.

The rest of the paper is structured as follows. The second section focuses on the main research questions and the third reports on general features of the design. Section four provides an overview of the four studies. In section five, we report the main results. In the last section, we discuss the results and the contribution of the paper.

2 Questions to be addressed

The main research question of this paper is:

Q1: Do hypothetical payments (H) provide the same outcome as real payments (R) in TD elicitation tasks?

Recently, several studies have used an intermediate solution as a cost-saving alternative to real payments: paying one subject out of X rather than all of them (i.e., BRIS). Hence, only a fraction of individuals is randomly selected to receive the real payments, and participants are aware of this probability ex-ante (Baltussen et al., 2012). The associated monetary and logistics costs decrease proportionally, but the rest of the problems (lack of trust, inflation, privacy, and financial arbitrage) remain untouched.

As a second goal of this research, we will compare this probabilistic mechanism with real payments. We rephrase Q1 as follows to account for BRIS:

Q2: Do one-out-of-ten payments (B) provide the same outcome as real payments (R) in TD elicitation tasks?

To address Q1 and Q2, we conduct four experiments. We run a lab experiment with university students in Seville (Spain), a lab-in-the-field experiment in the Kano province of Nigeria, and two online experiments in Prolific Academic (in which the time horizons in the long-term block of the TD task differ; see below). These will be referred to as studies I to IV, respectively.

3 Treatments, balance, and MPL task

In the four studies, we follow the same protocol: participants are randomly assigned to one of the treatment arms (real, BRIS, or hypothetical: R/B/H). The randomization allows us to evaluate the *causal impact* of different payment schemes over the estimated TD.

3.1 Treatments

We compare 3 treatments that differ in the probability of being paid:

- **R**: Earnings with probability $P = 1$, where all subjects get a real payment.
- **B**: Earnings with probability $P = 1/10$, where 1 subject out of 10 gets a real payment.

- **H:** Earnings with probability $P=0$, where none of the subjects get a real payment.

All the subjects were informed of their payment scheme ex-ante, but were unaware of the existence of other payment schemes (i.e., treatments).

3.2 Sample and balance across experiments

Section A (OA) provides details about the sample (Lab: $n=119$, Field: $n=721$; Prolific 1: $n=606$; Prolific 2: $n=592$). This section also shows in detail the results of the randomization and balance between treatments (see Table A.1).

3.3 Eliciting time preferences

Our instrument to measure time preferences was an MPL adapted from Collier and Williams (1999) and Espín et al. (2012). Similar tasks have been used for instance in Burks et al. (2012), Espín et al., (2015, 2019), and Martín et al. (2019). Participants made a total of 20 binary choices between a sooner smaller amount of money and a later but larger amount in two blocks of ten decisions each. The first block (short-term block) involves choosing between a no-delay option (“today”) and a one-month delay option in all the studies. The second block (long-term block) is the same in the first three studies but different in Study IV: it involves choosing between a one-month delay option and a seven-month delay option in studies I-III, while it involves choosing between a one-month delay option and a two-month delay option in Study IV. Study IV aimed to analyze whether using the same one-month delay for both the short-term and long-term blocks, rather than different delays (i.e., one month and six months) as in the first three studies, makes any difference. We used the same amounts in both blocks, whereas interest rates vary according to the time horizon considered in each block. The amount of the sooner payoff was fixed across decisions whereas the amount of the later payoff increased in interest rates from decision 1 to decision 10 (see Table 1).

The protocol described above allows us to compute the beta and delta parameters (β , δ) of a quasi-hyperbolic discount function (Burks et al., 2012; Laibson, 1997; McClure et al., 2004; Phelps & Pollak, 1968). The beta-delta model formalizes the individual’s discount function as $V_d = \beta\delta^t V_u$, where V_d is the discounted psychological value of a reward with (undiscounted) value V_u which will be received in t time units. β and δ are the “beta” and “delta” discount factors, respectively. Theoretically, β and $\delta \in (0, 1]$. The *higher these discount factors, the more patient* the individual is since delayed rewards are valued more (i.e., they are discounted less). The beta discount factor refers to present bias, that is, the value of any non-immediate reward is discounted by a fixed proportion β , regardless of the delay. The delta discount factor captures “long-term discounting” in an exponential functional form, that is, for each unit of time that constitutes the delay to delivery, the value of a reward is discounted by δ . This model thus allows for a possible difference between short- and long-term

Table 1 MPL design across experiments

Study I (euros)		Study II (nairas)		Studies III-IV (pounds)		Monthly interest rate		
Sooner	Later	Sooner	Later	Sooner	Later	Short (%)	Long	Long (Study IV) (%)
10	10.00	400	400	3	3.00	0.00	0.00	0.00
10	10.70	400	427	3	3.20	6.70	1.12	6.70
10	11.30	400	453	3	3.40	13.40	2.23	13.40
10	12.00	400	480	3	3.60	20.10	3.35	20.10
10	12.70	400	507	3	3.80	26.70	4.45	26.70
10	13.30	400	533	3	4.00	33.30	5.55	33.30
10	14.00	400	560	3	4.20	40.00	6.67	40.00
10	14.70	400	587	3	4.40	46.70	7.78	46.70
10	15.30	400	613	3	4.60	53.40	8.90	53.40
10	16.00	400	640	3	4.80	60.00	10.00	60.00

Monthly simple interest rates are displayed. In all studies except Study IV, the interest rates differ between the short-term and long-term blocks because the delays considered are one month and six months, respectively. In Study IV, the delay in both the short-term and long-term blocks is one month

discounting and has been shown to predict outcomes better than other formulations (Burks et al., 2012).

We opted for the non-delayed option (“today”) because we wanted to test whether there are differences in beta (i.e., present bias) between treatment H and treatment R. Present bias refers to the apparent tendency of (some) individuals to assign a premium to immediate rewards (McClure et al., 2004; Takeuchi, 2011). It is reasonable to expect that the “today” option will induce stronger differences between hypothetical and real rewards because the immediacy premium might partly capture differences in uncertainty or transaction costs between immediate and non-immediate rewards (Chabris et al., 2008), which are absent in hypothetical scenarios. There is evidence that delaying the sooner option by one day helps to avoid possible confounds such as differential transaction costs between payment dates or trust issues (Sozou, 1998). Note that having payments today may make it more likely to capture present bias than having a front-end delay, assuming that the relevant threshold for immediacy is one day.

However, without a truly immediate option, the beta parameter cannot be accurately estimated because any non-immediate reward is automatically discounted by beta. Technically speaking, for choices between non-immediate rewards, the beta parameter in $V_d = \beta \delta^t V_u$ cancels out between the two sides of the equation. In our design with a “today” option, we therefore expected to find the strongest differences between the H and R treatments in present bias β , or short-term discounting.

In each block, we obtained the switching point and interpreted that, at that point, a participant was indifferent between both options. Following the protocol introduced by Burks et al. (2012), we computed the β and δ for each participant. The time units were defined in months. As is standard in the literature, we assume that utility

Table 2 Discount factors and number of later allocations by study

Variable	Mean	SD	Min	Max	Incons (%)
<i>Study I: Lab</i>					
Beta	0.827	0.097	0.652	1.049	–
Delta	0.937	0.020	0.918	0.990	–
# later alloc. (short)	5.233	2.503	0	10	3.33
# later alloc. (long)	2.617	2.577	0	9	0.00
<i>Study II: Field</i>					
Beta	0.744	0.142	0.612	1.089	–
Delta	0.931	0.027	0.918	1.000	–
# later alloc. (short)	2.617	3.898	0	10	0.55
# later alloc. (long)	1.648	3.390	0	10	0.28
<i>Study III: Prolific 1</i>					
Beta	0.801	0.124	0.606	1.089	–
Delta	0.939	0.026	0.918	1.000	–
# later alloc. (short)	4.516	3.505	0	10	1.98
# later alloc. (long)	2.853	3.374	0	10	1.16
<i>Study IV: Prolific 2</i>					
Beta	1.004	0.128	0.600	1.667	–
Delta	0.786	0.127	0.600	1.000	–
# later alloc. (short)	5.313	3.320	0	10	1.51
# later alloc. (long)	5.379	3.310	0	10	1.34

is linear over the relevant range of payoffs given that they are rather small. Note, however, that previous research suggests that risk aversion (i.e., the concavity of the utility function) should be accounted for to correctly estimate discount factors (e.g., Andersen et al., 2008). Since our goal is not to estimate aggregate discount factors but to compare the time preferences elicited using different payment methods, we assume that any effect of risk preferences will be compensated between treatments. Moreover, in the samples in which risk preferences were elicited we also control for them.

Table 2 provides the descriptive statistics for the short-term (β) and long-term (δ) discount factors obtained in each study. Note that, following standard methodology, we do not restrict β to be smaller than one, so that some individuals might display “future bias” (Balakrishnan et al., 2020; Jackson & Yariv, 2014; Takeuchi, 2011).

The proportion of inconsistent choices (multiple switching or non-monotonic patterns) differs slightly across studies and blocks of the task, as can be seen in Table 2.

In Study I, 3.33% of participants made inconsistent choices in the short-term block, whereas none did so in the long-term block. In Study II, only 0.6% of the participants made inconsistent choices in the short-term block and 0.28% in the long-term block. It should be noted that the field data of Study II was gathered using enumerators who were trained to avoid inconsistencies. In Study III, 1.98% of the subjects made inconsistent choices in the short-term block and 1.16% in the long-term block. Finally, in Study IV the figures were 1.51% and 1.34%, respectively. For the sake of completeness, in the empirical analysis we also consider the number of

later allocations in the short- and long-term blocks as an alternative measure of individuals' patience. Unlike beta and delta, these measures are not parameterized and consider individuals who make both consistent and inconsistent decisions (β and δ cannot be computed for inconsistent individuals).

Figure 1 shows the distribution of the number of later allocations by study. In Study II (Nigeria), 65% and 80% of the subjects chose the sooner option in all 10 decisions (i.e., number of later allocations = 0) of the short-term and long-term block, respectively. In the literature it is common to find such a high percentage of people always choosing the sooner reward, particularly in field experiments. For example, Martín et al. (2019) found that 48% of Spanish Gitanos chose the sooner option. Of course, increasing the interest rate associated with the delayed reward would have probably reduced the number of sooner allocations in this study. Yet this would have been counterproductive in the experiments conducted with participants from Spain and the UK (i.e., studies I, III, and IV) where the data are quite well distributed.

Our procedure provides comparable decisions, and therefore comparable estimates, across samples. Note that the GPS developed by Falk et al. (2018) ranks Nigeria in 49th position (out of 76 countries) in terms of patience, whereas Spain and the UK rank 18th and 11th, respectively. Regardless of the time-horizon considered, participants in Nigeria appear to be more impatient than participants in Spain and the UK.

Yet, the goal of this study is not to compare patience between countries. As mentioned, an alternative method would be to adjust the interest rates to minimize censoring in each study (especially in the field experiment). However, if distinct tasks are set in the different experiments, it would be impossible to know if the differences between studies in terms of the effect of incentives were due to differences between the tasks rather than differences between experimental settings (i.e., lab, field, online).

4 Implementation and sample

4.1 Study I: The lab experiment

In principle, the lab provides the most controlled setting to test whether different reward schemes affect TD measures. In the lab, experimenters have a higher degree of control over the environment and can ensure greater credibility for future payments. The lab typically also has some drawbacks though: participants are university students, self-selected into the experiment, and with a relatively high socioeconomic status.

We ran the lab experiment at the University of Seville and the Pablo de Olavide University, Spain, from April to May 2019. Participants were recruited in the two campuses using flyers and the School of Economics website. Among the 473 subjects who signed up, 120 were randomly assigned to the study and then randomly assigned to treatments R, H, B.

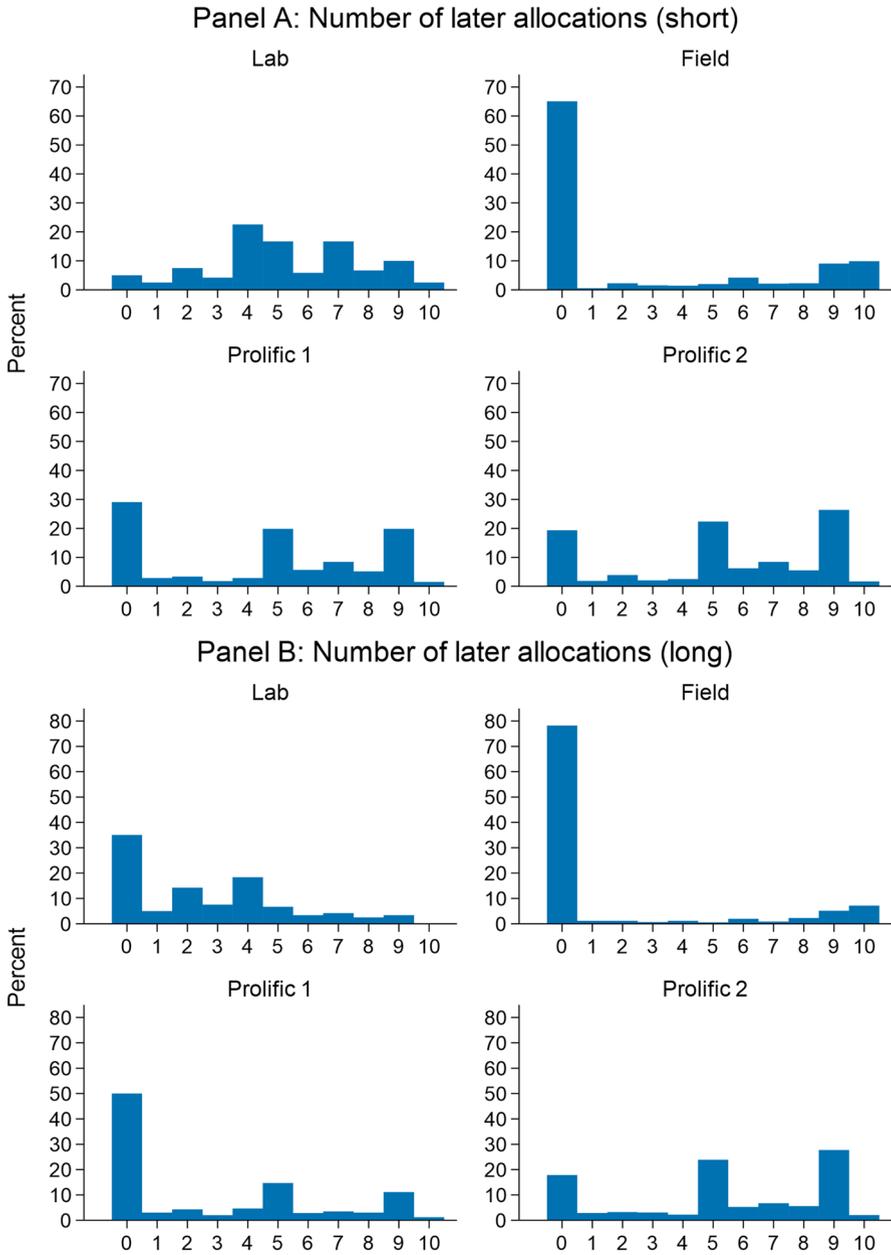


Fig. 1 Short- and long-term later allocations by study

The sample was composed of students from undergraduate degree programs in business (31%), law and economics (24%), marketing (20%), economics (16%), and others. The average age of the participants was 22 and 39% were female. See Section A (OA) for details.

For the final payments, as is standard in the literature, one out of the 20 decisions (ranging 10 to 16 euros)⁶ was randomly selected for payments (see Charness et al., 2016, for a discussion on the validity of this payment method). In the R condition, all participants were paid the amount associated to their choice in that decision on the corresponding date (either “today,” or in one month, or in seven months), whereas in treatment B we randomly selected 10% of the participants to receive the money. No participant was selected for payment in the H treatment.

We offered participants the possibility of bank transfers, but only about 40% selected this option. The remaining 60% was paid in cash at the university on the day corresponding to the randomly selected decision. All participants received a show-up fee of 4 euros, were informed about the content of the experiment prior to participating and signed a consent form. The study was approved by the Ethics Committee of Loyola University.

4.2 Study II: The field experiment

In Study II, we explore the effect of hypothetical and BRIS methods in the field, with a more heterogeneous sample compared to the lab. In addition, Study II provides a much larger sample and therefore a more powerful analysis.

We ran a lab-in-the-field experiment in Kano province, Northern Nigeria. The experiment was conducted in seven villages in the province (Albasu, Daho, Dorayi, Gidan Maharba, Farantama, Jáen, and Panda) from November 2018 to April 2019. A total of 721 households were randomly selected to obtain a representative sample of the study area with the eligibility criterion of having at least one child between the age of 6 and 9 years old.⁷ Each household in the total sample was randomly assigned to one of the three treatments (R/B/H).

As is standard in the field, the experiment was conducted by enumerators, which implies that the instructions were read—and often explained—by the enumerator. Sixty-two enumerators were hired and trained for the fieldwork. Each one received a list of households to visit and a tablet to conduct the interviews. The random allocation of households to treatments was computerized and the enumerators did not have any influence on the selection. Enumerators conducted face-to-face interviews in the households and only one person was interviewed per household.

The resulting sample size was $n = 721$ (by treatments, R: 239, B: 246, H: 236). Subjects were fully aware of their payment scheme. See Section A (OA) for details.

⁶ The current average wage per hour in Spain was about 15 euros. Given the average duration of the experiment (30 min), subjects were paid at least 1.5 times the average wage.

⁷ We followed this criterion because the experiment was part of a much larger intervention conducted by the DIME (The World Bank). In addition, the random selection of households followed a geographical criterion based on their distance from the catchment areas of the local schools.

The experiment consisted of four tasks: coordination game, expectations, time discounting, and risk preferences. The TD task was always performed in third place. The payment scheme was held constant across the entire experiment: participants in the R treatment performed all four tasks with real money, whereas participants in the H (B) treatment performed all four tasks with hypothetical (BRIS) money.

To elicit time preferences, we used the same MPL (same interest rates) as in the lab experiment. Table 1 shows the payments. We re-calculated the payments to be able to pay about one-day average wages for the entire experiment (equal to 1080 nairas or 3 USD). This resulted in a minimum payment of 400 nairas in the TD task. We randomly selected one of the 20 MPL choices for payment. For all participants in the R treatment and for the randomly selected 10% in the B treatment, we made the payments by charging their cell phones with the chosen amount on the date of the selected decision.

The study was approved by the Ethics Committees of Middlesex University London and IRB Solutions (US). All participants signed a consent form.

4.3 Study III: Online experiment 1 (Prolific 1)

An increasing number of papers run experiments online using platforms such as Amazon Mechanical Turk, Behave4 Diagnosis, and Prolific Academic, among others. In general, these experiments involve a more heterogeneous sample pool than traditional laboratory subjects and there is less control over what subjects are doing when completing the experiment. While recent evidence suggests that online data are reliable (Horton et al., 2011; Rand, 2012; Arechar et al., 2018; Prissé & Jorrat, 2022), the payment mechanism could still influence the behavior of these experimental subjects and hence the elicitation of time preferences.

We therefore ran an online experiment using Prolific Academic. Prolific is a crowdsourcing platform for behavioral research. Regarding transparency, in Prolific subjects know that they have been recruited to participate in an experiment and are aware of the expected payments. Researchers can also screen the subject pool in a range of dimensions before inviting subjects (for more details, see Palan & Schitter, 2018).

The experiment was published on Prolific on July 15th and lasted for four hours. Subjects were randomly assigned to treatments R, B, or H with probability 1/3. We restricted the sample to subjects living in the UK because it was 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 socio-economic questions to avoid losing observations with respect to the control variables. The experiment consisted of time, risk preferences, and dictator game tasks (always in this order). As in Study II, the payment scheme was held constant across the entire experiment, and we used the exact same MPL with adjusted payments (see Table 1).

The resulting sample size was $n = 606$ (by treatments, R: 187, B: 204, H: 215). Subjects were fully aware of their payment scheme. See Section A (OA) for details.

As in studies I and II, we randomly selected one out of the 20 MPL decisions to compute the final payoffs. However, as Prolific forces researchers to pay a completion fee, all the participants received a fixed payment of £1.2. In addition, all the participants in treatment R and 1 out of 10 in B received a bonus payment corresponding to the selected decision. These procedures were clearly explained in the instructions and the participants signed a consent form.

4.4 Study IV: Online experiment 2 (Prolific 2)

For Study IV, we conducted a parallel replication of Study III in Prolific but changed the long-term block: the delay considered in Study IV was one month rather than six months (see Table 1). That is, subjects in this study had to choose between a one-month delay option and a two-month delay option in the long-term block. This allowed us to test the sensitivity of the payment scheme to a shorter delay in long-term decisions, which may have an effect (Cohen et al., 2020).

The selection of subjects, randomization, implementation, and payment procedures were the same as in Study III. The resulting sample size was $n = 592$ (R: 193, B: 203, H: 196). See Section A (OA) for details.

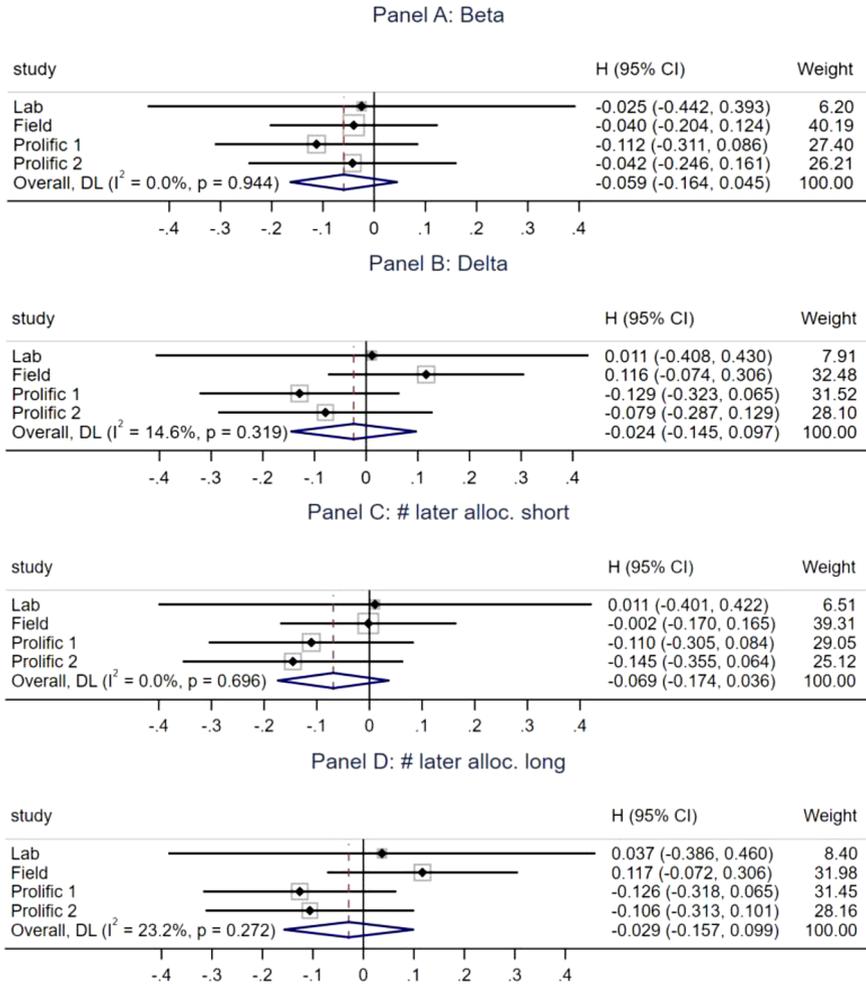
5 Results

In this section, we analyze the four studies combined to answer Q1 and Q2. More specifically, we first standardize the dependent variables (beta, delta, and #later allocations) to estimate the effect of H (Q1) and B (Q2) in terms of the SD of each variable in R. Then, we provide an interpretable and comparable measure for each effect size and an overall effect for all the four studies combined based on random-effects meta-analysis. An analysis of the unstandardized effects sizes is provided in Section B (OA). In addition, one particular concern regarding the use of hypothetical decisions is that they might increase variance, that is, noisy decision making (for example, due to lack of attention), but not necessarily change the average response (Camerer & Hogarth, 1999). Thus, we analyze whether response variance differs across treatments.⁸ In the OA we run a number of additional analyses for robustness.

5.1 Are hypothetical and real choices different (Q1)?

Figure 2 shows the results of the random-effects meta-analysis for the effect of treatment H on the four dependent variables considered, as expressed in terms of the SD of the variable in treatment R in the same study (similar to a Cohen's d value). In all cases, we provide an estimate (with 95% CI) for the effect in

⁸ Since the number of individuals making inconsistent choices, which is an alternative way of looking at noise, is very low in our experiments (see Table 2), and we even have enumerators who were trained to avoid inconsistencies in Study II, we relegate the analysis of potential between-treatment differences in inconsistencies to Section B (OA), where we examine each study in detail.



Note: Random-effects weights calculated using the DerSimonian-Laird method. I^2 between-studies heterogeneity statistics and p-values are reported.

Fig. 2 Meta-analytic results for the H vs. R treatments

each study and for the overall effect using the DerSimonian and Laird method for weighting the samples. The resulting weights are shown in the last column. In addition, we provide the I^2 statistic of heterogeneity between studies and its P -value. If heterogeneity does not yield significance (as is the case in all the analyses), we can say that the effects in the different studies come from the same distribution and that the variation can be attributed to chance. The results are controlled for gender, age, education level, and enumerator in the case of Study II.

Table 3 Variance ratio and KS tests for the H vs. R treatments

	(1) Beta	(2) Delta	(3) #later alloc. (short)	(4) #later alloc. (long)
(i) <i>Standard deviation by treatment</i>				
SD(R)	0.171	0.096	3.747	3.596
SD(H)	0.165	0.101	3.677	3.646
(ii) <i>R vs. H</i>				
SD(R)/SD(H)	1.041	0.948	1.019	0.986
<i>P</i> (ratio < 1)	0.849	0.082*	0.688	0.358
Kolmogorov–Smirnov (<i>P</i> -val)	0.847	0.229	0.996	0.998

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 group is smaller than 1. * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$

The overall effect of H on beta is -0.059 SD (95% CI = $[-0.164, 0.045]$), which is not significant ($P = 0.263$), and it is not significant either in any of the studies. This also applies to delta, on which the overall effect of H is -0.024 SD (95% CI = $[-0.145, 0.097]$, $P = 0.694$). In terms of the (monthly) discount factors beta and delta, based on SDs of 0.172 and 0.096 as estimated in treatment R, these effects are -0.010 (95% CI = $[-0.028, 0.008]$) and -0.002 (95% CI = $[-0.014, 0.009]$), respectively. For the number of later allocations in the short-term and long-term blocks, the overall effects are -0.069 SD (95% CI = $[-0.174, 0.036]$, $P = 0.200$) and -0.029 SD (95% CI = $[-0.157, 0.099]$, $P = 0.656$), respectively.

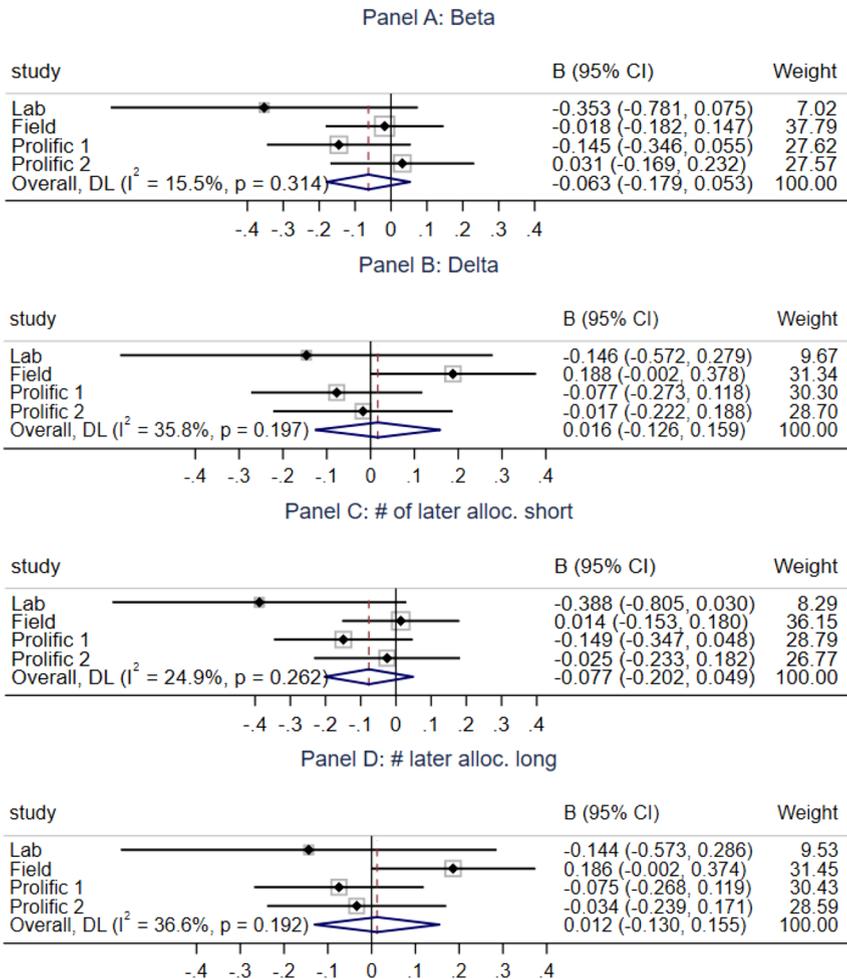
All these effects are very small in size. In sum, for the four dependent variables, the comparison between treatments H and R is not significant either overall or in any individual study. Heterogeneity statistics suggest that the variation between studies is due to chance in all variables (all $P > 0.27$). More specifically, only between 0% (beta and short-term later allocations) and 23.2% (long-term later allocations) of the observed variability occurs across studies.

To analyze differences in variance, Table 3 shows the results of a series of variance ratio tests for all the four studies combined. Since the hypothesis is that real incentives trigger less noisy decisions, we conduct one-tailed tests against this hypothesis, that is, against $SD(R)/SD(H) < 1$. The results indicate that the hypothesis of different variances in treatments H and R is not supported for any of the TD variables considered, except for a marginally significant difference in delta, for which H shows a greater variance than R ($P = 0.082$; all remaining $P > 0.35$).

According to the above results, neither the mean estimates nor the variances differ between treatments H and R. Still, the distributions may vary as a result of more complex factors. To test for potential differences in the observed distributions, we conduct Kolmogorov–Smirnov tests and report the P -values in Table 3. The results suggest that the distributions are not different for any of the variables (all $P > 0.22$). Figures B.1 to B.4 (OA) show the distribution of choices under H and R in the two blocks in studies I to IV, respectively.

5.2 Are BRIS and real choices different (Q2)?

Figure 3 displays the meta-analytic results for the comparison between BRIS and real payment methods (i.e., treatment B vs. treatment R) using the same protocol as for treatments H vs. R. The overall effect of B on beta is -0.063 SD (95% CI = $[-0.179, 0.053]$, $P = 0.288$), while on delta the estimation yields $+0.016$ SD (95% CI = $[-0.126, 0.159]$, $P = 0.821$). In terms of the (monthly) discount factors beta and delta, based on aforementioned estimated SDs of 0.172 and 0.096, these effects are -0.011 (95% CI = $[-0.031, 0.009]$) and $+0.001$ (95% CI = $[-0.012, 0.015]$), respectively. Regarding the number of later allocations, for treatment B we find an overall effect of -0.077 SD (95% CI = $[-0.202, 0.049]$, $P = 0.230$) and $+0.012$



Note: Random-effects weights calculated using the DerSimonian-Laird method. I^2 between-studies heterogeneity statistics and p-values are reported.

Fig. 3 Meta-analytic results for the B vs. R treatments

Table 4 Variance ratio and KS tests for the B vs. R treatments

	(1) Beta	(2) Delta	(3) #later alloc. (short)	(4) #later alloc. (long)
(i) <i>Standard deviation by treatment</i>				
SD(R)	0.171	0.096	3.747	3.596
SD(B)	0.172	0.100	3.768	3.696
(ii) <i>R vs. H</i>				
SD(R)/SD(B)	0.999	0.963	0.994	0.973
<i>P</i> (ratio < 1)	0.486	0.165	0.443	0.238
<i>Kolmogorov–Smirnov</i> (<i>P-val</i>)	0.918	0.315	0.998	0.868

The null hypothesis is that the ratio between the standard deviation of the variable in the R group and the standard deviation in the B group is smaller than 1. **P* < 0.1, ***P* < 0.05, ****P* < 0.01

SD (95% CI = [− 0.130, 0.155], *P* = 0.864) for the short- and long-term blocks, respectively.

Therefore, all the overall effects are very small and non-significant, thus suggesting that there are no relevant differences between treatments. Although we find marginally significant effects of about +0.19 SD in Study II for both delta and the number of allocations in the long-term block (*P* = 0.05), the heterogeneity statistics suggest that all between-study variation can be attributed to chance in all cases (all *P* > 0.19). Yet, compared to H vs. R, B vs. R yields larger heterogeneity in all cases (between 15.5% and 36.6% of the variability occurs across studies).

As regards differences in variation (i.e., noise), using the same protocol as above, Table 4 shows that for all the variables considered the aggregate SDs do not differ between treatments B and R (*P* > 0.16). Again, the above results indicate that neither the mean estimates nor the variances differ between these two treatments. The Kolmogorov–Smirnov test comparing both distributions does not yield significance for any of the variables either (all *P* > 0.31; see Table 4). Figures B.1 to B.4 (OA) show the distribution of choices under treatments B and R in the two blocks in studies I to IV, respectively.

5.3 Additional analyses

To check the robustness of our results, the OA extends the analysis in five ways. Section B revises results study by study and shows that all the aggregate results reported above are not dramatically different across studies. 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 typically 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).

Section C focuses on ranges (in terms of Cohen’s *d* or SD) and *equivalence tests*. Applying this method, we can conclude 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. In other words, we can reject with 95% confidence that hypothetical or BRIS payments change beta by more than 0.16 SD with respect to real payments (i.e., the beta discount factor is not altered by more than 0.028). For delta, the interval that deems H and B equivalent to R is 0.14 SD (i.e., the delta discount factor is not altered by more than 0.013). 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 *equivalent* TD values as treatment R.

Section D analyzes alternative regression specifications (interval and binomial regressions) and shows that the results of the individual studies are robust to alternative regression specifications.

Section E, by means of a previously conducted online experiment (Study V), studies the potential effects of several design features typical of large-scale surveys/experiments on hypothetical TD. The main results are: (i) having played other games before the TD hypothetical task leads to increased patience, especially for short-term discounting; and (ii) the order of the blocks in the hypothetical TD task do not seem to influence choices.

Finally, 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 uniquely driven by Study V.

6 General discussion

This paper systematically studies the impact of different incentive schemes in the elicitation of time preferences using MPLs. We cover lab, field, and online experiments with very different subject pools in four experiments.

Our results show that for mean estimates, variances, and observed distributions, hypothetical incentives result in very similar time preferences as real incentives. This also applies to the BRIS payment method, in which only 10% of the participants are randomly selected to get the real payment. However, in some studies and for some variables, the BRIS method yields weak differences with respect to real rewards. In this sense, the BRIS results are slightly more ambiguous. Further studies should analyze more systematically different probabilities of getting paid.

These findings are in line with the lab (Johnson & Bickel, 2002; Lawyer et al., 2011; Madden et al., 2003, 2004; Matusiewicz et al., 2013), and field literature (Harrison et al., 2002; Ubfal, 2016), as well as with a recent meta-analysis (Matousek et al., 2022).

Therefore, our analysis suggests that existing hypothetical MPL tasks, often used to gather TD data in the field, are indeed informative of individual time preferences. At least, the resulting preference measures are essentially not different from those elicited with real incentives. Our findings thus indicate that hypothetical time

preferences are a valid proxy for incentivized ones in “money earlier or later” (MEL) experiments. Yet, it must be said that whether choices over money are equivalent to choices over time-dated utility flows (i.e., what time preference models actually refer to) is debatable (Cohen et al., 2020).⁹

Apart from the obvious budget implications, eliminating real payoffs also reduces other problems: children can participate more easily in experiments, subjects do not need to disclose private information, and problems such as differing transaction costs and inflation rates are alleviated, among other advantages. This is good news for field studies that may include this sort of TD tasks with a minimal impact on their budgets.

There are, at least, two important limitations. First, our study focuses on MPLs. While there is an intense debate over whether CTBs (Convex Time Budgets) work better than MPLs, the latter have been used more extensively. A precise analysis of the impact of hypothetical decisions on CTBs is necessary. An exception is the recent work by Prissé (2022) who compared hypothetical, real, and BRIS payments using continuous MPL (a procedure in between MPL and CTB) and found that hypothetical decisions yield similar results as incentivized ones. This suggests that the current findings can be extended to other elicitation tasks. Second, our results are necessarily silent about the effect of different payment schemes in other tasks, particularly those tasks involving strategic interaction. Recent evidence on risk preference elicitation (Brañas-Garza et al., 2021a) is aligned with the current finding of no difference between hypothetical and real rewards (and weak effects for the BRIS method). Moreover, measures of social preferences appear to be affected by the use of hypothetical rewards and stake sizes. In particular, people tend to be less prosocial when incentives are real or when stakes are high (Brañas-Garza et al., 2021b; Clot et al., 2018; Larney et al., 2019).

This leads us to speculate that the validity of hypothetical rewards might be restricted to tasks in which there are no obvious socially desirable responses or demand effects (Zizzo, 2010) that may act as intrinsic motivators for behavior (and which need to be overridden by real incentives) as is the case in time and risk preferences tasks, but not in social preferences tasks. Future research should explore this hypothesis in greater depth.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10683-022-09776-5>.

Acknowledgements We thank Giuseppe Attanasi, Michele Belot, Klarita Gërghan, Nagore Iriberry, Rosemarie Nagel, Pedro Rey-Biel, Amparo Urbano, and the audience at EUI-Florence, ESA2019, ESA2020, ECREEW-Seville, and SAE-Alicante for their comments and suggestions. We also thank A. Amorós, J. F. Ferrero, E. Mesa, and A. Núñez for conducting the lab experiment and Laura Costica and Edwin Daniel (Hanovia Ltd) for data collection in Nigeria. Nigeria data were gathered during the pilots of a World Bank project (PIs: Ericka Rascon and Victor Orozco). Funding provided by the Spanish Ministry of Science, Innovation and Universities (PGC2018-093506-B-I00, PGC2018-098186-B-I00).

⁹ It is unclear whether income or consumption is the appropriate argument for the utility function. Considering these criticisms, it remains to clarify what the MEL methodology is actually measuring. It is important to note, however, that most of the empirical literature relies on MEL tasks (Cohen et al., 2020).

and PID2021-126892NB-I00), Junta de Andalucía (PY18-FR-0007), PRACTICO-CM (Comunidad de Madrid), CAVTIONS-CM-UC3M (Comunidad de Madrid/Universidad Carlos III de Madrid), and Athena3i - Marie Skłodowska-Curie grant 754446 (European Union's Horizon 2020/Universidad de Granada). The replication material for the study is available at <https://doi.org/10.5061/dryad.vt4b8gtvn>.

References

- Åkerlund, D., Golsteyn, B. H. H., Grönqvist, H., & Lindahl, L. (2016). Time discounting and criminal behavior. *Proceedings of the National Academy of Sciences*, *113*(22), 6160–6165.
- Alan, S., & Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, *126*(5), 1865–1911.
- Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583–618.
- Andreoni, J., & Sprenger, C. (2012a). Estimating time preferences from convex budgets. *The American Economic Review*, *102*(7), 3333–3356.
- Andreoni, J., & Sprenger, C. (2012b). Risk preferences are not time preferences. *The American Economic Review*, *102*(7), 3357–3376.
- Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, *21*(1), 99–131.
- Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, *121*(2), 635–672.
- Balakrishnan, U., Haushofer, J., & Jakiela, P. (2020). How soon is now? Evidence of present bias from convex time budget experiments. *Experimental Economics*, *23*(2), 294–321.
- Baltussen, G., Post, G. T., Van Den Assem, M. J., & Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. *Experimental Economics*, *15*(3), 418–443.
- Bickel, W. K., Pitcock, J. A., Yi, R., & Angtuaco, E. J. (2009). Congruence of BOLD response across intertemporal choice conditions: Fictive and real money gains and losses. *Journal of Neuroscience*, *29*(27), 8839–8846.
- Borghans, L., & Golsteyn, B. H. (2006). Time discounting and the body mass index. *Economics and Human Biology*, *4*(1), 39–61.
- Bosch-Domènech, A., Branas-Garza, P., & Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D: 4D) predict cognitive reflection? *Psychoneuroendocrinology*, *43*, 1–10.
- Brañas-Garza, P., Estepa, L., Jorrat, D. A., Orozco, V., & Rascón Ramírez, E. (2021a). To pay or not to pay: Measuring risk preferences in lab and field. *Judgment and Decision Making*, *16*(5), 1290–1313.
- Brañas-Garza, P., Jorrat, D., Kovářik, J., & López, M. C. (2021b). Hyper-altruistic behavior vanishes with high stakes. *PLoS ONE*, *16*(8), e0255668.
- Burks, S., Carpenter, J., Götte, L., & Rustichini, A. (2012). Which measures of time preference best predict outcomes: Evidence from a large-scale field experiment. *Journal of Economic Behavior and Organization*, *84*(1), 308–320.
- Cadena, B. C., & Keys, B. J. (2015). Human capital and the lifetime costs of impatience. *American Economic Journal Economic Policy*, *7*(3), 126–153.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1/3), 7–42.
- Castillo, M., Ferraro, P. J., Jordan, J. L., & Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, *95*(11–12), 1377–1385.
- Castillo, M., Jordan, J. L., & Petrie, R. (2019). Discount rates of children and high school graduation. *The Economic Journal*, *129*(619), 1153–1181.
- Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, *37*(2–3), 237–269.
- Charness, G., Gneezy, U., & Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, *131*, 141–150.
- Clot, S., Grolleau, G., & Ibanez, L. (2018). Shall we pay all? An experimental test of random incentivized systems. *Journal of Behavioral and Experimental Economics*, *73*, 93–98.

- Cohen, J., Ericson, K. M., Laibson, D., & White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, 58(2), 299–347.
- Coller, M., & Williams, M. B. (1999). *Experimental Economics*, 2(2), 107–127.
- Dewitte, S., & Cremer, D. D. (2001). Self-control and cooperation: Different concepts, similar decisions? A question of the right perspective. *The Journal of Psychology*, 135(2), 133–153.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, 100(3), 1238–1260.
- Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939–944.
- Epper, T., Fehr, E., Fehr-Duda, H., Kreiner, C. T., Lassen, D. D., Leth-Petersen, S., & Rasmussen, G. N. (2020). Time discounting and wealth inequality. *The American Economic Review*, 110(4), 1177–1205.
- Espín, A. M., Brañas-Garza, P., Herrmann, B., & Gamella, J. F. (2012). Patient and impatient punishers of free-riders. *Proceedings of the Royal Society B Biological Sciences*, 279(1749), 4923–4928.
- Espín, A. M., Correa, M., & Ruiz-Villaverde, A. (2019). Patience predicts within-group cooperation in an ingroup bias-free way. *Journal of Behavioral and Experimental Economics*, 83, 101456.
- Espín, A. M., Exadaktylos, F., Herrmann, B., & Brañas-Garza, P. (2015). Short-and long-run goals in ultimatum bargaining: Impatience predicts spite-based behavior. *Frontiers in Behavioral Neuroscience*, 9, 214.
- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., & Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Netspar Discussion Paper No. 01/2016-003*.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645–1692.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2), 351–401.
- Giné, X., Goldberg, J., Silverman, D., & Yang, D. (2017). Revising commitments: Field evidence on the adjustment of prior choices. *The Economic Journal*, 128(608), 159–188.
- Golsteyn, B. H., Grönqvist, H., & Lindahl, L. (2014). Adolescent time preferences predict lifetime outcomes. *The Economic Journal*, 124(580), F739–F761.
- 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.
- Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399–425.
- Jackson, M. O., & Yariv, L. (2014). Present bias and collective dynamic choice in the lab. *The American Economic Review*, 104(12), 4184–4204.
- Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, 77(2), 129–146.
- Kim, H. B., Choi, S., Kim, B., & Pop-Eleches, C. (2018). The role of education interventions in improving economic rationality. *Science*, 362(6410), 83–86.
- Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences*, 15(3), 213–222.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443–478.
- Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 151, 61–72.
- Lawyer, S. R., Schoepflin, F., Green, R., & Jenks, C. (2011). Discounting of hypothetical and potentially real outcomes in nicotine-dependent and nondependent samples. *Experimental and Clinical Psychopharmacology*, 19(4), 263–274.
- Lührmann, M., Serra-Garcia, M., & Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, 10(3), 309–332.

- Macchia, L., Plagnol, A. C., & Reimers, S. (2018). Does experience with high inflation affect intertemporal decision making? Sensitivity to inflation rates in Argentine and British delay discounting choices. *Journal of Behavioral and Experimental Economics*, 75, 76–83.
- Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology*, 11(2), 139–145.
- Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., & Wegener, A. A. (2004). Delay discounting of potentially real and hypothetical rewards: II. Between- and within-subject comparisons. *Experimental and Clinical Psychopharmacology*, 12(4), 251–261.
- Martín, J., Brañas-Garza, P., Espín, A. M., Gamella, J. F., & Herrmann, B. (2019). The appropriate response of Spanish Gitanos: Short-run orientation beyond current socio-economic status. *Evolution and Human Behavior*, 40(1), 12–22.
- Matusiewicz, A. K., Carter, A. E., Landes, R. D., & Yi, R. (2013). Statistical equivalence and test-retest reliability of delay and probability discounting using real and hypothetical rewards. *Behavioural Processes*, 100, 116–122.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Sepa-rate neural systems value immediate and delayed monetary rewards. *Science*, 306(5695), 503–507.
- Meier, S., & Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal Applied Economics*, 2(1), 193–210.
- Meier, S., & Sprenger, C. D. (2013). Discounting financial literacy: Time preferences and participation in financial education programs. *Journal of Economic Behavior and Organization*, 95, 159–174.
- Matousek, J., Havranek, T., & Irsova, Z. (2022). Individual discount rates: A meta-analysis of experimental evidence. *Experimental Economics*, 25(1), 318–358.
- Non, A., & Tempelaar, D. (2016). Time preferences, study effort, and academic performance. *Economics of Education Review*, 54, 36–61.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.
- Paola, M. D., & Gioia, F. (2014). Does patience matter in marriage stability? Some evidence from Italy. *Review of Economics of the Household*, 15(2), 549–577.
- Paola, M. D., & Gioia, F. (2017). Impatience and academic performance: Less effort and less ambitious goals. *Journal of Policy Modeling*, 39(3), 443–460.
- Perez-Arce, F. (2017). The effect of education on time preferences. *Economics of Education Review*, 56, 52–64.
- Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2), 185–199.
- Prissé, B. (2022). Visual continuous time preferences: Lab, field and high schools. *Mimeo*.
- Prissé, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics*, 100, 101910.
- Rachlin, H. (2002). Altruism and selfishness. *Behavioral and Brain Sciences*, 25(02), 788.
- Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179.
- Schaner, S. (2015). Do opposites detract? Intrahousehold preference heterogeneity and inefficient strategic savings. *American Economic Journal Applied Economics*, 7(2), 135–174.
- Sozou, P. (1998). On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society B Biological Sciences*, 265(1), 2015–2020.
- Sunde, U., Dohmen, T., Enke, B., Falk, A., Huffman, D., & Meyerheim, G. (2022). Patience and comparative development. *The Review of Economic Studies*, 89(5), 2806–2840.
- Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents field behavior. *The American Economic Review*, 103(1), 510–531.
- Takeuchi, K. (2011). Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, 71, 456–478.
- Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *The American Economic Review*, 100(1), 557–571.
- Ubfal, D. (2016). How general are time preferences? Eliciting good-specific discount rates. *Journal of Development Economics*, 118, 150–170.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75–98.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.