Experimental Economics (2022) 25:593–623
https://doi.org/10.1007/s10683-021-09723-w

ORIGINAL PAPER

Magnitude effect in intertemporal allocation tasks

Chen Sun[1] · **Jan Potters**[2]

Received: 7 November 2019 / Revised: 18 May 2021 / Accepted: 31 May 2021 /
Published online: 10 June 2021
© The Author(s) 2021

Abstract
Intertemporal choices are affected by both discount rate and utility curvature. We investigate how the two aspects of time preference are affected by the size of the total budget using an intertemporal allocation task. At the aggregate level as well as at the individual level, we find magnitude effects both on the discount rate and on intertemporal substitutability (i.e., utility curvature). Individuals are more patient when dealing with larger budgets and also regard larger budgets to be more fungible. The latter effect suggests that the degree of asset integration is increasing in the stake.

Keywords Time preference · Magnitude effect · Convex Time Budget method · Asset integration

JEL Classifications C91 · D12 · D91

1 Introduction

The prediction of the standard consumption-savings model, that people always discount an income at the market interest rate, has been found inconsistent with empirical results.[1] One important anomaly, dating back to Thaler (1981), is the magnitude

[1] To be more precise, people discount at the market rate unless they do not have enough liquidity to smooth consumption between relevant periods. Loewenstein and Thaler (1989) elaborate on why this is the prediction of the standard theory and summarizes three related anomalies. The prediction implies that atemporal utility function should appear to be linear when the outcomes are small and the discount rate should equal the market interest rate.

✉ Chen Sun
chen.sun@hu-berlin.de

Jan Potters
j.j.m.potters@uvt.nl

[1] Humboldt-Universität zu Berlin, Berlin, Germany

[2] Department of Economics and CentER, Tilburg University, Tilburg, Netherlands

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

effect: when comparing a smaller-sooner reward with a larger-later reward, people favor the later reward more often as the amounts of the two rewards are scaled up. Studies on the magnitude effect help us better understand how people make intertemporal choices differently for small and large rewards, for instance, how people allocate inheritance or lottery income over time.

While several experiments have found a magnitude effect, no study has explored the underlying mechanisms.[2] Intertemporal choices are affected by both the discount rate and the curvature of the atemporal utility function, where the latter determines intertemporal substitutability. Accordingly, the magnitude effect could have two potential causes: people are more patient for larger rewards, and/or people find larger rewards more substitutable across time. The two mechanisms are indistinguishable in a *single-reward task* where a decision-maker can only receive a reward on a single date (either sooner or later), but they have different implications and lead to different behavioral patterns in more general situations.

One such situation is the *intertemporal allocation task*, where a decision-maker allocates a fixed budget between two dates, given a (usually positive) return to the share allocated to the later date. Choices in intertemporal allocation tasks can be characterized by two attributes: the average shares allocated to a particular date given a set of return rates, and the responsiveness of these shares to changes in the rate of return. The former attribute is determined by the discount rate and the latter one by the intertemporal substitutability. Thus, studying the magnitude effect in intertemporal allocation tasks can help us identify the mechanism underlying the magnitude effect.

In this paper, we perform a lab experiment to investigate how choices in intertemporal allocation tasks change with the size of the total budget, and in particular, whether the stakes impact intertemporal preferences through patience (the discount rate) or intertemporal substitutability (the atemporal utility function).

Some theories provide an explanation for the magnitude effect in single-reward tasks. Benhabib et al. (2010) posit that a fixed cost of waiting makes people impatient to small outcomes, but it matters less when outcomes are large. Noor (2011) establishes a magnitude-dependent discounting model where the discount rate of a dated outcome is decreasing in the size of the outcome. Fudenberg and Levine (2006) predict that people exert costly self-control when stakes are high but indulge themselves when stakes are low, which generates a magnitude effect. Holden and Quiggin (2017) assume that people take into account more background consumption when experimental rewards are larger, which also explains the magnitude effect in single-reward tasks. When those theories (with proper extension) are applied to intertemporal allocation tasks, Benhabib et al. (2010) and Noor (2011) predict a magnitude effect on the discount rate, while Fudenberg and Levine (2006) and Holden and Quiggin (2017) predict a magnitude effect on the utility curvature. Our

[2] See, for instance, Holcomb and Nelson (1992), Kirby (1997), Kirby et al. (1999), Andersen et al. (2013), Sutter et al. (2013), Halevy (2015). Frederick et al. (2002) summarize the early literature on the magnitude effect of time preferences. Andersen et al. (2013) review the more recent literature.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

experiment helps to further distinguish the explanatory power of these different models.

We employ the Convex Time Budget (CTB) method introduced by Andreoni and Sprenger (2012). It allows subjects to form a portfolio of a sooner reward and a later reward given a budget constraint. The possibility for subjects to make interior choices (and not only corner choices as in single-reward tasks) enables researchers to simultaneously identify the discount rate and the intertemporal substitutability. In particular, the identification of the intertemporal substitutability comes from subjects' allocation of rewards as a response to changes in the interest rate, which is exactly the definition of the elasticity of intertemporal substitution, and hence the method provides a robust measure against misspecification.[3]

The design of our experiment has three main features. First, all subjects receive equal amounts of participation fees on the sooner date and the later date regardless of their choices, and the payment conditions are constant across time. Thus, the transaction costs and the trustworthiness of the payments are equalized across periods. Second, we implement two treatments. In one treatment subjects allocate between today and 4 weeks later, while in the other treatment subjects allocate between 4 weeks later and 8 weeks later. This allows us to assess whether the magnitude effect is affected by the inclusion of a front-end delay. Finally, by assuming a simple yet popular model, the CTB method allows us to identify the discount rate and the atemporal utility function simultaneously. As a result, we can disentangle the channels of the magnitude effect.

We find evidence of the magnitude effect in intertemporal allocation tasks: the budget share allocated to the later date is increasing in the total budget. The size of the magnitude effect is found to be decreasing in the stake. The pattern is not affected by whether or not a front-end delay is present. At the aggregate level as well as at the individual level, we find magnitude effects both on the discount rate and on intertemporal substitutability. Both channels have considerable impacts on predicted choices. We find that the latter effect is not the same as the magnitude effect on risk attitudes found in previous studies, and hence it might be problematic to correct for the curvature of utility functions by risk attitudes. Instead, the magnitude effect on intertemporal substitutability is consistent with theories proposing that people integrate experimental rewards with more background wealth as the size of rewards gets larger.

The remaining part of this paper is structured as follows: We introduce our experimental design in Sect. 2. In Sect. 3 we formulate our hypotheses. In Sect. 4, we investigate non-parametrically the magnitude effect and its relation with the front-end delay. We explore the channels by parametric estimation both at the aggregate

[3] Abdellaoui et al. (2013a) provided another method for measuring intertemporal preferences parametrically. Their method identifies utility curvature from marginal utilities for different quantities on the same date, while the CTB method identifies utility curvature from sensitivities of choices to interest rates. The two methods are equivalent if the true model is with a power period utility function and a magnitude-independent discount function. If the condition is not satisfied, the former method is better at measuring curvature of a period utility function, and the CTB method is better at measuring sensitivity of choices to interest rates (or, elasticity of intertemporal substitution).

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

level and at the individual level in Sect. 5. In Sect. 6, we discuss the interpretations of our findings. We conclude in Sect. 7.

2 Experimental design

2.1 The Convex Time Budget method, parameters, and implementation

The basis of our experimental design is the Convex Time Budget method introduced by Andreoni and Sprenger (2012). The method consists of a set of intertemporal allocation tasks. In each *decision,* subjects are asked to allocate N tokens to two dates: t days from today, and $(t + \tau)$ days from today. Each token allocated to t is worth P_t euros, while each token allocated to $(t + \tau)$ is worth $P_{t+\tau}$ euros. If a subject allocates n_t tokens to the sooner date and $n_{t+\tau}$ to the later date, the sooner reward is $z_t = P_t \cdot n_t$ euros and the later reward is $z_{t+\tau} = P_{t+\tau} \cdot n_{t+\tau}$ euros.

Choices are subject to the budget constraint, $n_t + n_{t+\tau} \leq N$, and the non-negativity constraints, $0 \leq n_t, n_{t+\tau} \leq N$. Subjects are told that they can allocate any number of tokens they like to one of the two dates. Examples of both corner choices and interior choices are given to remove any hesitation in making either type of choice.

Decisions with the same total budget, N, are grouped in one *decision form*, which is displayed on one page. There are seven decisions in each decision form. The return to each token allocated to the later date is fixed as $P_{t+\tau} = €0.20$, while the return to each token allocated to the sooner date is varied and takes the values $P_t = €0.20$, €0.19, €0.18, €0.17, €0.16, €0.15, and €0.14. Hence, those returns imply seven gross interest rates, $R = 1$, 1.05, 1.11, 1.18, 1.25, 1.33, and 1.43, respectively, over a period of τ days. The constraints can be rewritten as

$$R \cdot z_t + z_{t+\tau} \leq m$$

$$z_t, z_{t+\tau} \geq 0$$

where m is the total budget and $m = P_{t+\tau} \cdot N$.

We implement the CTB method by a zTree program (Fischbacher, 2007). Figure 1 shows the interface of a typical decision form. Each decision takes a row. Decisions can be made by scrolling the bars. Once an adjustment is made for one decision, the amounts of the sooner reward and of the later reward in that decision are automatically calculated and displayed.

To avoid any possible effects of initial values, the amounts of rewards are initially blank. Decisions cannot be submitted until all the scroll bars have been adjusted at least once.

2.2 Procedures

There are two parts to our experiment. Part I consists of five decision forms, with $N = 100$, 200, 300, 400, and 800, respectively. The order is randomly drawn for each

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

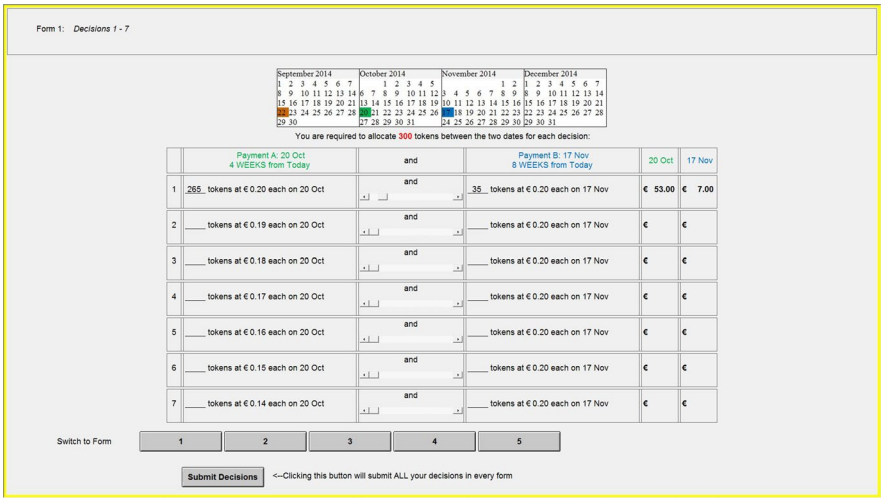


Fig. 1 The interface of a typical decision form in Part I

subject. Subjects can move to a specific decision form by clicking the button with the corresponding number. One can go to any decision form at any time, regardless of whether the current decision form is completed. Decisions are automatically stored when one switches to another decision form. This makes comparisons across total budgets very easy to the subjects in case they would want to make such comparisons. Decisions can only be submitted when all the 35 decisions are completed.

We randomly assign subjects to one of two treatment groups. In the Present Group, the sooner date is today while the later date is 4 weeks from today, i.e., $t = 0$ and $\tau = 28$. In the Delayed Group, the sooner date is 4 weeks from today while the later date is 8 weeks from today, i.e., $t = 28$ and $\tau = 28$. Comparing the two groups enables us to check if there exists a present bias on average and, more importantly, if there exists a magnitude effect when no rewards are available in the present.[4]

We also test the intertemporal independence of preferences over monetary rewards.[5] One alternative hypothesis is that a subject in the Delayed Group may allocate less to the sooner date if she has allocated a large amount of money to an even sooner date since the desire for extra consumption has already been partly satisfied. A similar hypothesis applies to the Present Group: a subject in the Present Group

[4] Identifying present bias would be possible under weaker assumption by varying the presence of a front-end delay within subjects. We choose for a between subject design on this dimension to keep the experiment relatively short and to focus on within-subject variations of the interest rate and the budget size which are central to our analysis.

[5] The property that preferences over alternatives are not affected by a common increase in outcome in the same period when uncertainty is not involved is often called "time separability" in early literature of time preference. However, when uncertainty is considered, the term time separability also has the meaning of correlation neutrality (as opposed to correlation aversion or seeking) and hence the use of it may cause confusion. We thank an anonymous referee for pointing this out.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 1 Restrictions on the number of tokens and returns to one token allocated to a specific date in Part II

	Today	Four weeks from today	Eight weeks from today
Present Group			
Returns to one token	€0.20, €0.19, €0.18, €0.17, €0.16, €0.15, €0.14	€0.20	€0.26
Restriction on the number of tokens	No additional restriction	No additional restriction	0 or 200
Delayed Group			
Returns to one token	€0.08	€0.20, €0.19, €0.18, €0.17, €0.16, €0.15, €0.14	€0.20
Restriction on the number of tokens	200 or 400	No additional restriction	No additional restriction

may allocate less to the later date if she has already allocated a large amount of money to an even later date since the guilt for not saving has been partially released. If preferences are intertemporal dependent, the use of a model with a time-separable preference is likely to be problematic. Thus, we want to test the hypothesis of intertemporal independence before we perform a parametric estimation with a time-separable model.

We use Part II to test the intertemporal independence. It is composed of an extended CTB decision form with seven decisions. Subjects are asked to allocate 400 tokens to three dates, today, 4 weeks from today, and 8 weeks from today. One additional restriction is imposed, depending on which group one is in. A subject in the Present Group can allocate either 0 or 200 tokens to 8 weeks from today; she cannot choose other numbers. But she is still free to allocate any number of tokens between today and 4 weeks from today. Similarly, a subject in the Delayed Group can allocate either 200 or 400 tokens to today. She is still free to allocate any number of tokens (if there remains some) between 4 weeks from today and 8 weeks from today. The restrictions and the returns to one token allocated are shown in Table 1.

The additional date (8 weeks from today for the Present Group or today for the Delayed Group) is accompanied by a very high return for the Present Group and a very low return for the Delayed Group so that subjects are induced to allocate 200 tokens to this additional date. If they do so, the remaining task is equivalent to the one with a total budget of 200 tokens in Part I. This characteristic makes the two decision forms comparable.

We do not directly give a fixed reward on the additional date. This is because a fixed reward might be mentally isolated from the allocation task due to narrow bracketing, and hence the test of intertemporal independence in the allocation task may be invalid.

At the end of the experiment, subjects were asked to finish a questionnaire. As in previous studies with the CTB method, we asked about subjects' expenditures in a typical week. The average response was €55.22 per week or €7.89 per day.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

2.3 Experimental payments

The payments are composed of two parts. First, all subjects receive a €5 *participation fee* on each of the two dates scheduled in Part I. Second, each subject has a 10% chance to receive *earnings from decisions*. Before the experiment starts, each subject is randomly given a lottery number, ranging from 0 to 9. After all subjects in a session finish the questionnaire, the experimenter invites one of the subjects to roll a ten-sided die in front of all subjects in the session. Subjects who have a lottery number that equals the die roll get the earnings from decisions. One decision is randomly selected from the 42 decisions in the two parts as the *decision that counts*. If the decision that counts is from Part I, the allocation in that decision will be realized as the earnings from decisions. If the decision that counts is from Part II, the allocation will be realized and the subject will also receive a €5 participation fee on the additional date in Part II; hence a subject will receive three participation fees if a decision in Part II is realized. All the rules above were articulated in the instructions, and the instructions were always read aloud before either part of the experiment.

The earnings were paid by bank transfer to subjects' checking accounts. We made orders of transfers soon after the experiment and sent reminder emails with information about the incoming amounts on the day of the experiment and all the payment dates. Given the reliability of the banking service, subjects can expect to receive all delayed payments exactly on the appropriate payment dates, while some of the present payments might be received one day after the experimental day due to the inter-bank processing.

We believe the payment tool we used was as good as cash in terms of liquidity. Checking accounts are used in private transactions such as paying for rents. Checking accounts are also linked to debit cards. In the Netherlands, debit cards are widely used for daily transactions in almost all kinds of stores including supermarkets, university restaurants, and bookstores without any transaction fees. We held a survey about subjects' use of debit cards in the questionnaire. The responses show that bank transfers give high liquidity to the rewards so that no isolation effect should be expected due to the payment method.[6]

2.4 Transaction costs and credibility of payments

For our experiment, it is extremely important to equalize the transaction costs and the trustworthiness of the payments across periods, because a difference in the transaction costs over the two periods can be a confounding factor of the magnitude effect.

Several facilities were employed to equalize the transaction costs across periods and to increase the credibility of the payments. The transaction costs include the

[6] 84.7% of the subjects pay at least 50% of their expenditure in general by debit card, while 91.1% pay at least 30%. Among those who pay less than 30% of their expenditure in general by debit card, 61.1% pay at least 30% of their expenditure in university restaurants or in supermarkets by debit card. Among the remaining seven subjects, four withdraw cash at least 3 times per month.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

costs to collect rewards, to confirm that the rewards have been received with correct amounts, and to remember the earnings so that they can be consumed on the expected dates.

First, we sent reminder emails with information about the incoming amounts on the day of the experiment and all the payment dates. Subjects knew this from the instructions, so they did not need to worry about forgetting the earnings on the payment dates, a situation in which the expected marginal utility of the delayed rewards might be lowered.

Second, as Andreoni and Sprenger (2012) did, we delivered our business card and told the subjects to contact us immediately in case they would not receive a payment on time. It increased the credibility of payments and meanwhile served as a reminder of the payments.

Third, we asked subjects to fill in a payment reminder card with the amounts of their rewards on the corresponding dates just after their earnings were displayed. This served as a second reminder in case they forget to check emails.

In sum, the characteristics that one will receive a participation fee on each payment date and that all payments will be received by bank transfer help equalize the transaction costs of receiving payments on all dates. At the same time, the business cards, the payments reminder cards, and the reminder emails reduced the risk of forgetting the rewards. The business cards also lowered the perceived default risks. Even though the risk might still be perceived by some subjects, it should be equal across periods since the payment tools and all auxiliary facilities were the same.

2.5 Sample

Our experiment was conducted at the CentERlab, Tilburg University.[7] 203 students of the university participated in one of the 11 sessions, 94 in the Present Group and 109 in the Delayed Group. Each subject made 42 decisions. One session took one hour and ten minutes on average. 22 subjects got the earnings from decisions, which averaged €69.16. The overall average earning was €17.49.

3 Hypotheses

Most previous studies define the magnitude effect in single-reward tasks. Denote a reward z_t on a sooner date t by (z_t, t). In a single-reward task, a subject chooses between a sooner reward (z_t, t) and a later reward $(z_{t+\tau}, t+\tau)$, where $\tau > 0$ is the delay. A subject displays a *(positive) magnitude effect* if for all $z_t > 0$, $z_{t+\tau} > 0$ and $\eta > 1$,

[7] The payment dates were in September, October and November. The fall semester in Tilburg University started from the end of August and ended in early December. Hence the payment dates were earlier than the final exam weeks and the Christmas vocation, which keeps our experiment from their probably large impacts on the subjects' demand of money.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

$$\left(z_t, t\right) \sim \left(z_{t+\tau}, t+\tau\right) \Rightarrow \left(\eta z_t, t\right) \prec \left(\eta z_{t+\tau}, t+\tau\right).$$

In words, the later reward is more favorable if the amounts are scaled up.

We adapt the definition of the magnitude effect to the intertemporal allocation task. A subject makes a choice $z^*(R,m)$ out of a linear budget set $\left\{\left(z_t, z_{t+\tau}\right) : Rz_t + z_{t+\tau} = m\right\}$, where $z^* = \left(z_t^*, z_{t+\tau}^*\right)$. She displays a *(positive) magnitude effect* if for all $m > 0$ and $\eta > 1$,

$$\frac{z_{t+\tau}^*(R,m)}{m} \leq \frac{z_{t+\tau}^*(R,\eta m)}{\eta m}$$

and

$$z_t^*(R,m) > 0 \Rightarrow \frac{z_{t+\tau}^*(R,m)}{m} < \frac{z_{t+\tau}^*(R,\eta m)}{\eta m}.$$

In words, people put a larger share of the budget on the later date as the total budget is increased. She may only fail to do so in case the sooner reward is already zero. The adapted definition is consistent with the original one. Options with a larger fraction of later reward become more favorable if all options in the menu are scaled up by the same factor.

Hypothesis 1 (magnitude effect on budget share): $\frac{z_{t+\tau}^*}{m}$ is increasing in m.

We are also interested in whether the magnitude effect is affected by the presence of a front-end delay. Benhabib et al. (2010) suggest that a fixed cost of delaying rewards can account for the magnitude effect in single-reward tasks since the fixed cost matters less in case the rewards are scaled up. However, it is not clear if this cost is incurred only when a present reward is delayed or if it applies equally to delaying a future reward. We thus test whether the magnitude effect is smaller or even non-existent if the sooner reward is also in the future.

Hypothesis 2 (a front-end delay leads to a smaller magnitude effect): $\frac{z_{t+\tau}^*}{m}$ changes less with m in the Delayed Group than in the Present Group.

The two hypotheses above can be tested without assuming a specific model.

Conditional on finding a positive magnitude effect, we wish to explore the channels of the magnitude effect. If intertemporal independence is supported by our results (which it will, as we show below), we will estimate the parameters of preferences, with the assumption that subjects maximize a time-separable utility function with power atemporal utility functions and quasi-hyperbolic discounting, i.e., subjects maximize

$$U\left(z_t, z_{t+\tau}\right) = \delta^t \frac{\left(z_t + \omega\right)^\alpha - 1}{\alpha} + \beta\delta^{t+\tau} \frac{\left(z_{t+\tau} + \omega\right)^\alpha - 1}{\alpha}, \tag{1}$$

where β is the present bias parameter, δ is the daily discount factor, α is the exponent parameter. z_t and $z_{t+\tau}$ are the sooner reward and the later reward, respectively. $\omega > 0$

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

is the background consumption mentally integrated with the experimental reward when the decision is made.[8]

When the power utility function is assumed, the elasticity of intertemporal substitution in consumption, $e_c \equiv -\frac{\ln\left(\frac{c_{t+\tau}}{c_t}\right)}{\ln\left(\frac{u'(c_{t+\tau})}{u'(c_t)}\right)}$, is equal to $\frac{1}{1-\alpha}$ (c_t and $c_{t+\tau}$ are the consumption on the sooner date and on the later date, respectively.). Thus, the exponent parameter, α, is a positive transformation of e_c. If $\alpha \to 1$, the atemporal utility function becomes linear, and the elasticity goes to infinity. In that case, subjects just go for the largest present value, and hence rewards are perfectly substitutable between dates. In case $\alpha \to -\infty$, the atemporal utility function is Leontief, and the elasticity goes to zero. In that case, subjects always divide the total budget into two equal amounts. In general, the larger the value of α, the more substitutable the subject considers the two rewards to be. Therefore, α is a measure of intertemporal substitutability.

It brings several advantages to assume such a model. First, the parameters in this model have important economic meanings. The discount factor determines the average choice across interest rates and hence measures the patience of the subject; if a subject is more patient, she will allocate more tokens to the later date for all interest rates. The intertemporal substitutability of consumption between different points in time relates to the dispersion of the choices across interest rates since it measures how sensitive the choices are to the interest rate. These behavioral measures are hard to estimate without assuming a model. Due to the non-negativity constraint, choices are censored at the corners if the preference parameters are extreme. As a result, directly measuring the average choice (as a measure of patience) and the dispersion of choices (as a measure of intertemporal substitutability) leads to biases. In contrast, the model we assume is tractable and easy to estimate. Moreover, the model is widely used in both theoretical and empirical applications.[9]

Given the model above, we test the following two hypotheses.

Hypothesis 3 (magnitude effect on the discount factor): δ is increasing in m.

Hypothesis 4 (magnitude effect on intertemporal substitutability): α is increasing in m.

[8] The atemporal utility function here satisfies the subproportionality defined by Loewenstein and Prelec (1992) if and only if $\omega > 0$. Thus, the magnitude effect on monetary discount rate is guaranteed as long as the background consumption is positive.

[9] To address the concern about misspecification, in Online Appendix D, we check the robustness of our results by estimating the background consumption parameter. The estimate can be a mixture of real background consumption and an inherent shift of the utility function. Depending on the sign of the shift, the elasticity of intertemporal substitution could be increasing or decreasing in the consumption, or be a constant. This kind of flexibility is especially important when the stake is varied in the experiment. The results are the same.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

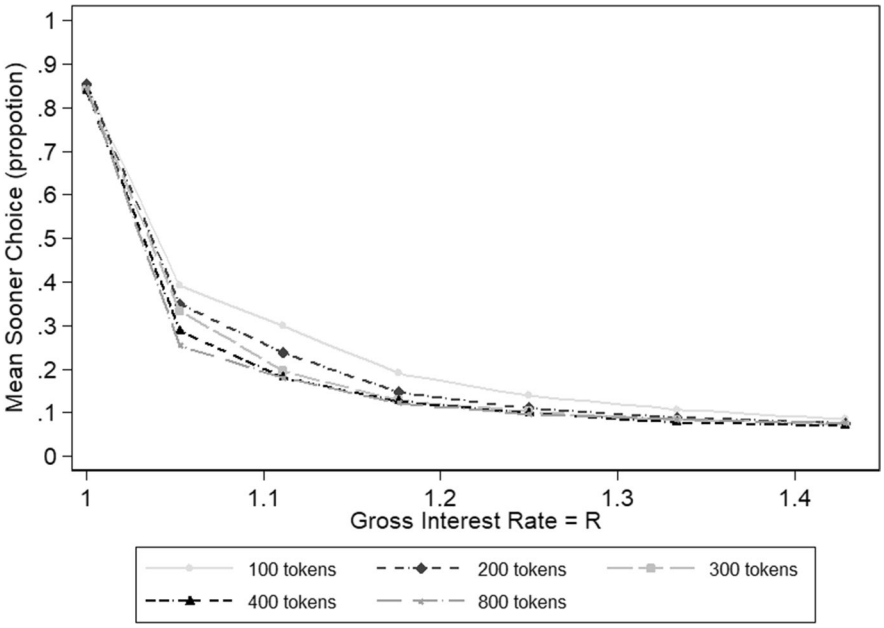


Fig. 2 Mean budget share on the sooner date in Part I

4 Overall effects

4.1 Magnitude effect on budget share

In our data, 28% of the choices are interior, and 62% of our subjects make at least one interior choice. This is very close to the 30% and 63%, respectively, in Andreoni and Sprenger (2012). The relationships between the budget shares and the interest rates are also similar. A discussion about the rationality of subjects in the CTB task is provided in Appendix A.

In Fig. 2 we plot the mean budget share allocated to the sooner date against the gross interest rate, R, of each CTB decision in Part I. We plot separate points for the five total budgets ($m = €20, €40, €60, €80, €160$). The budget share allocated to the sooner date declines with the total budget.

The difference seems to be larger when the interest rate is smaller but still positive. This is mainly due to censoring. When the interest rate is zero ($R = 1$) or the highest ($R = 1.43$), most choices are at the corners for both smaller and larger total budgets.

To judge whether there is a significant magnitude effect, we perform Hotelling's T-squared tests on the mean differences in budget shares between total budgets, taking seven choices with the same total budget as a vector (see Table 2).[10] The null

[10] Hotelling's T-squared test is asymptotically nonparametric, so it can be applied to a large sample in nonnormal cases. We also perform a multivariate signed-rank test (Oja and Randles 2004) and the results are basically the same: the magnitude effects are significant between the stakes of €20 and €40

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 2 Multivariate mean difference tests between total budgets

		€40	€60	€80	€160
€20	F-statistic	3.3184***	5.1162***	6.3670***	6.0477***
	p value	0.0023	0.0000	0.0000	0.0000
€40	F-statistic		1.5477	3.9336***	3.9433***
	p value		0.1533	0.0005	0.0005
€60	F-statistic			1.5761	3.0328***
	p value			0.1444	0.0047
€80	F-statistic				1.7104
	p value				0.1084

Hotelling's T-squared tests on the mean differences in the budget shares between two total budgets for all gross interest rates. 203 sets of observations for each total budget. The degrees of freedom of the F-statistics are (7,196). ***, **, and *Indicate significance at the 1% level, 5% level, and 10% level, respectively

The results are robust against multiple hypotheses testing problems since performing a Holm-Bonferroni correction on any family of independent hypotheses does not change the significance at the 1% level

hypothesis is that the means of choices are the same across total budgets, taking into account the correlation within-subject. This class of tests makes sense because individual heterogeneity may have made different subjects reveal magnitude effects on tasks with different interest rates (e.g., Subject 1 on Interest Rate 1 while Subject 2 on Interest Rate 2) so that the magnitude effects on all choices would be jointly significant, but the effect on choices with any single interest rate might be insignificant. The results show that the magnitude effect is significant between the total budgets of €20 and €40 and between any two non-adjacent total budgets. These results support Hypothesis 1, which states that a larger share of the budget is allocated to the later date when the size of the budget increases.[11]

The results also show that the differences are insignificant between adjacent total budgets larger than €20. Since the allocation is monotonic in the total budget and the differences are significant between non-adjacent total budgets, the insignificance suggests that the magnitude effect is the greatest when comparing between the smallest total budgets (€20 and €40) and becomes smaller for larger stakes. The pattern is consistent with the fact that Andersen et al. (2013) found a "statistically significant" but "not economically significant" magnitude effect when they elicited time preferences using very high stakes.[12]

Footnote 10 (continued)
and between any two non-adjacent stakes at the 5% level for the entire sample and significant at the 10% level for either group. The Holm-Bonferroni correction is already performed.

[11] In Online Appendix B, we present the results of group-wise tests. The results are the same.

[12] Andersen et al. (2013) compare stakes of 1,500 Danish kroner and 3,000 Danish kroner, which are roughly equivalent to €200 and €400, respectively.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

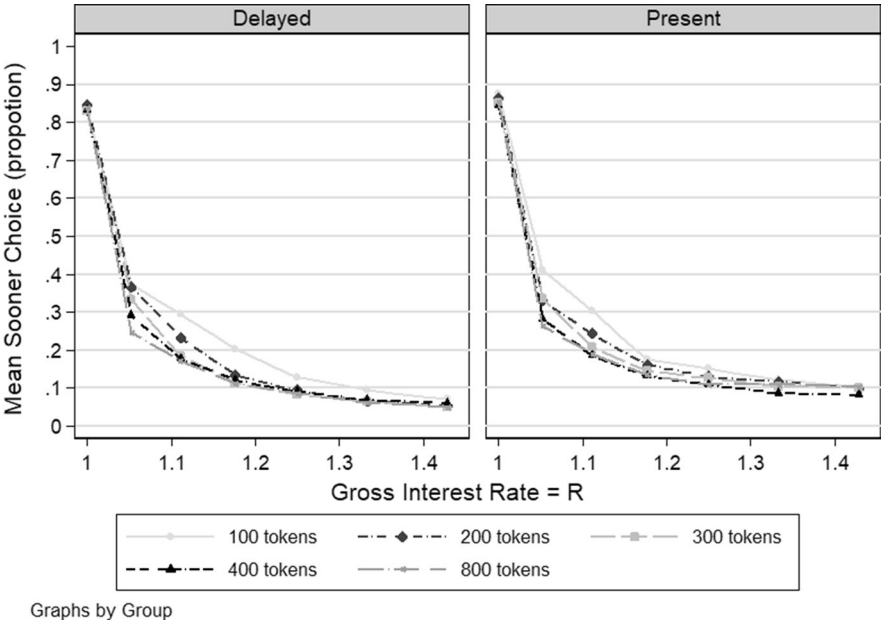


Fig. 3 Mean budget shares in the present group and the delayed group

In Online Appendix B, we show group-wise results of the same tests. The results are the same.

4.2 Conditional on the presence of an immediate reward?

We test if the magnitude effect is affected by the existence of a front-end delay. First, Table B.1 shows the results of Hotelling's T-squared tests on the magnitude effects for the Present Group and the Delayed Group, respectively. We find significant magnitude effects in both groups. This implies that the presence of an immediate reward is not a necessary condition for the magnitude effect.

Second, we plot separate graphs for the two groups in Fig. 3. Subjects in the Delayed Group seem to be slightly more patient than those in the Present Group. However, when we perform Hotelling's T-squared test on all the 35 decisions in Part I between groups, the null hypothesis that the two groups have the same mean responses is not rejected. The p value is 0.2424 when the degree of freedom is (35, 167). Thus, we find no evidence that the magnitude effect is affected by the existence of a front-end delay, and we reject Hypothesis 2.[13]

[13] This result also implies that a present bias is not detected. The present bias here refers to non-stationarity of preferences according to the categorization of Halevy (2015). The result of no present bias is consistent with Sutter et al. (2013) and does not necessarily contradict the stylized fact that the discount rate is decreasing in the time distance between the sooner reward and the later reward, as in Benhabib et al. (2010). That stylized fact and stationarity can hold simultaneously if there is subadditivity in discounting [see Read (2001), Dohmen et al. (2017)].

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 3 Multivariate mean difference tests between Part I and Part II

Subsample	Present	Delayed	Total
F-statistic	1.5560	1.4192	1.0979
Degree of freedom	7, 79	7, 101	7, 187
p value	0.1609	0.2058	0.3662

Hotelling's T-squared tests on the mean differences in the budget shares in the decisions with the total budget of €40 between Part I and Part II. Subjects who chose a different number from 200 tokens for the additional date in Part II such that their choices were not comparable between the two parts have been removed from the sample. ***, **, and *Indicate significance at the 1% level, 5% level, and 10% level, respectively

This finding has an implication for the modeling of the magnitude effect: no matter what generates the magnitude effect, it applies equally to situations with an immediate reward and those without. For instance, if it is a fixed cost of delaying rewards that generates the magnitude effect, as proposed by Benhabib et al. (2010), the cost applies equally to delaying an immediate reward and to delaying a future reward.

4.3 Intertemporal independence

Our results show that Part II is a valid test of intertemporal independence since most subjects chose 200 tokens for the additional date in Part II. Only nine out of 203 subjects selected a different number than 200 to the additional date, which involved 41 (2.9%) out of 1415 decisions.

After removing those decisions, we compare the choices with the total budget of €40 between Part I and Part II, separately for each group. Table 3 shows that Hotelling's T-squared tests fail to reject the null hypothesis that responses to the two parts have the same means. Those results support intertemporal independence, which will be assumed in the next section.[14]

5 Channels

To disentangle the magnitude effect into two channels, we perform parametric estimations both at the aggregate level and at the individual level. We then test if the preference parameters change with the size of the total budget.

[14] Our test of intertemporal independence is not informative of whether subjects are correlation neutral. Since we focus on settings where outcomes are perfectly correlated across periods, for simplicity, we assume a time separable utility function in our estimation.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

5.1 Aggregate-level estimation

5.1.1 Estimation strategy

In our main specification, we assume a time separable utility function with power atemporal utility functions as in Eq. (1). We set ω (background consumption) equal to the average response to the question about one's typical daily expenditure, €7.89, as Andreoni and Sprenger (2012) did in two of their specifications.[15]

Given the intertemporal utility function, solving the optimization problem yields the tangency condition

$$\frac{z_t+\omega}{z_{t+\tau}+\omega}=\begin{cases}(\beta\delta^{\tau}R)^{\frac{1}{\alpha-1}}, & \text{if } t=0\\ (\delta^{\tau}R)^{\frac{1}{\alpha-1}}, & \text{if } t>0\end{cases}.$$

Taking logs gives a linear equation

$$\ln\left(\frac{z_t+\omega}{z_{t+\tau}+\omega}\right)=\left(\frac{\ln\beta}{\alpha-1}\right)\cdot 1_{t=0}+\left(\frac{\ln\delta^{\tau}}{\alpha-1}\right)+\left(\frac{1}{\alpha-1}\right)\cdot\ln R$$

where $1_{t=0}$ is the indicator for the Present Group.

The parameters to be estimated are the present bias parameter, β, the discount factor, δ, and the power curvature parameter, α. The present bias parameter is identified by the differences in allocation between the Present Group and the Delayed Group. If there is a present bias, subjects in the Present Group will allocate more tokens to the sooner date than those in the Delayed Group. The discount factor is identified by one's average choice across different experimental interest rates. A more patient subject will allocate more tokens to the later date in all decisions. The curvature parameter is identified by the dispersion of one's choices across interest rates. Those who consider rewards highly substitutable over time are likely to make corner choices in all decisions, while those with lower elasticity of intertemporal substitution will make choices closer to equal splits.

Following the practice in previous studies (Andreoni & Sprenger, 2012, and Augenblick et al., 2015), we assume a normally distributed error term additive to the log-consumption ratio and consider censoring, to yield the two-limit Tobit model:

[15] To fix the background consumption across subjects brings the advantage that all effects come from the variation in choices rather than also from the variation in the self-reported background consumptions, which may be noisy. We check the robustness by setting ω as individual background consumption and average/individual background consumption combined with the participation fee (See Online Appendix C). The results are basically the same.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

$$l^*_{i,j,k} \equiv \ln\left(\frac{z^*_{t;i,j,k}+\omega}{z^*_{t+\tau;i,j,k}+\omega}\right)$$
$$= \left(\frac{\ln\beta}{\alpha-1}\right)\cdot 1_{t=0;i} + \left(\frac{\ln\delta^\tau}{\alpha-1}\right) + \left(\frac{1}{\alpha-1}\right)\ln R_j + \varepsilon_{i,j,k}, \varepsilon_{i,j,k} \sim N(0,\sigma_k)$$

$$l_{i,j,k} = \begin{cases} \ln\frac{\omega}{m_k+\omega}, & \text{if} \quad l^*_{i,j,k} \le \ln\frac{\omega}{m_k+\omega} \\ l^*_{i,j,k}, & \text{if} \quad \ln\frac{\omega}{m_k+\omega} < l^*_{i,j,k} < \ln\frac{\frac{m_k}{R_j}+\omega}{\omega} \\ \ln\frac{\frac{m_k}{R_j}+\omega}{\omega}, & \text{if} \quad l^*_{i,j,k} \ge \ln\frac{\frac{m_k}{R_j}+\omega}{\omega} \end{cases}$$

where $i = 1,\ldots,203$ denotes Subject i, $j = 1,\ldots,7$ denotes Interest rate j, and k=1,...,5 denotes Total budget k. The error term is allowed to vary across total budgets since giving a larger number of tokens might induce a larger noise, which might be a competing explanation of a larger sensitivity to the interest rate.

The Tobit model can predict corner choices with a natural interpretation. When the background consumption ω is positive, the marginal utility at a zero reward is finite. If the implied interest rate is much higher than the discount rate, the model predicts a latent choice with a negative budget on the sooner date. The individual would be willing to give up part of her background wealth on the sooner date to earn a larger amount on the later date. But in the experiment, she is not allowed to do that. This is naturally captured by the Tobit model by censoring at the later corner. The opposite case occurs when the implied interest rate is much lower than the discount rate. Then the individual would be willing to borrow money from the experimenter if she could, but her choice is censored at the sooner corner.[16]

The model is estimated by the quasi-maximum-likelihood method: when performing the estimation, the error term, ε, is assumed to be i.i.d., while in computing the standard errors, the error term is assumed to be independent across subjects, but might be correlated within-subject. Estimates of the parameters can be recovered and standard errors can be inferred by the delta method.

Since we are interested in the magnitude effect, we also perform the estimation with interaction terms of the parameters and the budget dummies. Thus, tests can be performed on the differences between the parameters for different total budgets.

To see why the measure of the intertemporal substitutability is robust against misspecification, notice that α is identified from the sensitivity of the log-ratio of the consumption (i.e., $\ln\left(\frac{z^*_{t;i,j,k}+\omega}{z^*_{t+\tau;i,j,k}+\omega}\right)$) to the logarithm of the gross interest rate (i.e., $\ln R_j$). This is approximately the sensitivity of the percentage change in the ratio of consumption to the percentage change in the relative price, which is exactly the

[16] An alternative way to model the idiosyncratic shock is to assume the Logit model. We perform estimation for both error structures and find that the Tobit model predicts better than the Logit model, especially in predicting corner choices. Details are available upon request.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 4 Discounting and curvature parameter estimates in the aggregate-level estimation

Model:	Tobit	Tobit				
Total budget:	All	€20	€40	€60	€80	€160
Present bias:$\hat{\beta}$	0.989	0.989	0.989	0.986	0.997	0.986
	(0.018)	(0.021)	(0.018)	(0.018)	(0.017)	(0.020)
Discount factor over four weeks:$\widehat{\delta^{\tau}}$	0.968	0.948	0.961	0.971	0.972	0.982
	(0.012)	(0.016)	(0.013)	(0.012)	(0.011)	(0.012)
Curvature:$\hat{\alpha}$	0.955	0.928	0.947	0.952	0.958	0.968
	(0.004)	(0.007)	(0.005)	(0.005)	(0.004)	(0.003)
S.e. of the error term:$\hat{\sigma}$	3.699	2.294	2.986	3.369	3.857	5.314
	(0.281)	(0.200)	(0.245)	(0.269)	(0.307)	(0.454)
Log-likelihood	− 13,678.51			− 13,538.56		
Observations	7,105			7,105		
Uncensored	1,969			1,969		
Clusters	203			203		

Two-limit Tobit estimators. Power utility function with $\omega = 7.89$ (average reported background consumption). Column 1: assuming that parameters are invariant to total budgets. Column 2–6: assuming that parameters vary with total budgets. Clustered standard errors in parentheses. Log-likelihood has been corrected for the transformation of dependent variables. Standard errors are calculated via the delta method.

definition of the elasticity of intertemporal substitution. Therefore, even if the utility function is misspecified, α is still a measure of intertemporal substitutability.

In Online Appendix D, we assume a more flexible specification, in which the utility function has one additional free parameter. There the background consumption, ω, is also a parameter to be estimated. In this way, we address the concern that the average self-reported background consumption may not match the true background consumption integrated with the experimental rewards in decision making, or the elasticity of intertemporal substitution of the utility function may not be constant (i.e., the power utility function with a fixed consumption shift is misspecified). The results are basically the same.

5.1.2 Results

Table 4 reports the magnitude-invariant estimates and the magnitude-specific estimates of the parameters, respectively. A salient feature is that none of the estimates of β is significantly different from 1, implying no evidence of present bias, which is consistent with our finding in the model-free analysis. The annual discount rate for all the budgets together is 52.7%, which is in the range found by previous studies. The curvature parameters are always significantly smaller than 1, implying that the subjects on average consider the monetary rewards received on different dates imperfectly substitutable, which is also consistent with other studies (e.g. Andreoni & Sprenger, 2012; Andreoni et al., 2015; Augenblick et al., 2015; Cheung, 2020).

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 5 Estimates of parameter differences between total budgets in the aggregate-level estimation

Total budget:		€40	€60	€80	€160
Present bias: $\hat{\beta}$	€20	− 0.000	− 0.003	0.008	− 0.003
		(0.008)	(0.010)	(0.011)	(0.013)
	€40		− 0.003	0.008	− 0.003
			(0.006)	(0.008)	(0.009)
	€60			0.011*	0.000
				(0.006)	(0.009)
	€80				− 0.011
					(0.009)
Discount factor over four weeks: $\widehat{\delta^4}$	€20	0.014**	0.024***	0.025***	0.035***
		(0.006)	(0.008)	(0.009)	(0.010)
	€40		0.010*	0.011*	0.021***
			(0.005)	(0.006)	(0.007)
	€60			0.001	0.011*
				(0.005)	(0.006)
	€80				0.010**
					(0.004)
Curvature: $\hat{\alpha}$	€20	0.018***	0.024***	0.029***	0.040***
		(0.003)	(0.004)	(0.004)	(0.005)
	€40		0.006***	0.011***	0.022***
			(0.002)	(0.002)	(0.003)
	€60			0.006***	0.016***
				(0.002)	(0.002)
	€80				0.011***
					(0.002)
S.e. of the error term: $\hat{\sigma}$	€20	0.692***	1.075***	1.562***	3.019***
		(0.106)	(0.138)	(0.179)	(0.322)
	€40		0.383***	0.871***	2.327***
			(0.092)	(0.125)	(0.267)
	€60			0.487***	1.944***
				(0.108)	(0.258)
	€80				1.457***
					(0.218)

Estimates of parameter differences are inferred from the Two-limit Tobit estimation by the delta method. The estimation assumes power utility with $\omega = 7.89$. Separate parameters are estimated for each total budget among €20, €40, €60, €80, and €160. There are 1,421 observations (203 clusters) for each total budget. Clustered standard errors in parentheses. Standard errors are calculated via the delta method. ***, **, and *Indicate significance at the 1% level, 5% level, and 10% level, respectively

Most importantly, both the discount factor and the curvature parameter are increasing in the stake. To judge if these magnitude effects are significant: Table 5 presents Wald tests over the differences of parameters between total budgets. We find significant magnitude effects both on the discount factor, δ, and on the exponent

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 6 Marginal effects of allowing a parameter to vary with total budgets in the aggregate-level estimation

Total budget:		€40	€60	€80	€160
Parameter values used in prediction:	$\beta_1, \delta_k, \alpha_1$(Delayed):	21.3	34.7	36.6	48.6
	$\beta_1, \delta_k, \alpha_1$(Present):	21.8	36.5	38.1	50.7
	$\beta_1, \delta_1, \alpha_k$(Delayed):	24.8	33.9	44.4	65.5
	$\beta_1, \delta_1, \alpha_k$(Present):	27.7	38.0	49.2	72.8

The changes in choices predicted by the Tobit model using the parameter values indicated by the row title compared with $(\beta_1, \delta_1, \alpha_1)$, for the two groups separately. k in the row titles stands for the total budget in the column title. For instance, the first cell in the first row is the difference between the choices made in the seven decisions with the total budget of €40 predicted by the model with parameter values $(\beta_1, \delta_2, \alpha_1)$ and those predicted by the model with parameter values $(\beta_1, \delta_1, \alpha_1)$. In other words, it is the marginal effect of allowing δ to vary with the total budget from €20 to €40. The unit is 1% of the total budget

parameter, α, which is a positive transformation of the elasticity of intertemporal substitution. The discount factor is increasing in the total budget, meaning that the decision weights on later rewards shift upward when the total budget increases. The elasticity of intertemporal substitution is increasing in the total budget, meaning that the rewards on the two dates are more substitutable to the subjects when a larger total budget is provided. This results in choices closer to the two corners (to which corner depends on whether $\delta R > 1$). Thereby, we verify Hypothesis 3 and Hypothesis 4.

To get an idea about the size of the magnitude effects, we compare the discount rates and the predicted monetary discount rates, respectively, between stakes. The continuous annual discount rate at stakes of €20, €40, €60, €80, and €160 are 0.696, 0.519, 0.384, 0.370, and 0.237, respectively. Thus, the discount rate is 1.3 to 1.6 times larger when the stake is halved. This is larger than the effects found by Andersen et al. (2013). Their discount rate at a stake of 1,500 Danish kroner is only 1.0 to 1.1 times larger than that at a stake of 3,000 Danish kroner. As we mentioned earlier, the difference is consistent with the pattern that the magnitude effect becomes smaller at higher stakes. To incorporate the magnitude effect on the utility curvature, we predict the monetary discount rates at the lowest and the highest stakes. Take the monetary discount rate at the stake of €20 as an example, it is the continuous annual discount rate implied by an indifference relation between a later reward of €20 and an equally good sooner reward, assuming linear utility. We find the monetary discount rate at €20 is 1.058 and that at €160 is 0.257. The former is 4.1 times larger than the latter. Therefore, the two channels of the magnitude effect between the stakes of €20 and €160 are both large. Halevy (2015), for instance, finds that the monetary discount rates measured with a budget of $10 are 1.4 to 1.9 times larger than those with a budget of $100.

To illustrate the relative importance of the two channels of the magnitude effect, we use the estimates above to predict choices in the 35 questions for both the Present Group and the Delayed Group. Table 6 presents the marginal effects of allowing one parameter to vary with the total budget: in each row, we allow only one parameter,

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

either δ or α, to vary with the total budget of the decisions (as indicated by the column title), but fix the other two parameters at the value estimated from the budget of €20. Each number in a cell is the total change (in units of $\frac{N_k}{100}$, the percentage of the total budget) in the seven decisions with the corresponding total budget. The results show that the marginal effect of allowing α to vary with the total budget is at least as large as the marginal effect of allowing δ to vary. This suggests that the magnitude effect on the elasticity of intertemporal substitution is at least as important as the magnitude effect on the discount rate.

5.2 Individual-level estimation

The aggregate-level estimation provides evidence of positive magnitude effects on the discount factor and intertemporal substitutability. One may wonder whether these results are purely a compositional effect: a bias resulting from forcing all subjects to have the same preferences and the same distribution of noise. To deal with this concern, we also perform individual-level estimation and tests.

5.2.1 Estimation and testing procedure

We keep all the assumptions that underlie Eq. (1) except for β since it is not identified in individual-level estimations. We estimate the discount factor (δ) and the intertemporal substitutability (α) for each combination of subject and stake, and then test if the two parameters are increasing in the magnitude within-subject.

One important difference from the aggregate-level estimation is that under-identification occurs when a subject made no or only one interior choice at one stake. There are 627 out of 1015 (62%) subject-stake combinations for which this is the case. We thereby adopt a conservative way to test the magnitude effect. First, we obtain point estimates of δ and α if possible by estimating the Tobit model specified in Sect. 5.1. Whenever there is an under-identification problem, we remove the error term from the model and then infer the intervals of δ and α that can generate the observations. Second, we perform a one-sided sign test on the two parameters, respectively, with the null hypothesis that they do not change with the magnitude. The sign test is flexible in that it does not impose assumptions on the distributions of idiosyncratic shocks to the parameters. For a comparison between a point estimate and an interval estimate, we recognize a difference only if the point is not in the interior of the interval. For a comparison between two interval estimates, we recognize a difference if the two intervals do not overlap.[17]

[17] To illustrate the procedure of our test, we take a numerical example. Suppose a subject made seven interior choices when the total budget is €20 and made seven corner choices when the total budget is €160. In the former case, point estimates can be obtained through the Tobit estimation: for example, $\widehat{\delta^\tau} = 0.761$ and $\hat{\alpha} = 0.830$. In the latter case, the observations can only be generated by $\widehat{\delta^\tau} \in [0.95, 1]$ and $\hat{\alpha} \in [0.992, 1]$. Since the point estimate of $\hat{\alpha}$ at the stake of €20 is left to the range of feasible parameter values of $\hat{\alpha}$ at the stake of €160 (i.e., $0.830 < 0.992$), we take it that $\hat{\alpha}$ increases as the total budget rises from €20 to €160. Suppose, instead, that the point estimate of $\hat{\alpha}$ is 0.970 at the stake of €20 while keeping the possible range of $\hat{\alpha}$ at the stake of €160 the same. Since in this case the point estimate falls into the latter range of feasible values (i.e., $0.992 < 0.970 < 1$), we take $\hat{\alpha}$ to be "unchanged" as the total budget rises from €20 to €160

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

We also categorize the subjects based on how their preference parameters change with the stake. For each parameter, a subject presents values for five stakes, allowing for 10 pairwise comparisons across stakes. We use the following two criteria for our categorization: first, we define subjects who present no magnitude effect as those who have "unchanged" results in all the 10 comparisons across stakes, subjects who present positive magnitude effects as those who have no "decrease" and at least one "increase", and subjects who present negative magnitude effects as those who have no "increase" and at least one "decrease", and call the rest as subjects who present a mixed magnitude effect. Arguably this criterion is too stringent in that it does not allow for any error. Our second criterion allows for error: we define subjects who present positive magnitude effects as those who have more "increases" than "decreases" and subjects who present negative magnitude effects as those who have more "decreases" than "increases".

5.2.2 Results

Table 7 shows the results of the tests at the individual level. We reject the null hypotheses of no magnitude effect on the two parameters, in favor of positive magnitude effects. This shows that the two channels of the magnitude effect on intertemporal choices are robust against individual heterogeneity.

Table 8 shows the results of subject categorization. Under both criteria, subjects who present positive magnitude effects are much more frequent than those who present negative magnitude effects. This is consistent with our main finding that positive magnitude effects exist.[18]

6 Interpretations

The results above imply that when an average subject faces a larger budget in an intertemporal allocation task, she behaves more patiently, but also she regards rewards to be more substitutable between dates.

6.1 Relation with the magnitude effect on risk aversion

According to the Discounted Expected Utility (DEU) theory, when the atemporal utility function is a power function, the risk attitude and the elasticity of intertemporal substitution are represented by the same parameters, since risk aversion and imperfect fungibility both originate from diminishing marginal utility. Therefore,

[18] In both the individual-level test and the categorization, there are a large fraction of choices or subjects that present no magnitude effect. This is mainly because many subjects present no or only one interior choice for a stake, and hence we are often unable to sign the comparison. We then take a conservative procedure: we mark inconclusive comparisons as unchanged and perform sign tests, which only rely on comparisons between increases and decreases

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 7 Sign tests on preference parameters between total budgets

		€40	€60	€80	€160
Discount factor over four weeks: $\widehat{\delta^{\tau}}$	#increase/unchanged/decrease relative to €20	31%/48%/21%	35%/48%/17%	38%/48%/14%	36%/47%/17%
	p value	0.0255**	0.0002***	0.0000***	0.0002***
	#increase/unchanged/decrease relative to €40		31%/51%/19%	34%/50%/16%	36%/49%/15%
	p value		0.0107**	0.0003***	0.0000***
	#increase/unchanged/decrease relative to €60			26%/57%/17%	29%/54%/17%
	p value			0.0347**	0.0086***
	#increase/unchanged/decrease relative to €80				23%/58%/19%
	p value				0.1928
Curvature: $\hat{\alpha}$	#increase/unchanged/decrease relative to €20	32%/58%/11%	33%/57%/10%	32%/58%/10%	33%/60%/7%
	p value	0.0000***	0.0000***	0.0000***	0.0000***
	#increase/unchanged/decrease relative to €40		30%/58%/13%	31%/58%/11%	32%/59%/9%
	p value		0.0002***	0.0000***	0.0000***
	#increase/unchanged/decrease relative to €60			24%/60%/16%	31%/60%/9%
	p value			0.0374**	0.0000***
	#increase/unchanged/decrease relative to €80				29%/61%/10%
	p value				0.0000***

Right-tailed sign tests on the differences of parameters between two total budgets. 203 sets of observations for each budget. $\omega = 7.888$. When under-identification occurs, interval estimates are yielded for the two preference parameters. A point estimate and an interval estimate are considered to be different only if the point is not in the interior of the interval. Two interval estimates are considered to be different only if their intersection is empty or a singleton. ***, **, and *Indicate significance at the 1% level, 5% level, and 10% level, respectively.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 8 Categorization of subjects

		Positive magnitude effect	Negative magnitude effect	Mixed magnitude effect	No magnitude effect
Criterion 1: No error allowed	Discount factor over four weeks: $\hat{\delta}^{\tau}$	27 (13%)	4 (2%)	85 (42%)	87 (43%)
	Curvature: $\hat{\alpha}$	30 (15%)	3 (1%)	64 (32%)	106 (52%)
Criterion 2: Error allowed	Discount factor over four weeks: $\hat{\delta}^{\tau}$	75 (37%)	31 (15%)	10 (5%)	87 (43%)
	Curvature: $\hat{\alpha}$	76 (37%)	15 (7%)	6 (3%)	106 (52%)

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

one may wonder whether the magnitude effect on intertemporal substitutability is the same as the magnitude effect on risk attitudes.

We find evidence against this equivalence when we compare our results with previous findings. Studies using risky prospects find that the Arrow–Pratt measure of relative risk aversion (Pratt, 1964) is larger when outcomes are scaled up (Binswanger, 1981; Kachelmeier & Shehata, 1992; Holt & Laury, 2002; Harrison et al., 2005; Fehr-Duda et al., 2010; Bouchouicha & Vieider, 2017). This is in the direction that is opposite to what we find. Increasing relative risk aversion suggests an increase in the concavity of a power utility function as the stake increases while our results show a movement towards linearity. This contradiction suggests that the magnitude effect on relative risk aversion is not driving the magnitude effect on intertemporal substitutability.

Some other studies also suggest that risk aversion and intertemporal substitutability should be separated. Andreoni and Spregner (2012) found no significant correlation at the individual level between the curvature estimated by the CTB method and the risk attitude elicited by the MPL method. Abdellaoui et al. (2013b), Miao and Zhong (2015), Cheung (2015). also found that the utility curvature elicited from intertemporal tasks is quantitatively different from that elicited from tasks with risk. We provide evidence from a different perspective: while the previous studies showed that the degrees of concavity are different for the two kinds of utility functions, we show that the degrees of concavity change in opposite directions when the stake is varied.

This finding has implications for both theories and experimental methods. First, it lends support to the theories which separate intertemporal substitutability from risk aversion (Kihlstrom & Mirman, 1974; Richard, 1975; Kreps & Porteus, 1978; Epstein & Zin, 1989; Weil, 1990; Bommier, 2007; Ebert and van de Kuilen 2016, etc.) Second, it casts doubt on the use of a risk-elicitation task to correct for the curvature when eliciting time preferences.

6.2 Relation with borrowing constraints

In theory, a binding borrowing constraint can lead to a magnitude effect on the monetary discount rate in a single-reward task if the background consumption is expected to grow over time, as shown by Epper (2015). However, Meier and Sprenger (2010) found that experimentally elicited long-run discount rates are uncorrelated with credit constraints, suggesting that on average, whether the borrowing constraint is binding does not affect intertemporal choices in experiments where all outcomes are in the future.

Moreover, given the fact that subjects may have savings that provide some but small liquidity, the number of subjects whose borrowing constraints are binding should be increasing in the stake. For this reason, if the borrowing constraint is the main issue, we should observe that the intertemporal substitutability is decreasing in the stake, which is inconsistent with our results. Therefore, we believe that a binding borrowing constraint is not the main driver of our results.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

6.3 Relation with existing theories

We discuss the implications of our empirical findings for some theories that may explain the two channels of the magnitude effect.

One model that can account for the magnitude effect on the discount factor was proposed by Benhabib et al. (2010). They developed a model with a fixed cost of delaying rewards. The idea is that whenever a delayed reward is chosen, a fixed cost is incurred so that as the stake increases, the cost becomes relatively less important and hence the subject appears more patient.

Noor (2011) proposed a model of magnitude-dependent discounting, which leads to similar predictions. In his model, the discount function is increasing in the utility at the later period. As the stake gets larger, the discount function converges to 1.[19]

One theory that can explain the magnitude effect on intertemporal substitutability is an extended version of the dual-self bank-nightclub model of Fudenberg and Levine (2006). In the original model, the agent first chooses the amount of pocket cash when no temptation is present, and then she chooses the amount of consumption when a windfall is available and temptation plays a role. The strategy for utility maximization is to spend the complete windfall when it is small but try to save some money out of the windfall when it is large. A small windfall is not integrated with the lifelong wealth, because the agent does not bother to perform self-control, but it is worth controlling oneself when the windfall is large. As a result, the utility function for windfalls is much more concave when the size of the windfall is below a certain threshold than when it is above the threshold.

The model can explain a magnitude effect on intertemporal substitutability if we impose the assumption that an agent who anticipates a reward in the future does not immediately adjust her cash allocation plan. Instead, she keeps the anticipated reward in the mental account of windfalls until it is received and part of it is consumed. Only after the remainder is moved into the mental account of savings does she reschedule her future consumption.

When this assumption is used, what the model predicts in intertemporal allocation tasks is consistent with our results. A subject is likely to make interior choices when the budget is small, i.e., below the threshold induced by the self-control costs. Since the utility function for windfalls is very concave, the subject balances extra consumption on the sooner date and the later date. As the budget increases above the threshold, the subject likes to save part of it for consumption smoothing. Since the utility function for savings is much less concave (close to linear), these savings will be allocated fully to either the sooner date (when the interest rate is small) or the later date (when the interest rate is large). Hence, as the budget increases the intertemporal substitutability increases and it will appear as if the utility function has become less concave (see Online Appendix E for a simulation).

Another model that can explain the magnitude effect on intertemporal substitutability is the mental zooming theory proposed by Holden and Quiggin (2017). The theory presumes that people integrate more background consumption with the

[19] Baucells and Heukamp (2012) extend the magnitude-dependent discounting model to allow risks.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

experimental reward as the size of the reward increases. If the budget increases, individuals 'zoom out' as it were and take a broader perspective on the decision problem. One reason may be that individuals are likely to divide and use up a bigger windfall over a longer period. Based on the data collected from their field experiment with Malawian peasants, Holden and Quiggin showed that the magnitude effect on the discount rate in single-reward tasks would disappear if the unobserved background consumption is assumed to be an increasing function of the stake.

In intertemporal allocation tasks, the increasing background consumption can generate a magnitude effect on intertemporal substitutability. To see why, denote the observed elasticity of intertemporal substitution in experimental rewards by e_z. The relationship between e_z and preference parameters is

$$e_z = \frac{1}{1-\alpha} \cdot \frac{\log\left(\frac{z_{t+\tau}}{z_t}\right)}{\log\left(\frac{z_{t+\tau}+\omega}{z_t+\omega}\right)}.$$

Since e_z is increasing in both α and ω, an increase in α and an increase in ω are competing explanations for the magnitude effect on intertemporal substitutability. If subjects take into account more background consumption as the total budget increases, we would observe a greater sensitivity to the interest rate, i.e., a greater e_z. When we assume a fixed background consumption, however, the pattern will be attributed to a magnitude effect on α.

Both the mental-accounting Fudenberg-Levine model and the mental zooming theory point to partial integration with lifelong wealth, which seems to be an important mechanism of the magnitude effect on intertemporal substitutability. Andersen et al. (2018) showed empirically that subjects only partially integrate experimental rewards with wealth in risk preference tasks. While they provide evidence of partial asset integration by exploiting variation in personal wealth, we provide within-subject evidence suggesting that the degree of asset integration is increasing in the stake.

None of the current models can explain both a magnitude effect on the discount factor and a magnitude effect on the intertemporal substitutability. Of course, the two channels can be explained by a mode-switching model in which individuals are assumed to have different preferences for different stakes. However, a truly unified explanation is still lacking.

7 Conclusion

Our study investigates the magnitude effect on intertemporal choices in a setting that is more general than a single-reward task, namely the intertemporal allocation task. After adapting the definition of the magnitude effect to the new task, we verify its existence: people allocate a larger share of the budget to the later reward as the total budget increases. The magnitude effect is not affected by whether the sooner reward is immediate or in the future. The size of the magnitude effect is smaller when the stakes are higher.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

We then look deeper into the effect, by exploring the channels. The results underscore the importance of a dimension that is often overlooked, namely, the intertemporal substitutability. We find evidence that both discount factor and intertemporal substitutability change with the magnitude of rewards. The magnitude effect on intertemporal substitutability is not driven by the magnitude effect on risk attitudes, since the two magnitude effects are in the opposite directions.

Some theories may provide explanations for one of the two channels. A cost-of-delay model (Benhabib et al., 2010) or a magnitude-dependent discounting model (Noor, 2011; Baucells & Heukamp, 2012) can account for a magnitude effect on the discount factor. Models which allow the degree of asset integration (mental accounting) to vary with the size of the budget can explain a magnitude effect on intertemporal substitutability. However, a new theory would be needed to account for both channels simultaneously and in a unified way.

Appendix

Appendix A. Rationality of subjects in the Convex Time Budget method

The CTB method allows subjects to make interior choices and hence makes it possible to measure discount rates and utility curvature simultaneously. However, Chakraborty et al. (2017) found that a proportion of subjects, especially those who make interior choices, violate wealth monotonicity in the CTB dataset of Andreoni and Sprenger (2012) and that the magnitude of wealth monotonicity violations conditional on violating at least once is as large as that generated by uniform random choice, and hence questioned the rationality of subjects in making CTB decisions.

In this appendix, we follow Chakraborty et al. (2017) to examine price monotonicity and wealth monotonicity of our dataset. In specific, we look at fractions of monotonicity violations among all subjects and among subjects who make at least one interior choice, respectively. We also measure the average magnitude of wealth monotonicity violations for those who violate wealth monotonicity at least once, and we compare it with the distribution of the magnitude generated by uniform random choice.

One advantage of our sample in testing wealth monotonicity is that we have 70 pairs of tasks that only differ in total budget for each subject, compared to one pair in Andreoni and Sprenger (2012). This endows our test with greater power than that in Chakraborty et al. (2017).

Table 9 shows the rationality indices of the full sample and the subsamples as well as those generated by uniform random choice. The fractions of price monotonicity violations (2–3%) are less than the fractions of wealth monotonicity violations (10–20%). A possible reason is that prices are varied within decision forms but total budgets are varied across decision forms, and hence it is easier to make comparisons across prices than across total budgets. Nevertheless, those fractions and the magnitudes of violations are much lower than those generated by uniform random choice, suggesting that our dataset and the subsample of subjects who make interior choices are highly informative.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Table 9 Rationality of subjects compared to uniform random choice

	Fraction of price monotonicity violations	Average magnitude of price monotonicity violations (euros)	Fraction of wealth monotonicity violations	Average magnitude of wealth monotonicity violations (euros)
Full sample	0.022	0.13	0.101	0.82
Subjects who make at least one interior choice	0.037	0.23	0.155	1.03
Subjects who violate wealth monotonicity at least once	0.033	0.26	0.208	1.67
Mean of uniform random choice (Standard deviation)	0.429 (0.006)	8.12 (0.21)	0.297 (0.008)	2.68 (0.10)

The first three rows present fractions of price monotonicity violations and wealth monotonicity violations as well as average magnitudes of violations in terms of euros in the full sample, the subsample of subjects who make at least one interior choice, and the subsample of subjects who violate wealth monotonicity at least once. The last row presents the means and the standard deviations of the same four indices generated by uniform random choice. The means and standard errors are calculated by simulating 10,000 times

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10683-021-09723-w.

Acknowledgements We are grateful for the helpful discussions with Arthur van Soest, and the insightful comments of Mohammed Abdellaoui, Elena Cettolin, Stephen Cheung, Mark Dean, Sebastian Ebert, Glenn Harrison, Gijs van de Kuilen, Guy Mayraz, Bin Miao, Wieland Müller, Kirsten Rohde, Karl Schlag, Stefan Trautmann, Yuehui Wang, Yuxin Yao, Songfa Zhong, and participants at the Tilburg mEETing, the Xiamen Workshop on Experimental Economics, the ENTER Jamboree, the Tilburg Economics Workshop, the MBEES Maastricht, the EEA Annual Congress, the ESA European meeting, and the IMEBESS in Rome. We thank the Netherlands Organization of Scientific Research (NWO) for financial support. Chen Sun further acknowledges financial support from the German Science Foundation (Deutsche Forschungsgemeinschaft) through the project CRC TRR 190.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abdellaoui, M., Bleichrodt, H., & l'Haridon, O. (2013a). Sign-dependence in intertemporal choice. *Journal of Risk and Uncertainty, 47*(3), 225–253.

Abdellaoui, M., Bleichrodt, H., l'Haridon, O., & Paraschiv, C. (2013b). Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time. *Management Science, 59*(9), 2153–2169.

Andersen, S., Cox, J. C., Harrison, G. W., Lau, M. I., Rutström, E. E., & Sadiraj, V. (2018). Asset integration and attitudes toward risk: Theory and evidence. *Review of Economics and Statistics, 100*(5), 816–830.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2013). Discounting behaviour and the magnitude effect: Evidence from a field experiment in Denmark. *Economica, 80*(320), 670–697.

Andreoni, J., Kuhn, M. A., & Sprenger, C. (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior and Organization, 116*, 451–464.

Andreoni, J., & Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review, 102*(7), 3333–3356.

Augenblick, N., Niederle, M., & Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics, 130*(3), 1067–1115.

Baucells, M., & Heukamp, F. H. (2012). Probability and time trade-off. *Management Science, 58*(4), 831–842.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-Bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior, 69*(2), 205–223.

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural India. *The Economic Journal, 91*(364), 867–890.

Bommier, A. (2007). Risk aversion, intertemporal elasticity of substitution and correlation aversion. *Economics Bulletin, 4*(29), 1–8.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Bouchouicha, R., & Vieider, F. M. (2017). Accommodating stake effects under prospect theory. *Journal of Risk and Uncertainty, 55*(1), 1–28.
Chakraborty, A., Calford, E. M., Fenig, G., & Halevy, Y. (2017). External and internal consistency of choices made in convex time budgets. *Experimental Economics, 20*(3), 687–706.
Cheung, S. L. (2015). Comment on "Risk preferences are not time preferences": On the elicitation of time preference under conditions of risk. *American Economic Review, 105*(7), 2242–2260.
Cheung, S. L. (2020). Eliciting utility curvature in time preference. *Experimental Economics, 23*(2), 493–525.
Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2017). The robustness and pervasiveness of sub-additivity in intertemporal choice. Briq institute. https://www.briq-institute.org/wc/files/people/armin-falk/working-papers/the-robustness-and-pervasiveness-of-sub-additivity-in-intertemporal-choice.pdf.
Ebert, S., & van de Kuilen, G. (2016) Measuring Multivariate Risk Preferences. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2637964.
Epper, T. (2015). Income expectations, limited liquidity, and anomalies in intertemporal choice. *Economics Working Paper Series* 1519, University of St. Gallen, School of Economics and Political Science.http://ux-tauri.unisg.ch/RePEc/usg/econwp/EWP-1519.pdf
Epstein, L. G., & Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica, 57*(4), 937.
Fehr-Duda, H., Bruhin, A., Epper, T., & Schubert, R. (2010). Rationality on the rise: Why relative risk aversion increases with stake size. *Journal of Risk and Uncertainty, 40*(2), 147–180.
Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351–401.
Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *American Economic Review, 96*(5), 1449–1476.
Halevy, Y. (2015). Time consistency: Stationarity and time invariance. *Econometrica, 83*(1), 335–352.
Harrison, G. W., Johnson, E., McInnes, M. M., & Elisabet Rutström, E. (2005). Risk aversion and incentive effects: Comment. *American Economic Review, 95*(3), 897–901.
Holcomb, J. H., & Nelson, P. S. (1992). Another experimental look at individual time preference. *Rationality and Society, 4*(2), 199–220.
Holden, S. T., & Quiggin, J. (2017). Bounded awareness and anomalies in intertemporal choice: Zooming in google earth as both metaphor and model. *Journal of Risk and Uncertainty, 54*(1), 15–35.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655.
Kachelmeier, S. J., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review, 82*(5), 1120–1141.
Kihlstrom, R. E., & Mirman, L. J. (1974). Risk aversion with many commodities. *Journal of Economic Theory, 8*(3), 361–388.
Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General, 126*(1), 54–70.
Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*(1), 78–87.
Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica, 46*(1), 185.
Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics, 107*(2), 574–597.
Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic Perspectives, 3*(4), 181–193.
Meier, S., & Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics, 2*(1), 193–210.
Miao, B., & Zhong, S. (2015). Comment on "Risk preferences are not time preferences": Separating risk and time preference. *American Economic Review, 105*(7), 2272–2286.
Noor, J. (2011). Intertemporal choice and the magnitude effect. *Games and Economic Behavior, 72*(1), 255–270.

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press

Oja, H., & Randles, R. H. (2004). Multivariate nonparametric tests. *Statistical Science, 19*(4), 598–605.
Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica, 32*(1/2), 122.
Read, D. (2001). Is time discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23*(1), 5–32.
Richard, S. F. (1975). Multivariate risk aversion, utility independence and separable utility functions. *Management Science, 22*(1), 12–21.
Sutter, M., Kocher, M. G., Daniela, G. R., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review, 103*(1), 510–531.
Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters, 8*(3), 201–207.
Weil, P. (1990). Nonexpected utility in macroeconomics. *Quarterly Journal of Economics, 105*(1), 29–42.

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

https://doi.org/10.1007/s10683-021-09723-w Published online by Cambridge University Press