ISSN (Print) 0473-453X
Discussion Paper No. 1230
ISSN (Online) 2435-0982

OUTCOME- AND SIGN-DEPENDENT TIME PREFERENCES: AN INCENTIVIZED INTERTEMPORAL CHOICE EXPERIMENT INVOLVING EFFORT AND MONEY

Shohei Yamamoto
Shotaro Shiba
Nobuyuki Hanaki

Secondly Revised October 2026
Revised June 2025
February 2024

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

OUTCOME- AND SIGN-DEPENDENT TIME PREFERENCES: AN INCENTIVIZED INTERTEMPORAL CHOICE EXPERIMENT INVOLVING EFFORT AND MONEY

Shohei Yamamoto[1,*], Shotaro Shiba[2], Nobuyuki Hanaki[3,4]

[1]College of Economics, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan. Email: shohei.yamamoto@rikkyo.ac.jp

[2]Faculty of Economics, Toyo University, 5-28-20 Hakusan, Bunkyo, Tokyo 112-8606, Japan. Email: sibasyou2001@gmail.com

[3]Institute for Social and Economic Research, the University of Osaka, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. Email: nobuyuki.hanaki@iser.osaka-u.ac.jp

[4] University of Limassol, Cyprus

* Corresponding author

OUTCOME- AND SIGN-DEPENDENT TIME PREFERENCES: AN INCENTIVIZED INTERTEMPORAL CHOICE EXPERIMENT INVOLVING EFFORT AND MONEY[†]

Abstract

Previous research consistently identified differences in time preferences between effort and monetary decisions. However, the root cause of this difference—whether it stemmed from the intrinsic nature of the outcomes or the associated pleasurable or unpleasurable experiences—remained undefined. In response, we conducted novel two-stage experiments employing a 2 $\times$ 2 design contrasting outcomes (money and effort) and domains (gain and loss). This approach allowed for the incentivization of all decisions, including those involving future monetary losses. Our study reveals that present bias is absent at the median for both monetary and effort-based choices, whereas the size of individual deviations from time-consistent choice, in either direction, is significantly larger for effort, indicating outcome-dependent preferences. Effort choices are also smoother than monetary choices and far more heterogeneous across individuals in their time preferences. (JEL: C91, D91)

Keywords: time preferences, losses, incentivized experiment

[†]This research has benefited from the financial support of (a) the Joint Usage/Research Center, the Institute of Social and Economic Research (ISER), Osaka University, and (b) Grants-in-aid for Scientific Research Nos. 18K19954, 20H05631, and 23H00055 from the Japan Society for the Promotion of Science. The experiment reported in this paper has been approved by the IRB of ISER, Osaka University (No. 20221103). Our two experiments were preregistered at AsPredicted.org (#113813 & #205027).

1. Introduction

Time preferences have long been central to the study of individual decision-making, with decades of research exploring how people evaluate outcomes across time (Abdellaoui et al., 2013; Andreoni & Sprenger, 2012; Frederick et al., 2002; Laibson, 1997; Thaler, 1981). A core contribution of this literature is the distinction between impatience (a long-run discount factor) and present bias (overvaluation of present outcomes) captured in hyperbolic discounting models.

Empirical work has primarily focused on time preferences on monetary outcomes (e.g., Ashraf et al., 2006; Dohmen et al., 2010; Tanaka et al., 2010), which have been linked to behaviors such as credit card debt (Meier & Sprenger, 2010), body mass index (BMI) (Courtemanche et al., 2015), and impulse control disorders (Brewer & Potenza, 2008; Ida, 2014; Odum et al., 2002).

However, existing studies have also identified more nuanced features: time preferences are *sign-dependent* (Benzion et al., 1989; Thaler, 1981) and *outcome-dependent* (Odum et al., 2006; Reuben et al., 2010). These findings highlight the need to move beyond pleasant, monetary outcomes and examine non-monetary and aversive domains to fully understand intertemporal decision-making. Moreover, researchers have discussed how the fungibility of money and the delay between its receipt and consumption may affect time preferences (Cubitt & Read, 2007; O'Donoghue & Rabin, 2015). Consequently, recent studies (e.g., Augenblick et al., 2015; Augenblick & Rabin, 2019; Carvalho et al., 2016) have begun to examine effort-based choices because the time dedicated to specific tasks may serve as a more immediate reflection of consumption (Meissner & Pfeiffer, 2022). These studies show more pronounced present bias in effort choices than in monetary choices (see Imai et al., 2021, for a meta-analysis). However, monetary rewards and effort exertion are inherently different in types of outcomes and the associated experiences, whether gain or loss. Consequently, the observed differences in present bias could be influenced by the type or sign of the outcomes, or both.

This study investigates both outcome- and sign-dependence of time preferences through two incentivized experiments using a 2×2 design (money vs. effort; gain vs. loss). Building on Bosch-Domènech and Silvestre (2010), our method allows for the incentivization of future monetary losses—an area often studied using hypothetical scenarios (Thaler, 1981; Abdellaoui et al., 2013; Shiba & Shimizu, 2020). While few prior studies have implemented real-loss incentives (e.g., Gutierrez et al., 2024; Krause & Schmidt, 2023), our approach differs in that it is designed to minimize both *hypothetical bias* and the *house money effect*—the distortion of preferences that can arise when provided endowments are perceived as windfall money (Frederick et al., 2002; Thaler & Johnson, 1990).

In our experiments, participants made intertemporal choices involving either monetary gains/losses or effort gains (task exemptions) and effort losses (task additions). In the Main Experiment, participants were randomly assigned to one of the four conditions (between-subjects). In the Within-Subject Experiment, each participant was assigned to either the money or effort condition and completed both gain and loss domains (i.e., within-subjects for domains while between-subjects for outcome types). All monetary transactions were conducted via mobile payment to ensure timely and smooth implementation as in Andreoni et al. (2018).

This study contributes to the literature by providing a rigorous, fully incentivized test of sign- and outcome-dependence in time preferences. First, in contrast to the meta-analytic evidence on effort choices (Imai et al., 2021), neither experiment found economically meaningful present bias: the median present-bias parameter is approximately 1 in every condition, and the only economically visible deviation lies above 1, on the future-bias side. Second, we found evidence of outcome dependence. While neither outcome showed present bias — where β differed between outcomes, the effort median lay above 1 — time inconsistency (the size of deviations from time-consistent choice in either direction) was significantly higher for effort. Furthermore, effort-based choices tended to favor balanced allocations (i.e., demonstrated a preference for smoothness) over monetary choices. Third, sign dependence in impatience appeared in the Within-Subject Experiment in both monetary and effort choices, though in opposite directions in the *Impatience* index, but not to any economically

meaningful degree in the Main Experiment. Although the sign effect has been widely reported, this result points to a potentially modest role of sign dependence. We further explore this in the Discussion section.

The rest of the paper is organized as follows. The relevant literature is reviewed and our research questions are derived from it in Section 2. The design of the two experiments is described in Section 3, followed by the results in Section 4. Section 5 discusses the relationship between time preference parameters and everyday economic behaviors. Section 6 summarizes the findings and offers some discussion.

2. Literature Review and Research Questions

2.1. Present Bias and Impatience

Present bias and patience have become central concepts in behavioral economics, particularly through the quasi-hyperbolic discounting model (Laibson, 1997; Frederick et al., 2002). Many early studies relying on binary choice or matching tasks found present bias in monetary decisions. However, these studies typically assume linear utility and rely on cross-sectional designs, raising concerns about measurement validity (Andreoni & Sprenger, 2012; Halevy, 2015).

The Convex Time Budget (CTB) method developed by Andreoni and Sprenger (2012) has become one of the most widely used methods for measuring time preferences in incentivized experiments. Each CTB decision is an allocation problem: the participant divides an experimental budget between a sooner and a later date, choosing an allocation on a budget line whose interest rate varies across decisions. Because it observes how allocations respond to these varying interest rates, the CTB allows utility curvature to be estimated jointly with time preferences, which addresses the linear-utility concern noted above. Moreover, when the same allocation problems are posed again once the sooner date has arrived, any change in a participant's own allocations reveals time inconsistency directly (Halevy, 2015).

According to the meta-analysis of Imai et al. (2021), studies that employ the CTB often fail to detect present bias in monetary contexts, whereas present bias is found more often when the intertemporal object is not money: Augenblick et al. (2015) report strong present bias for effort but not for money in the same participants, and the meta-analytic effort estimates point the same way. Such evidence has been read as indicating that non-monetary outcomes such as effort are more diagnostic of underlying intertemporal preferences. This empirical pattern, rather than any particular explanation of it, motivates our focus on effort.

Money typically represents a gain or a consumption resource, whereas effort is experienced as disutility or a loss. This distinction suggests that time preferences may vary not only by outcome type (effort vs. money) but also by domain (gain vs. loss). Prior research—most of it based on elicitation methods other than the CTB—has indeed reported variation along both dimensions. However, these studies typically explore only one dimension (outcome or sign), especially in incentivized settings. Thus, the joint effect of outcome- and sign-dependence on present bias and impatience remains unclear. We review studies employing a range of methodologies because, to the best of our knowledge, no CTB-based studies have examined both outcome-dependent and sign-dependent present bias.

2.1.1. Sign-dependence. Earlier studies in monetary contexts report higher discounting for gains than for losses—known as the *sign effect* (Benzion et al., 1989; Myerson et al., 2017; Thaler, 1981). Some studies even find negative discounting for future losses (Hardisty & Weber, 2009; Yoon & Chapman, 2016). Non-monetary outcomes such as health, pain, consumer goods and effort show sign effects as well, with magnitudes that appear to vary across contexts (Baker et al., 2003; Chapman, 1996; Hardisty & Weber, 2009; Yamamoto & Navarro-Martinez, 2022).

Research on sign-dependent present bias is more limited and inconsistent. Thaler (1981) found present bias in monetary gains but not in losses, Shiba and Shimizu (2020) in both, and Abdellaoui et al. (2013) found it stronger in losses than in gains. The only incentivized study to directly compare

present bias between gains and losses is Andreoni et al. (2018), which subtracted losses from a provided endowment. They found present bias higher in the loss domain and nearly absent for gains. For effort, to the best of our knowledge, Abdellaoui et al. (2018) is the only comparison of present bias between gains and losses of working time, and they found similar levels in both domains.

2.1.2. Outcome-dependence. In the gain domain, discount rates for primary rewards such as food and time are generally higher than those for money (Cheung et al., 2022; Estle et al., 2007; Odum et al., 2006; Reuben et al., 2010; Tsukayama & Duckworth, 2010), suggesting that individuals tend to be more impatient when evaluating non-monetary outcomes than monetary ones. However, patience across monetary and non-monetary outcomes, such as consumer goods (Tsukayama and Duckworth, 2010) and working time (Abdellaoui et al., 2018), appears to be positively correlated.

In the loss domain, direct comparisons between monetary and non-monetary outcomes are limited, but existing evidence suggests distinct patterns. Harris (2012) found that while individuals often postponed monetary losses, responses to losses involving pain, embarrassment, or social rejection varied substantially. Experimental studies also show slower or even negative discounting for health and pain-related losses (Baker et al., 2003; Ganiats et al., 2000; Loewenstein, 1987). These findings imply that non-monetary losses, including effort tasks, may involve more complex time preferences.

Research also indicates that present bias varies by outcome type. Cheung et al. (2022) reported weaker present bias for food than for money, while Bleichrodt et al. (2016) observed stronger present bias for health outcomes than for monetary ones. Some studies have found greater present bias for effort than for monetary gains (Augenblick et al., 2015; Augenblick & Rabin, 2019). Abdellaoui et al. (2018) is the only study directly comparing present bias for monetary gains and gains of leisure time, finding greater bias for the latter. Notably, they also found no significant correlation between present bias in the two outcome types. These comparisons, however, have so far been conducted almost entirely in the gain domain. Present bias in the loss domain—especially effort versus monetary losses—remains largely unexplored.

These literatures leave two gaps: outcome type and sign have rarely been varied jointly within a single incentivized design, and the evidence is thinnest in the loss domain, above all for effort losses versus monetary losses. These gaps leave open the following research questions: (1) whether the degree of present bias — or, more broadly, of time inconsistency in either direction — differs between effort exertion choices and monetary choices; (2) whether the degree of impatience differs between gains and losses in both monetary and effort choices; (3) whether the sign effect differs between effort and monetary outcomes; (4) whether the level of impatience differs between effort and monetary outcomes in both gain and loss domains. We have preregistered (AsPredicted #113813 and #205027) the experimental design and the analyses but no directional hypotheses; accordingly, we pose these as open questions rather than as ex-ante directional hypotheses and present the corresponding results descriptively in Section 4.

2.2. Hypothetical Intertemporal Decisions

To address these questions, incentivized measurement is crucial for avoiding unwanted bias. Much of the existing literature—especially in the loss domain—relies on hypothetical scenarios, which may introduce hypothetical bias (Coller & Williams, 1999; Frederick et al., 2002). The evidence of such bias is mixed for monetary gains (Coller & Williams, 1999; Brañas-Garza et al., 2023; Matousek et al., 2022), and it may be more problematic for monetary losses. Common methods, such as providing initial endowments (e.g., Gutierrez et al., 2024), may trigger the house-money effect (Thaler & Johnson, 1990). This suggests that established findings, such as the sign effect, warrant re-examination under incentivized conditions.

For example, Krause & Schmidt (2023) compared a Hypothetical treatment with no real stakes, an Opportunity Loss treatment with losses deducted from a provided endowment, and a Debit treatment in which participants received a transfer in advance and signed a direct debit mandate for real losses. The Debit treatment produced smaller differences in discount rates between gains and

losses and lower overall discounting, suggesting that perceived real losses reduce bias and that traditional methods may overstate sign effects. Hypothetical bias itself may also differ between gains and losses (Etchart-Vincent & l'Haridon, 2011; Weber et al., 2004), which can compromise estimates of sign-dependent preferences.

To elicit underlying preferences across both gains and losses, we take a step further than Krause and Schmidt (2023). We aim to mitigate unwanted bias, particularly the house-money effect, by adopting a method inspired by Bosch-Domènech and Silvestre (2010). They did not simply endow participants with money but had them earn their endowment via unrelated tasks before the main experiment. Their post-experiment surveys suggested this income was less likely to be treated as windfall. Accordingly, we employ a two-stage design: participants first earn money to cover potential losses in an initial online session, and then, on a separate day, make incentivized intertemporal decisions in the main session.

3. Experimental Design

We conducted two experiments to examine the outcome- and sign-dependence of time preferences. Both experiments were conducted online rather than in a physical laboratory: participants took part remotely on their own computers, and all payments were delivered through PayPay, one of the most widely used mobile payment services in Japan. Our primary experiment, which we call the Main Experiment, used a fully between-subjects 2×2 design, where participants were randomly assigned to one of four conditions: Money-gain, Money-loss, Effort-gain, or Effort-loss.

In the other experiment, which we call the Within-Subject Experiment, participants were randomly assigned to either a money or effort condition and completed tasks involving both gains and losses. In this setup, the outcome type (money vs. effort) varied across participants, while the domain (gain vs. loss) varied within each participant. A key advantage of this structure is that it enables individual-level analysis of sign effects. Earlier within-subject investigations of sign effects, such as Benzion et al. (1989), Chapman (1996), and Ikeda et al. (2010), relied on hypothetical scenarios; the Within-Subject Experiment revisits this structure with incentivized decisions. The fully between-subjects design of the Main Experiment enabled clearer and more robust comparisons between outcome types and domains under a unified experimental framework, addressing concerns about comparability in the Within-Subject Experiment.

The Within-Subject Experiment was conducted first, beginning in November 2022; the Main Experiment began in December 2024. Because the two experiments share most of their procedure, Section 3.1 describes the Main Experiment in full (Sections 3.1.1–3.1.2) and the Within-Subject Experiment through its differences (Section 3.1.3).

In both experiments, participation required having a PayPay account and having made a transaction with it within the previous month; agreeing to participate in all sessions of the experiment lasting five weeks; and having a stable internet connection and a computer with which to take part. Participation was restricted to computers: in both experiments, every survey screened out mobile devices at its opening branch. To ensure smooth and reliable money transactions, we delegated all payment operations to Dosanite (https://dosanite.co.jp/), a third-party company. This arrangement helped establish trust in the execution of the transactions and encouraged participants to take their decisions seriously. Rewards varied depending on participants' decisions and the treatment to which they were assigned, with the average reward being 2,000 yen—roughly equivalent to two hours of work for university students. Besides that, they received their preferred item, which was valued at around 5,000 yen in the Within-Subject Experiment or up to that amount in the Main Experiment, if they completed all sessions in the experiment.[1] The substantial completion reward played a crucial

[1]On the first day of the experiment, we presented participants with a list of 30 and 21 attractive items in the Within-Subject Experiment and the Main Experiment, respectively. Participants were asked to rank these items in order of preference. The items were then distributed according to these rankings. A week after the last session of the experiment, the items were dispatched to the participants' designated locations. We opted for tangible items rather than money as com-

role in minimizing attrition, which was essential to preserving sample integrity and ensuring valid inference.

3.1. Design and Procedure

The experiment lasted 5 weeks, and participants needed to complete all sessions in Week 1 (the start of the experiment), Week 3 (two weeks later), Week 4 (three weeks later), and Week 5 (four weeks later). In Week 5, all participants answered a set of questions related to their cognitive skills[2] and behaviors potentially correlated with their time preferences,[3] as well as demographics such as age, gender, and student status (undergraduate or graduate). An attention check was included in each of the Week-3 and Week-4 surveys to exclude inattentive participants.

Figure 1 shows the schedule over five weeks (for example screens of the allocation decisions, see Online Appendix OA.11; the complete instructions are reproduced in the supplementary file Experimental Instructions, Sections EI.1–EI.16). In both experiments, participants in the Money treatment made intertemporal allocation decisions involving money, while those in the Effort treatment made similar decisions involving effort.

Insert Figure 1 about here.

3.1.1. Money Treatment. The participants completed a task in Week 1 involving the transcription of 4,000 numbers displayed on screen. Each page showed ten blurred numbers (see Figure 2 for an example). Thus, participants needed to complete 400 pages, and they received a reward of 2,000 yen for doing so. This reward was designed to offset any potential losses in subsequent sessions. The participants were informed that the reward was the participants' own payment and could be used freely. To ensure that participants understood the payment procedure and rules, they needed to correctly answer the comprehension question before proceeding.

Insert Figure 2 about here.

In each of Weeks 3 and 4 (the second and third sessions), participants made a total of 18 intertemporal allocation choices, answering two sets of nine allocation questions in the domain of their assigned condition: participants in the Money-gain condition made decisions exclusively involving monetary gains, whereas those in the Money-loss condition dealt only with monetary losses. In Set 1, participants faced allocation problems with a budget constraint,

$$(1 + R)x_s + x_l = B$$

where x_s and x_l are the amounts of outcome and B is a (future-value) budget. The set of questions is constructed with nine different interest rates: R$\in\{-0.1, -0.06, -0.023, 0, 0.037, 0.083, 0.143, 0.22, 0.33\}$ with a budget of $B = 1500$. Set 2 employed a budget constraint of

pletion rewards to minimize the likelihood of participants integrating them with the monetary rewards obtained during the experiment.

[2]Cognitive skills were measured by four matrix reasoning items and four three-dimensional rotation items from the International Cognitive Ability Resource (Condon & Revelle, 2014). In the three-dimensional rotation items, cubes were displayed at the top of the screen. Participants were tasked with selecting, from the given cube options, the one that could match the displayed cube when rotated. In the matrix reasoning items, a matrix featuring different geometric patterns—with one part missing—was displayed at the top of the screen. Participants were tasked with selecting the most fitting geometric pattern from the given options to complete the missing part.

[3]Namely, questions about their drinking habits (Bradford et al., 2017): how many days they drink alcohol per month; how many times they drink large quantities per month (five or more drinks for men and four or more drinks for women), and how many glasses of alcohol they drink on average. We also asked other questions including how many times they played sport this year (did not exercise, less than once a month, 1–3 days per month, more than once a week, more than twice a week, more than three times a week, more than five times a week); if they feel they are usually healthy (I think very healthy, I think healthy, Not so healthy, Not healthy); their height and weight; their GPA; and the saving proportion of their monthly income (in the Within-Subject Experiment, of the monthly salary). In addition, we asked about their eating and shopping habits from the Shorter PROMIS Questionnaire (Christo et al., 2003).

$$x_s + \frac{1}{1+R} x_l = B$$

with parameters R$\in\{-0.11, -0.07, -0.03, 0, 0.023, 0.083, 0.127, 0.2, 0.3\}$ and $B = 1200$.

We based the format of our CTB questions on the implementation in Cheung et al. (2022). Namely, participants chose one option from nine split bundles on the budget line as shown in Figure 3.[4] Furthermore, to maintain equal transaction costs between sooner and later dates, we excluded bundles where either the sooner or later allocation was zero (i.e., $x_s = 0$ or $x_l = 0$). The minimum values for x_s and x_l were set at 5 in both Sets 1 and 2. We did not explicitly show the interest rate to the participants.

Insert Figure 3 about here.

In the example questions shown in Figure 3, the payments on the sooner dates always decrease from the left to the right options (refer to Figures OA.11.1, OA.11.2, and OA.11.3 in Online Appendix OA.11 for example questions in the other conditions). To avoid the order effect, we created two versions of each set: in one, sooner-date payments decreased from left to right; in the other, the order was reversed. One of the two versions was randomly shown to the participants. In both experiments, the nine budget lines within each set were presented in a fixed order.

In Week 3, participants made allocation decisions for payments to be made one and two weeks later—i.e., Weeks 4 and 5—creating front-end-delay questions. In Week 4, they faced the same allocation problems between Weeks 4 and 5 again, but the sooner date now coincided with the decision day, creating a no front-end-delay setting. Payments due on the sooner date were scheduled for transfer through the app on that specific, named day, and effort due on the sooner date was to be performed on that specific, named day; the same holds for the later date one week afterwards. Payments and effort due on the decision day had to be completed by 23:59 that day.[5] Comparing the decisions in these two types of questions allows us to assess their level of present bias.

One question answered in Weeks 3 and 4 was then randomly selected, and the participant's decision in it was implemented. Participants in the Money-gain condition would receive x_s in Week 4 and x_l in Week 5; those in the Money-loss condition would pay x_s in Week 4 and x_l in Week 5.

Before making allocation decisions in Weeks 3 and 4, participants completed a comprehension quiz and could not proceed until answering it correctly.

3.1.2. Effort Treatment. In Week 1, participants agreed to transcribe 2,000 random numbers in both Weeks 4 and 5 and received 2,000 yen as compensation for doing so. As in the money treatment, they were informed that this compensation was their own payment, and they answered a comprehension question about the instructions. To reinforce the commitment, participants signed and returned a document detailing the contract. The purpose of this contractual agreement was to enhance the salience of the effort's reference point and encourage participants to internalize it. Without such a reference point, participants might integrate the 2,000 tasks into their intertemporal choices for effort-related tasks. As a result, their preferences could become unaffected by the gain and loss domains, which we aim to measure. We note that the absence of a corresponding device in the money treatment is deliberate rather than an asymmetry in the design: money may already carry a natural endowment reference point — the 2,000 yen the participant has earned and holds — in a way that effort does not, so the signed contract was intended to supply the effort task with a comparable reference point and to reduce any asymmetry in reference-point salience across the two outcome types. Imposing a symmetric but arbitrary monetary commitment device would have

[4] Augenblick et al. (2015), the first study to measure both money and effort with the CTB, implemented the method with a continuous slider task rather than discrete bundles; reanalyzing their data in a format analogous to our nine-bundle design yields results consistent with their core findings (available upon request).

[5] According to the payment vendor's records, every required transfer was completed without problems, and the survey timestamps show that, among the effort participants included in the analyses, every such effort task was completed on the decision day or within minutes after midnight, a median of 1.2 hours (Within-Subject Experiment) and 0.8 hours (Main Experiment) after the decision.

added artificiality without a clear benefit.[6] The intertemporal decisions in this treatment mirrored those of the money treatment. In each of Weeks 3 and 4, participants answered a total of 18 allocation questions regarding the amount of tasks in Weeks 4 and 5. Namely, participants transcribed $2000 + x_s$ and $2000 + x_l$ numbers in Weeks 4 and 5, respectively. The amount of numbers to transcribe was reduced in the Effort-gain condition (i.e., $x_s < 0$ and $x_l < 0$), while it was increased in the Effort-loss condition (i.e., $x_s > 0$ and $x_l > 0$).

One question was later randomly selected from all questions answered during Weeks 3 and 4 to determine the actual workload. The bundles of x_s and x_l, as well as the budget constraints (R and B), are identical between Money and Effort treatments, ensuring that the decisions are comparable.

3.1.3. The Within-Subject Experiment. The Within-Subject Experiment followed the same five-week procedure, with the same Week-1 tasks, and differed in two main respects. First, each participant completed both domains: in each of Weeks 3 and 4, participants answered a total of 36 allocation questions — two sets of nine each for gains and for losses — rather than 18; in the Money treatment they received x_s and x_l in the gain domain and paid them in the loss domain, and in the Effort treatment the amount of numbers to transcribe was reduced in the gain domain and increased in the loss domain. The order of the gain and loss domains was randomized, as was the left–right orientation of each set (Section 3.1.1). Both order randomizations were drawn independently at every level.[7] Second, the interest rates and budgets differed. The Within-Subject Experiment adopted the interest rate structure from Cheung et al. (2022): in Set 1, R$\in\{-0.3, -0.18, -0.07, 0, 0.11, 0.25, 0.43, 0.67, 1\}$ with $B = 1000$; in Set 2, R$\in\{-0.34, -0.21, -0.09, 0, 0.07, 0.25, 0.38, 0.6, 0.9\}$ with $B = 800$. Because Cheung et al.'s design used a three-week delay, whereas our experiments employed a one-week delay, these rates could have exaggerated impatience levels; for the Main Experiment we therefore recalibrated the interest rates and budgets to the values given in Section 3.1.1.

3.2. Estimation and Inference

Beyond the descriptive indices used for robustness in Section 4.5 (defined in Appendix A), we estimate time preferences using the quasi-hyperbolic discounting model (Laibson, 1997) combined with the constant relative risk aversion (CRRA) utility function to yield clearer differentiations among parameter values. Namely, we assume the utility from the alternative in the CTB questionnaire (x_s, x_l) is $U(x_s, x_l) = D(t_s)\,U(x_s) + D(t_l)\,U(x_l)$ where the discount function $D(.)$ and instantaneous utility $U(.)$ are $D(t) = \beta^{1_{t>0}}\delta^t$ and $U(x) = u(x)$ for $x > 0$ and $U(x) = -u(-x)$ for $x < 0$, respectively, where

$$u(x) = \begin{cases} \dfrac{x^{1-\alpha}}{1-\alpha} & (\alpha \neq 1) \\ \ln(x) & (\alpha = 1) \end{cases}.$$

This model allows us to jointly estimate three parameters: the weekly discount factor (δ), present bias (β), and utility curvature (α), using the maximum likelihood estimation (MLE) with the conditional (multinomial) logit probability of a specific choice

$$\Pr(choice) = \frac{e^{U^*/s}}{e^{U^1/s} + e^{U^2/s} + \ldots + e^{U^9/s}}$$

[6] The analyses reported in Online Appendix OA.7 are consistent with the contract functioning as intended.

[7] The gain and loss blocks were presented in a random, evenly weighted order determined between subjects — gain-first for 71 of the 165 participants in the Within-Subject Experiment and loss-first for the remaining 94 — and the left–right orientation of each set was drawn separately for the gain and the loss block; both randomizations were re-drawn independently in the Week-4 session. All eight observable agreement rates are consistent with independent draws. The two orientation draws agree for 55.8% of participants in Week 3 and 52.7% in Week 4; the Week-4 orientation agrees with the Week-3 orientation for 43.6% of participants in the gain block and 52.7% in the loss block; the Week-4 block order agrees with the Week-3 block order for 46.1% (48.2% for the corresponding orientation check in the Main Experiment); and the block-order and orientation draws are independent of one another (agreement of 54.5% and 46.7% in Week 3). All eight rates lie between 43.6% and 55.8%, close to the 50% expected under independent draws, and none approaches the 100% that fixing an order within subject would produce, or the 0% that systematic reversal would produce. One practical consequence is that roughly half of the participants faced a different block order in Week 4 than in Week 3; the order-dependence tests in Online Appendix OA.1 condition on the Week-3 block order.

where $U^1, U^2, \ldots, U^9$ are utilities of the nine alternatives in a choice set under δ, β, and α, and s is the scale of the stochastic-choice error: as s grows large, choice becomes noisier and the choice probabilities approach the uniform distribution over the nine alternatives, and as s approaches zero the model approaches deterministic maximization. U^* is the utility of the chosen alternative. We also calculate model-based time inconsistency as $\tau = |1-\beta|$, consistent with the model-free *TI* index (Appendix A).

We estimate the model by maximum likelihood at two levels. At the individual level, parameters are estimated separately for each participant and domain, subject to bounds following Cheung et al. (2022).[8] At the pooled level, reported in Appendix B, each parameter is allowed to shift with condition indicators.

For the pooled models we report delta-method standard errors; the table notes in Appendix B record which standard error applies to each specification. Individual-level estimates are summarized by medians and interquartile ranges and compared using nonparametric tests: the estimated β and δ are bounded, have medians at 1, and in the effort conditions carry mass at the estimation bounds, features for which Wilcoxon signed-rank and Mann–Whitney tests are appropriate, whereas the descriptive indices of Appendix A are approximately normal participant-level averages, for which t-tests are appropriate. Online Appendix OA.4 re-runs every comparison under a single test family in both directions; the direction and magnitude of each headline result are stable across families, while a minority of significance verdicts move with the family, concentrated in the bounded quasi-hyperbolic parameters.

---

4. Results

In both experiments, we recruited participants through ORSEE (Greiner, 2015) at the Institute of Social and Economic Research, the University of Osaka. In the Within-Subject Experiment, 200 participants signed up through ORSEE, and 184 of them completed the first session on November 22, 2022. Ultimately, 165 participants completed the experiment and were included in the analyses (35% female, age range: 18–36 years, with an average of 22.1 years). Of these 184, 18 did not complete the experiment appropriately: 15 in the effort conditions — eight who did not return after the first session and seven who did not complete the required effort tasks — and three in the money conditions who did not complete all sessions, none of whom refused a required payment (see Section 6.3); and one was excluded for failing the attention check. In the Main Experiment, a total of 325 participants completed the first session on December 23, 2024. Ultimately, 278 participants completed the experiment (46% female, age range: 18–43 years, with an average of 22.6 years). Of the 325 who completed the first session, 47 withdrew during the course of the experiment (12 in the Money-gain, 11 in the Money-loss, 14 in the Effort-gain, and 10 in the Effort-loss condition); according to the payment vendor's records, none of them refused a required payment (see Section 6.3), and no participant was excluded for failing the attention check. Online Appendix OA.2 documents attrition and selection in detail, including CONSORT flow diagrams for both experiments, a stage-by-stage breakdown, and analyses showing that attrition is unrelated to observable characteristics and early-stage behavior.

[8] In alignment with the methodology set forth by Cheung et al. (2022), we imposed bounds on the estimates for δ, β, and α within the ranges of 0.1 to 10, 0.05 to 20, and -10 to 10, respectively. The number of subjects whose estimated parameters approached the boundary (within a distance of 0.001) was as follows: in the Within-Subject Experiment, zero out of 89 for both gain and loss domains in the monetary choices. In the effort choices, the numbers were 14 and 12 out of 76 for the gain and loss domains, respectively. A further breakdown in the effort condition revealed that four and three subjects approached the boundary for δ, 11 and eight subjects for β, and zero and one subject for α in the gain and loss conditions, respectively. In the Main Experiment, the corresponding numbers were two out of 60 and two out of 75 for the gain and loss domains in the monetary choices, and nine out of 73 and 12 out of 70 for the gain and loss domains in the effort choices.

Table 1 outlines the demographic and other pertinent characteristics of the participants by condition in the Main Experiment, highlighting selected variables. Table OA.9.1 in the Online Appendix reports the same variables for the Within-Subject Experiment, together with a comprehensive list of additional variables for both experiments. None of the variables significantly differs between treatments (t-tests for the Within-Subject Experiment and Kruskal–Wallis tests for the Main Experiment; all $p > 0.1$). Both experiments drew from the same subject pool in different waves; Online Appendix OA.3 reports a Holm-corrected covariate comparison between the two experimental samples, together with a robustness check controlling for the attribute that differs after correction.

Insert Table 1 about here.

Throughout this section, we report the results of the Main Experiment and use the Within-Subject Experiment in two roles: as an internal replication and, because the same participants choose in both the gain and loss domains, as the source of individual-level cross-domain analyses that only a within-subject design can provide (Section 4.4). For each result we note whether the within-subject design shows the same pattern, and we report the Within-Subject Experiment in full, together with the within-design assumption checks, in Online Appendix OA.1. The section is organized around our three main findings. Because two terms that recur below are easily conflated, we use *present bias* only for a systematic overweighting of the present — β below 1 — and *time inconsistency* for any departure of β from 1, in either direction; our data will show the latter without the former. Section 4.1 documents the absence of economically meaningful present bias; Section 4.2 shows that time inconsistency is outcome-dependent; Section 4.3 shows that effort choices are smoothed and heterogeneous. Section 4.4 reports secondary results (sign-dependence and individual-level cross-domain correlations). Each finding rests primarily on the raw allocations (Figure 4) and the structural parameters (Table 3); Section 4.5 returns to the model-free indices (Table 2) and shows that they support the same three findings.

Table 2 reports the model-free indices (defined in Appendix A) by condition for the Main Experiment (Table OA.1.1 reports the Within-Subject Experiment, which Online Appendix OA.1 discusses), as a first look at the data before any model is imposed. We draw on these indices where a finding requires them and discuss them together in Section 4.5.

Insert Table 2 about here.

Figure 4 reports the mean proportions of allocations to sooner dates and their standard errors for each Main Experiment condition (Money-gain, Money-loss, Effort-gain, and Effort-loss).[9] The left and right panels show the results of Sets 1 and 2, respectively; the overall patterns between the two sets are similar, implying that preferences were not much affected by the two different budget sets. The proportion allocated to sooner dates decreases as R increases in the gain conditions, while the proportion allocated to later dates increases as R increases in the loss conditions, consistent with the law of demand. Panel (a) of Figure 4 shows the corresponding Within-Subject Experiment allocations, which follow the same patterns.

Insert Figure 4 about here.

As explained in Section 3.2, we use t-tests for the model-free indices and rank-based tests (sign-rank and rank-sum) for the bounded quasi-hyperbolic parameters; the direction and magnitude of every headline result are the same under both families (Online Appendix OA.4), and wherever a significance verdict is family-dependent we report both p-values.

[9]The proportion of the allocation is calculated by $\frac{1}{1+R}x_s^*/B$ for Set 1 and x_s^*/B for Set 2 where x_s^* is the amount allocated to the sooner date. Both equal 0 when a participant allocates everything to the later date and 1 when everything is allocated to the sooner date.

Our main analysis uses the structural parameters because they are estimated jointly with the utility-curvature parameter α, whereas the model-free indices may mix discounting with utility curvature — and Section 4.2 shows that curvature itself differs between money and effort (Section 5 applies the same reasoning to the choice of regressors). Their value as a robustness check lies precisely in their construction: computed directly from the choices, they do not depend on our parametric specification of discounting and utility, so a finding they reproduce does not hinge on that specification.

4.1. Finding 1: No Economically Meaningful Present Bias

Across all four conditions of the Main Experiment, the median present-bias parameter β is approximately 1, and no condition displays economically meaningful present bias.

The raw allocations already show this. In Figure 4, the solid lines (Week 3 allocations, with a front-end delay) and the dotted lines (Week 4 allocations, without a front-end delay) nearly coincide in every condition: removing the front-end delay did not shift allocations toward sooner gratification. The frequencies of present-biased, future-biased, and dynamically neutral participants are reported in Table OA.8.1; we note here only that non-neutral behavior is far more common in the effort conditions — fewer than 20% of effort observations are neutral, against more than 54% for money — and that it runs in *both* directions, a first sign of the pattern developed in Section 4.2.

Table 3 reports the individual-level quasi-hyperbolic parameter estimates for the Main Experiment (estimation details in Section 3.2): the median and interquartile range of δ, β, α, and $\tau = |1 - \beta|$ by condition, together with tests against the benchmark values $\delta = 1$, $\beta = 1$, $\alpha = 0$, and $\tau = 0$ (hereafter, the benchmark tests). The median β is 0.9998 (Money–gain), 0.9989 (Money–loss), 1.018 (Effort–gain), and 0.999 (Effort–loss). Sign-rank tests reject $\beta = 1$ in three cells (Money–gain $p = 0.006$; Money–loss $p = 0.007$; Effort–gain $p = 0.007$), while Effort–loss is not significant ($p = 0.470$; the t-test disagrees, $p = 0.004$). These rejections do not indicate present bias. In the money conditions the estimated deviations are of order 10^{-3} or smaller — the Money–gain interquartile range is 0.999–1.000 — and with heavy probability mass just below $\beta = 1$, the signed-rank statistic is driven by the sign pattern of these economically negligible deviations, so that tiny asymmetries reach conventional significance; under t-tests, neither money cell is anywhere near significant ($p = 0.370$ and $p = 0.816$). The only economically visible deviation, Effort–gain, lies *above* 1 — the future-bias side. These medians can be benchmarked against the literature. Effort studies that do detect present bias — designs in which participants allocate additional work, the counterpart of our effort–loss condition — report β well below 1: approximately 0.88 meta-analytically (Imai et al., 2021), 0.909–0.916 (Kölle and Wenner, 2023), 0.912 (Imas et al., 2022), and approximately 0.90 (Andreoni et al., 2023) — whereas all four of our condition medians (0.999–1.018) lie far closer to 1, the money medians by two orders of magnitude. The absence of present bias in our money conditions is likewise consistent with the monetary CTB literature (Andreoni and Sprenger, 2012). Median $\beta = 1$ with economically negligible deviations, not present bias, is therefore the reading that every cell supports.

Though Finding 1 concerns β, Table 3 also reports the long-run discount factor δ, which completes the benchmark picture. Discounting itself is essentially absent at the weekly horizon: the median δ is 1.000–1.001 in every condition. Because two of the four benchmark verdicts for δ are family-dependent (Online Appendix OA.4), we characterize δ by magnitudes rather than by significance: the money interquartile ranges are tight (0.970–1.000 in gain; 0.994–1.008 in loss), while the effort ones are wide (0.912–1.335; 0.737–1.508) — a preview of Finding 3.

The Within-Subject Experiment gives the same picture: the median β ranges from 1.000 to 1.002 across the four conditions, and $\beta = 1$ is not rejected at the 5% level in any condition under the sign-rank tests, though the two effort cells reject under t-tests (sign-rank $p = 0.527$ and $p = 0.088$ against t-test $p < 0.001$ and $p = 0.003$ in Effort–gain and Effort–loss, respectively; Online Appendix OA.1 and OA.4).

4.2. Finding 2: Time Inconsistency Is Outcome-Dependent

Our participants are not uniformly time-consistent — but their time inconsistency is two-sided, and it is concentrated almost entirely in effort. Because Section 4.1 shows that deviations from consistency run in both directions, we measure time inconsistency by a magnitude that does not cancel across individuals: $\tau = |1 - \beta|$.

The benchmark tests reject $\tau = 0$ in all four conditions under the rank family (Table 3; all $p < 0.001$) and in all but Money–gain under t-tests ($p = 0.204$), so individual departures from exact time consistency are detectable even for money; their size, however, differs sharply by outcome. The median τ is 0.0002 (Money–gain) and 0.0014 (Money–loss), against 0.249 (Effort–gain) and 0.201 (Effort–loss) — two to three orders of magnitude apart. Rank-sum tests of τ between outcomes reject equality in both domains ($p < 0.001$ in each; the t-tests agree, $p = 0.003$ and $p < 0.001$). By contrast, β itself differs between outcomes only in the gain domain (rank-sum $p = 0.012$, with the effort median above 1); in the loss domain the two medians coincide at 0.999 and the rank-sum test finds no difference ($p = 0.893$), although the t-test disagrees ($p = 0.002$), reflecting the wide effort dispersion documented in Section 4.3. This sensitivity of β-comparisons — and the fact that averaging β cancels two-sided deviations — is why we base Finding 2 on τ. The Money–gain versus Effort–loss comparison of Augenblick et al. (2015) points the same way: τ is higher in Effort–loss (rank-sum $p < 0.001$; the t-test agrees, $p = 0.002$), whereas β does not differ under the rank family (medians 0.9998 versus 0.999; rank-sum $p = 0.826$) — again the t-test disagrees ($p = 0.008$), reflecting the same wide effort dispersion. The model-free *TI* index gives the same picture (Section 4.5).

Finding 2 concerns time inconsistency, but the estimation behind Table 3 recovers utility curvature jointly with the time-preference parameters, so the same table reports α as well. Utility curvature is also outcome-dependent: α differs between money and effort in both domains (rank-sum, $p < 0.001$ in each). Within effort, the median α is positive in the gain domain (0.142; sign-rank $p < 0.001$) and negative in the loss domain (-0.788; $p < 0.001$) — the latter indicating concave utility for losses, consistent with Abdellaoui et al. (2013, 2018) — while the money medians are close to zero (0.002, $p = 0.073$, though the t-test disagrees, $p = 0.001$; -0.002, $p = 0.895$).

The Within-Subject Experiment shows the same qualitative pattern: τ is higher for effort than for money in both domains ($p < 0.001$ in each; Online Appendix OA.1).

4.3. Finding 3: Smoothing and Heterogeneity in Effort Choices

Effort allocations are interior and smoothed, and they differ widely across individuals; money allocations sit at the most extreme available bundles and are nearly homogeneous.

The allocation curves in Figure 4 make the first half of this point. The money curves approximate a step function at $(1 + R) = 1$: almost all money is allocated to the sooner date when the gross interest rate favors it and to the later date otherwise. The effort curves instead decline gradually over the whole range of R. This is the same money–effort contrast reported by Augenblick et al. (2015). The error bars in Figure 4 make the second half of the point: they are narrow for money and wide for effort, indicating far greater dispersion of behavior across individuals in the effort conditions.

The parameter estimates locate this heterogeneity in the time-preference parameters. In Table 3, the interquartile range of β spans 0.999–1.000 (gain) and 0.997–1.001 (loss) for money, against 0.961–1.432 and 0.840–1.338 for effort; for δ, the money ranges are 0.970–1.000 and 0.994–1.008, against 0.912–1.335 and 0.737–1.508 for effort. Brown–Forsythe tests reject equal dispersion between the Money and Effort treatments for β in both domains (gain $p = 0.003$; loss $p < 0.001$) and for δ in both domains (both $p < 0.001$); mean-centered Levene tests yield the same conclusions even more strongly (all four $p < 0.001$). By contrast, the dispersion of the curvature parameter α does not differ significantly between the treatments (Brown–Forsythe; gain $p = 0.364$, loss $p = 0.114$) — indeed, in the gain domain the money interquartile range for α is the wider one. The heterogeneity finding therefore concerns the time-preference parameters β and δ, not utility curvature. Full parameter distributions are shown in Figure OA.8.1.

Taken together, the three findings show that effort choices differ from money choices not in average present bias — no economically meaningful amount in either — but in their far stronger preference for smoothing and their pronounced individual heterogeneity in discounting behavior. The Within-Subject Experiment, with a separate sample and a within-subject design, reproduces all three patterns (Online Appendix OA.1); and the model-free indices and the structural parameters agree on the direction and magnitude of every headline result (Section 4.5 and Online Appendix OA.5).

4.4. Secondary Results: Sign-Dependence and Individual-Level Domain Correlations

Sign-dependence. In the Main Experiment, sign-dependence in the time-preference parameters appears only in the money treatment, and only at negligible magnitudes: δ is slightly higher in the loss domain (medians 1.000 versus 1.001; rank-sum $p = 0.003$), and τ is higher in the loss domain under the rank family (medians 0.0002 versus 0.0014; rank-sum $p < 0.001$) — though, as with the money $\beta = 1$ benchmarks in Section 4.1, the differences are economically negligible and the t-tests accordingly find nothing ($p = 0.241$ for δ and $p = 0.888$ for τ), the rank tests picking up tiny asymmetries. β differs at most marginally between domains (rank-sum $p = 0.076$). In the effort treatment, none of the time-preference parameters differs between domains (δ $p = 0.604$, β $p = 0.153$, τ $p = 0.698$); only curvature α does ($p < 0.001$), consistent with the sign-specific curvature pattern in Section 4.2. The model-free indices show no sign-dependence in either treatment (t-tests; Section 4.5). This contrasts with the Within-Subject Experiment, where the *Impatience* index showed a sign effect for money (greater impatience for gains, $p = 0.002$, although the sign-rank test does not reject, $p = 0.111$; Online Appendix OA.4) and a reversed one for effort ($p = 0.033$) — a divergence between the within- and between-subject designs to which we return in Section 6.2.

Individual-level cross-domain correlations (Within-Subject Experiment, Figure 5). Whether a person's time preference carries over from one domain to the other is itself outcome-dependent. In the money treatment, a participant's gain-domain and loss-domain parameters are nearly independent (pairwise Pearson correlations: $\rho = 0.10$ for δ; -0.02 for β): knowing how a person discounts monetary gains says almost nothing about how the same person discounts monetary losses. In the effort treatment, by contrast, the same individual is highly consistent across domains ($\rho = 0.53$ for δ; 0.69 for β), while the parameters themselves are far more dispersed across individuals. Figure 5 displays the corresponding index-level scatter plots: no clear cross-domain relationship for money except for *Smooth* ($\rho = 0.39$), strong positive relationships for effort. The cross-domain stability of time preference is thus not a property of the person alone but of the person–outcome pair.

Insert Figure 5 about here.

Insert Table 3 about here.

4.5. Robustness: The Model-Free Indices

The findings above rest on the raw allocations and the structural parameters. Here we return to the four model-free indices — simple choice-based measures constructed following Cheung et al. (2022), with *TI* added by us, defined in full in Appendix A and reported by condition in Table 2 and, for the Within-Subject Experiment, Table OA.1.1 — and their between-condition contrasts (Table 4), and show that they support the same three findings. *Impatience* is the mean share of each allocation assigned to the sooner date in the Week-3 session (to the later date for losses); *PB* is the change in that share from the Week-3 to the Week-4 session, when the front-end delay is removed; $TI = |PB|$ captures inconsistency in either direction; and *Smooth* measures how evenly allocations are split between the two dates.

In both experiments, the mean *Impatience* ranges from 0.39 to 0.44 in all conditions (Tables 2 and OA.1.1), indicating that our participants allocated a substantial amount to the later (sooner) dates when allocating gain (loss) outcomes. The median value of *Impatience* for monetary gains in Cheung et al. (2022) is 0.39, which is close to our Money-gain median in the Main Experiment (0.39; the mean in Table 2 is 0.41), indicating that our results in this condition are compatible with previous findings. The results across the two experiments are largely similar, except for a higher preference for smoothness in the Main Experiment, significant in three of the four conditions and most clearly for effort (Table OA.3.3).

Finding 1. At the index level, *PB* is not significantly different from zero in three of the four Main Experiment conditions (t-test, all $p \geq 0.120$); the exception is Effort–gain, where *PB* deviates marginally from zero (t-test $p = 0.073$; sign-rank $p = 0.031$) — in the *future*-bias direction, that is, allocations shifted away from the sooner date once the front-end delay was removed (the index means for the four conditions are in Table 2).

Finding 2. At the index level, *TI* is significantly positive in all four Main Experiment conditions (t-test, all $p < 0.001$). *TI* is larger for effort than for money in both domains (t-test, $p < 0.001$ in both; Columns (3)–(4) of Table 4, panel (b)), while average impatience does not differ between outcomes in either domain ($p \geq 0.163$). Comparing the Money–gain and Effort–loss conditions, as in Augenblick et al. (2015), gives the same conclusion: *TI* and *Smooth* are higher in Effort–loss (both $p < 0.001$), whereas *PB* does not differ ($p = 0.500$).

Finding 3. The *Smooth* index is higher for effort than for money in both domains (Columns (3)–(4) of Table 4, panel (b); both significant at the 1% level), matching the allocation curves discussed in Section 4.3.

Sign-dependence. The Main Experiment shows no sign-dependence in the descriptive indices: none of *Impatience*, *PB*, *TI*, or *Smooth* differs between the gain and loss domains in either treatment (t-tests, all $p \geq 0.209$; the single borderline exception under the rank family — *Smooth* in the money treatment, $p = 0.040$ — is listed in Online Appendix OA.4).[10]

Insert Table 4 about here.

The complete index comparisons under both test families are in Online Appendix OA.4, and Online Appendix OA.5 sets the index-based and parameter-based results side by side, effect by effect.

5. The Relationship Between Time Preferences and Economic Behaviors

We now turn to the relationship between the time-preference parameters measured in our experiments—for both money and effort, across the gain and loss domains—and behaviors outside the experimental setting, elicited through the questionnaire administered in Week 5. The preference parameters (Weeks 3 and 4) and these behaviors (Week 5) were measured within the same study wave; the associations reported in this section are therefore contemporaneous rather than predictive, a distinction we return to in Section 6.4.

Bartels et al. (2023) examine a wide range of behaviors and suggest their correlations with experimentally measured discount rates are frequently modest and sometimes even negligible. A plausible reason for this limited correlation is that Bartels et al. and many earlier studies measure time preferences predominantly with choices involving monetary gains. However, most real-world intertemporal choices encompass diverse outcomes and domains, for example, when choosing to exercise involves short-term effort for long-term health benefits.

[10]Note that while we employ paired t-tests in comparing gain and loss domains within each outcome in the Within-Subject Experiment because of the within-subject nature of the experiment, we employ two-sample t-tests for all the other comparisons in both experiments.

As our analyses have demonstrated, time preferences are dependent on outcome and sign to some extent and exhibit considerable heterogeneity at the individual level. Therefore, some behaviors may be correlated with time preferences only when outcome type and domain are taken into account.

We relate these behaviors to the structural parameters δ and β rather than to the model-free indices. We take the structural parameters as the main regressors because they are estimated jointly with the utility-curvature parameter α (Section 3.2), whereas the model-free indices may mix discounting with utility curvature—and Section 4.2 shows that curvature itself differs between money and effort. An association between an index and a behavior could therefore reflect either time preference or the curvature of utility, and the two cannot be separated. The indices nonetheless retain a virtue of their own: they are model-free and do not depend on our parametric specification of discounting and utility, a relevant consideration given the limitations of structural estimation under the CTB discussed in Section 6.1. Accordingly, our main analysis uses δ and β, and Online Appendix OA.10 reports the same regressions with the model-free indices as a robustness check. Of the three associations reported below, only the relationship between patience over effort gains and BMI is also detected with the indices; the GPA and binge-eating associations run in the same direction but are not statistically significant once the indices replace the structural parameters. Across all behaviors and conditions the two approaches agree in direction in 72.5% of cells and in the 5% verdict in 86.3%.

We estimate a separate regression for each behavioral outcome within each experimental condition. Continuous outcomes are analyzed with ordinary least squares and binary outcomes with logistic regressions. In every regression, the explanatory variables are the individual-level estimates of δ and β from the condition in question, with age, age squared, and dummy variables indicating female and graduate students as control variables. Table 5 presents, in standard regression format, the models corresponding to the three associations described below; the complete outputs for all behaviors and conditions—coefficients, standard errors, sample sizes, and R^2—are reported in Online Appendix OA.10, which also reports a variant that adds the magnitude of time inconsistency, $|1 - \beta|$, as a third regressor.

The Week-5 questionnaire covers behaviors that have previously been shown to correlate with experimentally measured time preferences—such as educational attainment (Duckworth & Seligman, 2005; Falk et al., 2018; Reed & Martens, 2011) and eating behaviors (Appelhans et al., 2012; McClelland et al., 2016)—as well as behaviors with mixed findings, such as BMI and obesity (Barlow et al., 2016; Cheung et al., 2022; Weller et al., 2008).

In the main text we discuss three associations (Table 5) that are statistically significant in the Main Experiment and carry the same signs in the Within-Subject Experiment; all estimates, significant or not, are reported in Online Appendix OA.10. First, δ in the Effort–gain condition is negatively correlated with BMI: participants with higher estimated patience over effort have lower BMI. Second, δ in the Effort–gain condition is positively correlated with grade point average (GPA), whereas δ in the money conditions is not. This pattern is not entirely aligned with previous results (e.g., Cheung et al., 2022), which found a positive correlation between patience in monetary choices and grades. Third, δ in the Money–loss condition is negatively correlated with binge eating, a negative association not found for the δ estimates from the other conditions. The Within-Subject Experiment shows the same three associations with the same signs (Online Appendix OA.10); the BMI and GPA associations are statistically significant there as well, whereas the binge-eating association is not.

These three behavioral measures—BMI, GPA, and binge eating—are thus significantly associated with time preferences only in specific conditions. We did not test whether the coefficients differ across conditions. Given the number of coefficients estimated, we regard these associations as exploratory. Results for the other behavioral variables are presented in Table OA.10.1, including behaviors previously linked to time preferences, such as alcohol use (Bradford et al., 2017; Sutter et al., 2013; Vuchinich & Simpson, 1998), savings (Angeletos et al., 2001; Bradford et al., 2017; Sutter et al., 2013), and exercise (Bradford, 2010; Bradford et al., 2017).

These results imply that the time-preference parameters are not robustly associated with certain

behaviors, in line with the findings of Bartels et al. (2023), even after accounting for outcome- and sign-dependent time preferences.

Insert Table 5 about here.

6. Summary and Discussion

To address our research questions on outcome- and sign-dependence of time preferences, we conducted two incentivized experiments with a 2×2 design contrasting two types of outcomes—money and effort—and two domains of experience—gain and loss. In the Main Experiment, participants were randomly assigned to one of the four conditions; in the Within-Subject Experiment, the outcome type varied across participants and the domain within participants, so that each participant made decisions over both gains and losses. Drawing on the framework by Bosch-Domènech & Silvestre (2010), our design permitted the incentivization of all decisions, even those involving future monetary losses, while minimizing potential biases.

Our analysis, combining model-free indices with individual-level quasi-hyperbolic parameter estimates, led to several key findings. In contrast to the meta-analytic evidence on effort choices (Imai et al., 2021), neither experiment found economically meaningful present bias.

Furthermore, our results indicate outcome-dependence. Although we found no economically meaningful present bias in either effort or monetary choices, time inconsistency was significantly higher in effort-based choices than in monetary ones. Participants also showed a stronger preference for smoothness in effort-based decisions than in monetary ones. At the individual level, in the Within-Subject Experiment, the correlation between gain and loss domains was notably stronger for effort choices than for monetary choices. The Within-Subject Experiment also identified sign-dependence in the *Impatience* index in both monetary and effort choices, though in opposite directions. However, we did not observe economically meaningful sign dependence in the Main Experiment.

6.1. Absence of Present Bias Across Monetary and Effort Decisions

Neither experiment shows economically meaningful present bias, in either monetary or effort choices, and the interpretation we find best supported — though, as we discuss below, not the only one the data permit — is a genuine null rather than an artifact of the design: the median β is approximately 1 in all four conditions, the result replicates across two independently conducted experiments, and it is robust to structural estimation (Section 4.1; Appendix B; Online Appendix OA.6). Across all specifications — pooled, individual, curvature-free, and tremble — no test rejects $\beta = 1$ in the present-bias direction by an economically meaningful margin; the only rejections below 1 are the two money cells of Table 3, where the deviations are of order 10^{-3} or smaller, and the remaining significant deviations lie above 1, on the future-bias side, and depend on the specification. Our condition medians also lie far closer to 1 than those reported by effort studies that detect present bias (Section 4.1). The remainder of this subsection asks whether the horizon, the immediacy of the sooner option, the elicitation method, the sample, or the stakes could instead have produced the null, and then quantifies the effect sizes our design can and cannot rule out.

A first candidate is the horizon: if present bias operates mainly over short delays, a design whose sooner–later gap is too long could fail to pick it up. A one-week gap between the sooner and later dates cannot explain the null, because it is the modal design in the effort-CTB literature: Augenblick et al. (2015), Abebe et al. (2021), Bartoš et al. (2021), and Kölle and Wenner (2023) all detect present bias with sooner–later gaps of about seven days, and Imas et al. (2022) detect it with a gap of one hour.

A second candidate is the immediacy of the sooner option: a sooner option not perceived as truly 'now' may fail to trigger present bias. Our implementation makes the sooner option as immediate as

the design permits: in the Week-4 block there is no front-end delay, and both payments and effort fall due on a specific, named day — the day of the decision itself. No implementation in this literature is literally instantaneous, and ours sits within the standard range: designs that execute the sooner option within the experimental session itself (Kölle and Wenner, 2023; Imas et al., 2022; Bartoš et al., 2021) are more immediate than same-day designs — Augenblick et al. (2015), Andreoni et al. (2023), and ours — which are in turn more immediate than the next-day implementation of Abebe et al. (2021). The last of these is the informative comparison: Abebe et al.'s sooner work session began only the day after the allocation decision — less immediacy than our design — yet their aggregate estimates show $\beta \approx 0.80$, so immediacy alone cannot explain our null. One caveat remains: app-mediated payment and same-week effort cannot be made perceptually instantaneous, and we cannot rule out that this residual attenuates measured present bias.

A third candidate is the elicitation method: the CTB itself might have masked present bias. Choice-consistency violations in CTB data can attenuate structural estimates (Chakraborty et al., 2017), and in CTB simulations a true β between roughly 0.9 and 1 cannot be reliably distinguished from 1 at the individual level, with estimates tending to shift toward 1 (Inukai et al., 2024); we accordingly acknowledge that a failure of detection cannot be fully excluded. At the same time, effort studies that also use the CTB do detect present bias, with a meta-analytic effort-pool β of approximately 0.88 (Imai et al., 2021), and the null is not likely to be a sampling accident, since it replicates across two independently conducted experiments. The effect sizes that could nonetheless have escaped detection are quantified at the end of this subsection.

Two further candidates are the sample and the stakes. Our participants were Japanese university students, and cross-cultural studies suggest that individuals from Eastern cultures tend to discount future rewards less than their Western counterparts (Du et al., 2002; Ishii et al., 2017; Kim et al., 2012; Takahashi et al., 2010); however, the monetary present bias estimated for a Japanese sample by Sawada and Kuroishi (2015) aligns closely with the meta-analytic average (Imai et al., 2021), so sample composition is unlikely to account for the null.[11] As for the stakes, the initial endowment of 2,000 units might seem large relative to the allocation options; but Abdellaoui et al. (2018) detect significant present bias in working time even with an initial value of four hours, and our additional analysis supports the interpretation that participants updated their reference points away from the initial endowment (Online Appendix OA.7).

Finally, we quantify which effect sizes remain compatible with our data. With 60–75 participants per cell ($\alpha = .05$, power $= .80$), the minimum detectable effect for the within-cell tests of present bias is $d \approx 0.33$–0.37 in standardized units (Cohen's d; 0.020–0.078 in raw index units), and for between-condition comparisons it is $d \approx 0.47$–0.50: the Main Experiment could reliably detect medium-sized present bias, but not small effects. Equivalence testing sharpens the bound in one cell: for present bias in the Effort–loss condition, a two one-sided tests (TOST) procedure rejects effects of $d = 0.3$ or larger ($p = 0.014$); in the remaining cells the equivalence test is not significant at that bound, so effects below the minimum detectable size cannot be excluded there. Section 6.2 discusses the statistical power of the between-subjects design for the sign-dependence comparisons.

6.2. Absence of Sign-Dependent Time Preferences in the Main Experiment

Although the Within-Subject Experiment revealed sign-dependent impatience in the *Impatience* index, this effect did not appear in the Main Experiment, which used a between-subjects design.

Notably, previous studies have identified sign effects using between-subjects designs: Mansell et al. (2024), Appelt et al. (2011), and Yoon and Chapman (2016) in monetary choices, and MacKeigan et al. (1993) in health-related decisions. However, all these studies relied on hypothetical scenarios,

[11] The participants in Sawada and Kuroishi's (2015) study resided in Iwanuma city, Miyagi prefecture. The city suffered damage from the 11th March 2011 Great East Japan Earthquake. Their field experiments were carried out in May, about half a year after a November 2013 survey from which the participants were drawn. They found that the impact of the earthquake-induced disaster damage on present bias was, in their words, 'not necessarily statistically significant'.

and the lack of real incentives is likely to inflate the observed effects (Krause & Schmidt, 2023). Our incentivized procedure may therefore have resulted in a smaller sign effect.

One explanation is salience. Within-subjects designs require participants to make decisions in both gain and loss domains, prompting direct comparisons between the two. This setup likely increases the salience of sign-related differences, thereby enhancing the detectability of the sign effect. In between-subjects designs, participants are only exposed to one domain, which may reduce their sensitivity to such differences. This explanation is consistent with findings on scope insensitivity (e.g., Hsee & Rottenstreich, 2004), where individuals fail to appropriately scale their responses when comparisons are not made salient.

Three further design differences bear on this divergence. First, block order: in the Within-Subject Experiment the effort sign contrast in *Impatience* is sensitive to which block came first (Online Appendix OA.1, Part 2), whereas the corresponding money contrast is not (Table OA.1.4); part of the effort sign effect is therefore consistent with an artifact of block order, which cannot arise when each participant faces a single domain. Second, the budget grid: for the Main Experiment we recalibrated the interest rates and budgets to the one-week delay (Section 3.1.3), so the two experiments do not present identical choice sets. Third, sample composition is unlikely to matter: after multiplicity correction the samples differ only in BMI (and its weight component), and controlling for BMI changes none of the cross-experiment comparisons of the index levels (Online Appendix OA.3). Online Appendix OA.5 sets out every effect under both measurement approaches in both experiments, and Online Appendix OA.4 records which significance verdicts move with the choice of test family.

A power analysis clarifies what the Main Experiment's null can and cannot tell us. For the between-subjects sign comparisons (60–75 participants per condition), the minimum detectable effect size at conventional levels ($\alpha = .05$, power $= .80$) is $d \approx 0.47$–0.49, while the sign-dependence effects observed in the Within-Subject Experiment are all smaller than this benchmark (the largest is $d \approx 0.45$, *Impatience* in the money conditions). Read together, the two experiments thus point to a consistent conclusion: the sign effect in time preferences, where it appears, is modest in magnitude, even though it has been documented in many earlier studies. The Main Experiment's null is what an effect of this modest size would be expected to produce in a between-subjects design with our samples. The data cannot fully separate two possibilities: a sign effect that genuinely weakens when domains are assigned between subjects, and an effect of similar, modest magnitude that our samples could not detect. The mechanisms discussed above therefore remain candidate explanations rather than established ones. This is a property of the effect size, not of an unusually small sample: our condition sizes are in line with comparable experiments in this literature, and detecting effects much below $d \approx 0.5$ between subjects requires samples far larger than is standard. The equivalence analysis in Section 6.1 quantifies what the data can affirmatively rule out for the present-bias nulls. Applying the same procedure to the sign comparisons here, with each comparison tested against $d = 0.3$ and against the size of the same index's sign effect in the Within-Subject Experiment, the data reject an effect as large as the Within-Subject money-*Impatience* effect ($d \approx 0.45$; $p = 0.038$) and effects of $d = 0.3$ or larger for *Impatience* in the effort conditions ($p = 0.039$); for the remaining comparisons, effects smaller than the minimum detectable size cannot be excluded.

6.3. Limitations of the Incentivization Method in the Loss Condition

One potential concern is that participants might try to avoid the required payments in the monetary loss condition—for example, by not completing the PayPay transfer after incurring a loss in Week 5. In practice, every participant who was required to make a payment did so: there were no non-repayers in either experiment (see Online Appendix OA.2 for the attrition and selection analysis). We believe this compliance was largely due to two factors: we clearly communicated during recruitment that participants were expected to complete all parts of the five-week study, and the completion reward, which included a preferred item valued at up to approximately 5,000 yen, was sufficiently attractive to motivate participation, including making the required payments. Although

our method relies on participant compliance, other incentivization methods—such as providing an initial windfall endowment—face similar challenges; we therefore consider our two-stage design a practical and effective method for incentivizing intertemporal loss decisions.

6.4. Time Preferences and Real-World Behaviors: Interpretation and Limitations

The correlations reported in Section 5 are contemporaneous; time preferences (Weeks 3 and 4) and behaviors (Week 5) were measured within the same study wave. How time preferences predict future behavior may differ substantially. It is therefore worthwhile to examine the predictive validity of outcome- and sign-dependent time preferences for actual future behaviors in longitudinal studies. A further step would be interventional: examining whether behaviors such as BMI respond when time preferences over effort are shifted—if such preferences can be shifted at all—would show whether the associations documented here have causal counterparts. If they do, interventions aimed at improving BMI might be more effective if they target individuals' time preferences regarding effort and leisure decisions, since BMI is only correlated with δ in the Effort-gain condition.

On the design side, a natural field experiment in a setting where an effort context already exists—for example, routine tasks in a real job, with the assigned workload varied upward or downward—would allow the reference point to be supplied by the real environment rather than by an experimenter-provided device, complementing the contract-based implementation used here.

Acknowledgements

The authors used generative AI tools (Anthropic's Claude) to assist with drafting and language editing of the manuscript text and with writing code for the analyses in this revision. The experiments were designed by the authors. The authors reviewed and edited all text and code and take full responsibility for the content of the manuscript.

Appendix

A. Measurements of Time Preference

We measure the participants' time preferences using choice-based indices. These descriptive measures are based on simple proportions of rewards allocated to sooner versus later payment dates, and therefore provide evidence regarding the behaviors of interest without relying on specific functional forms. The literature establishes that time preference can be decomposed into three distinct properties: impatience, present bias, and a preference for smoothing (Frederick et al., 2002). Based on Cheung et al. (2022), we measured these properties and added time inconsistency to capture time preferences in the four conditions: Money-gain, Money-loss, Effort-gain and Effort-loss.

Impatience describes the extent to which individuals discount delays, essentially measuring their willingness to wait for future outcomes. Impatience for gains is measured by the average proportions allocated to the sooner option in Week 3, the initial intertemporal decisions with a front-end delay. This is calculated as $Impatience = mean\left(\frac{x_s}{x_s+x_l}\right)$, where x_s and x_l are the sooner and later allocations at each decision problem, respectively. *Impatience* equals 1 if the participant always allocates all resources (the number of exempted tasks in Effort-gain) to the sooner dates (extremely impatient) and 0 if they always allocate them to the later dates (extremely patient).[12] For losses, the proportion of outcomes allocated to later dates serves as an indicator of impatience, computed as $Impatience = mean(\frac{x_l}{x_s+x_l})$. This is based on the rationale that an impatient individual would prefer to postpone losses to later dates.

[12]Note that the exclusion of corner bundles from our choice sets means that our descriptive measure *Impatience* does not reach these theoretical extremes in our experimental design. The same applies to *PB*, *TI* and *Smooth*, which are defined in this appendix.

Present bias (*PB*) indicates the extent to which individuals overweigh present outcomes. Present bias is measured by the difference between allocations at two periods, i.e., $PB = mean\left(\left(\frac{x_s}{x_s+x_l}\right)_{W4} - \left(\frac{x_s}{x_s+x_l}\right)_{W3}\right)$ in the gain domains and $PB = mean\left(\left(\frac{x_l}{x_s+x_l}\right)_{W4} - \left(\frac{x_l}{x_s+x_l}\right)_{W3}\right)$ in the loss ones. *PB* describes how much the allocations have changed from the questions with a front-end delay to those without. It takes values between -1 (extremely future-biased) and 1 (extremely present-biased).

Time inconsistency (TI) represents the extent of inconsistency in time preferences and is defined as the absolute value of PB: $TI = |PB|$.

Preference for smoothing (*Smooth*) describes how individuals prefer equal allocations between sooner and later dates. It is defined as $Smooth = mean(\frac{(x_s+x_l)-|x_s-x_l|}{x_s+x_l})$ for both gains and losses. It takes values between 0 (smoothing hater) and 1 (smoothing lover).

B. Pooled-Data Parameters Estimated by Maximum Likelihood Estimations (MLEs)

Within-Subject Experiment

Table B.1 reports the MLEs for the Money and Effort treatments, using pooled data. The four parameters of time preferences are estimated for each model. The estimated present bias β lies between 0.98 and 1.15 across the twelve models of both panels, and none of the Wald tests of $\beta = 1$ reported in the last row of each panel is significant at conventional levels (smallest p = 0.249). These pooled estimates are consistent with the individual-level results reported in Section 4.1.

Insert Table B.1 about here.

Models 1 and 4 use the gain domain, Models 2 and 5 the loss domain, and Models 3 and 6 both domains, with a 'loss' dummy variable to assess sign-dependent preferences.

We did not find that the discount factor δ was significantly different from one in either the Money treatment (p = 0.10, Model 1), or the Effort treatment (p = 0.30, Model 4) for gains. The utility curvature α does not significantly differ from 0 for either gains or losses in the Money treatment (p = 0.20, Model 1; p = 0.11, Model 2, respectively). The value is not significantly different from 0 in the Effort-loss condition (p = 0.11, Model 5), but is significantly negative in the Effort-gain condition (p = 0.02, Model 4).

This suggests that our subjects exhibited a linear utility function, except for gain outcomes in the Effort treatment where we observed a convex utility function. This result in this particular condition is not in line with previous studies, representing a unique feature of our methodology using CTB and MLE. It is critical, however, to note that this deviation occurs at the aggregate level: at the individual level, the median α in the Effort-gain condition is positive, though not significantly different from zero (0.012; Table OA.1.2), so the aggregate convexity is not a feature of the typical individual.

In the Money treatment, δ is significantly larger for the loss outcomes than for the gain ones (p = 0.002, Model 3). The results are consistent with the findings of the descriptive measures. However, no sign-dependent preference in terms of δ is found in the Effort treatment (p = 0.87, Model 6), whereas *Impatience* was significantly higher for losses in the descriptive measures. In addition, we do not find a significant difference for β between the gain and loss domains in either the Money or the Effort treatment (p = 0.58 for both treatments). This finding is again consistent with the results of our descriptive measures.

Main Experiment

In both the money and effort choices, δ is not significantly different between the gain and loss domains (p = 0.34 for the monetary choices; p = 0.83 for the effort choices). Similarly, no significant differences were found in β between the gain and loss domains for either money or effort choices (p = 0.69 for the monetary choices; p = 0.19 for the effort choices). This is consistent with our descriptive measures, which show no sign dependence in *Impatience* or *PB* in the Main Experiment.

References

Abdellaoui, Mohammed, Bleichrodt, Han, l'Haridon, Olivier, & Paraschiv, Corina. (2013). Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time. *Management Science*, *59*(9), 2153–2169. https://doi.org/10.1287/mnsc.1120.1690

Abdellaoui, Mohammed, Gutierrez, Cédric, & Kemel, Emmanuel. (2018). Temporal discounting of gains and losses of time: An experimental investigation. *Journal of Risk and Uncertainty*, *57*(1), 1–28. https://doi.org/10.1007/s11166-018-9287-1

Abebe, Girum, Caria, A. Stefano, & Ortiz-Ospina, Esteban. (2021). The selection of talent: Experimental and structural evidence from Ethiopia. *American Economic Review*, *111*(6), 1757–1806. https://doi.org/10.1257/aer.20190586

Andreoni, James, Callen, Michael, Hussain, Karrar, Khan, Muhammad Yasir, & Sprenger, Charles. (2023). Using preference estimates to customize incentives: An application to polio vaccination drives in Pakistan. *Journal of the European Economic Association*, *21*(4), 1428–1477. https://doi.org/10.1093/jeea/jvac068

Andreoni, James, Gravert, Christina, Kuhn, Michael A., Saccardo, Silvia, & Yang, Yang. (2018). *Arbitrage or narrow bracketing? On using money to measure intertemporal preferences* (Working Paper No. 25232). National Bureau of Economic Research. https://doi.org/10.3386/w25232

Andreoni, James, & Sprenger, Charles. (2012). Estimating time preferences from convex budgets. *American Economic Review*, *102*(7), 3333–3356. https://doi.org/10.1257/aer.102.7.3333

Angeletos, George-Marios, Laibson, David, Repetto, Andrea, Tobacman, Jeremy, & Weinberg, Stephen. (2001). The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation. *Journal of Economic Perspectives*, *15*(3), 47–68. https://doi.org/10.1257/jep.15.3.47

Appelhans, Bradley M., Waring, Molly E., Schneider, Kristin L., Pagoto, Sherry L., DeBiasse, Michelle A., Whited, Matthew C., & Lynch, Elizabeth B. (2012). Delay discounting and intake of ready-to-eat and away-from-home foods in overweight and obese women. *Appetite*, *59*(2), 576–584. https://doi.org/10.1016/j.appet.2012.07.009

Appelt, Kirstin C., Hardisty, David J., & Weber, Elke U. (2011). Asymmetric discounting of gains and losses: A query theory account. *Journal of Risk and Uncertainty*, *43*(2), 107–126. https://doi.org/10.1007/s11166-011-9125-1

Ashraf, Nava, Karlan, Dean, & Yin, Wesley. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, *121*(2), 635–672. https://doi.org/10.1162/qjec.2006.121.2.635

Augenblick, Ned, Niederle, Muriel, & Sprenger, Charles. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, *130*(3), 1067–1115. https://doi.org/10.1093/qje/qjv020

Augenblick, Ned, & Rabin, Matthew. (2019). An experiment on time preference and misprediction in unpleasant tasks. *The Review of Economic Studies*, *86*(3), 941–975. https://doi.org/10.1093/restud/rdy019

Baker, Forest, Johnson, Matthew W., & Bickel, Warren K. (2003). Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology*, *112*(3), 382–392. https://doi.org/10.1037/0021-843X.112.3.382

Barlow, Pepita, Reeves, Aaron, McKee, Martin, Galea, Gauden, & Stuckler, David. (2016). Unhealthy diets, obesity and time discounting: A systematic literature review and network analysis. *Obesity Reviews*, *17*(9), 810–819. https://doi.org/10.1111/obr.12431

Bartels, Daniel M., Li, Ye, & Bharti, Soaham. (2023). How well do laboratory-derived estimates of time preference predict real-world behaviors? Comparisons to four benchmarks. *Journal of Experimental Psychology. General*, *152*(9), 2651–2665. https://doi.org/10.1037/xge0001380

Bartoš, Vojtěch, Bauer, Michal, Chytilová, Julie, & Levely, Ian. (2021). Psychological effects of poverty on time preferences. *The Economic Journal*, *131*(638), 2357–2382. https://doi.org/10.1093/

ej/ueab007

Benzion, Uri, Rapoport, Amnon, & Yagil, Joseph. (1989). Discount rates inferred from decisions: An experimental study. *Management Science*, *35*(3), 270–284. https://doi.org/10.1287/mnsc.35.3.270

Bleichrodt, Han, Gao, Yu, & Rohde, Kirsten I. M. (2016). A measurement of decreasing impatience for health and money. *Journal of Risk and Uncertainty*, *52*(3), 213–231. https://doi.org/10.1007/s11166-016-9240-0

Bosch-Domènech, Antoni, & Silvestre, Joaquim. (2010). Averting risk in the face of large losses: Bernoulli vs. Tversky and Kahneman. *Economics Letters*, *107*(2), 180–182. https://doi.org/10.1016/j.econlet.2010.01.018

Bradford, David. (2010). The association between individual time preferences and health maintenance habits. *Medical Decision Making*, *30*(1), 99–112. https://doi.org/10.1177/0272989X09342276

Bradford, David, Courtemanche, Charles, Heutel, Garth, McAlvanah, Patrick, & Ruhm, Christopher. (2017). Time preferences and consumer behavior. *Journal of Risk and Uncertainty*, *55*(2–3), 119–145. https://doi.org/10.1007/s11166-018-9272-8

Brañas-Garza, Pablo, Jorrat, Diego, Espín, Antonio M., & Sánchez, Angel. (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *26*(2), 412–434. https://doi.org/10.1007/s10683-022-09776-5

Brewer, Judson A., & Potenza, Marc N. (2008). The neurobiology and genetics of impulse control disorders: Relationships to drug addictions. *Biochemical Pharmacology*, *75*(1), 63–75. https://doi.org/10.1016/j.bcp.2007.06.043

Carvalho, Leandro S., Meier, Stephan, & Wang, Stephanie W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *The American Economic Review*, *106*(2), 260–284. https://doi.org/10.1257/aer.20140481

Chakraborty, Anujit, Calford, Evan M., Fenig, Guidon, & Halevy, Yoram. (2017). External and internal consistency of choices made in convex time budgets. *Experimental Economics*, *20*(3), 687–706. https://doi.org/10.1007/s10683-016-9506-z

Chapman, Gretchen B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 771–791. https://doi.org/10.1037/0278-7393.22.3.771

Cheung, Stephen L., Tymula, Agnieszka, & Wang, Xueting. (2022). Present bias for monetary and dietary rewards. *Experimental Economics*, *25*(4), 1202–1233. https://doi.org/10.1007/s10683-022-09749-8

Christo, George, Jones, Susan L., Haylett, Samantha, Stephenson, Geoffrey M., Lefever, Robert M. H., & Lefever, Robin. (2003). The Shorter PROMIS Questionnaire: Further validation of a tool for simultaneous assessment of multiple addictive behaviours. *Addictive Behaviors*, *28*(2), 225–248. https://doi.org/10.1016/S0306-4603(01)00231-3

Coller, Maribeth, & Williams, Melonie B. (1999). Eliciting Individual Discount Rates. *Experimental Economics*, *2*(2), 107–127. https://doi.org/10.1023/A:1009986005690

Condon, David M., & Revelle, William. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Courtemanche, Charles, Heutel, Garth, & McAlvanah, Patrick. (2015). Impatience, incentives and obesity. *The Economic Journal*, *125*(582), 1–31. https://doi.org/10.1111/ecoj.12124

Cubitt, Robin P., & Read, Daniel. (2007). Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, *10*(4), 369–389. https://doi.org/10.1007/s10683-006-9140-2

Dohmen, Thomas, Falk, Armin, Huffman, David, & Sunde, Uwe. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*(3), 1238–1260. https://doi.org/10.1257/aer.100.3.1238

Du, Wanjiang, Green, Leonard, & Myerson, Joel. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *The Psychological Record*, *52*(4), 479–492. https://doi.org/10.1007/BF03395199

Duckworth, Angela L., & Seligman, Martin E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944. https://doi.org/10.1111/j.1467-9280.2005.01641.x

Estle, Sara J., Green, Leonard, Myerson, Joel, & Holt, Daniel D. (2007). Discounting of monetary and directly consumable rewards. *Psychological Science*, *18*(1), 58–63. https://doi.org/10.1111/j.1467-9280.2007.01849.x

Etchart-Vincent, Nathalie, & l'Haridon, Olivier. (2011). Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. *Journal of Risk and Uncertainty*, *42*(1), 61–83. https://doi.org/10.1007/s11166-010-9110-0

Falk, Armin, Becker, Anke, Dohmen, Thomas, Enke, Benjamin, Huffman, David, & Sunde, Uwe. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. https://doi.org/10.1093/qje/qjy013

Frederick, Shane, Loewenstein, George, & O'Donoghue, Ted. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/002205102320161311

Ganiats, Theodore G., Carson, Richard T., Hamm, Robert M., Cantor, Scott B., Sumner, Walton, Spann, Stephen J., Hagen, Michael D., & Miller, Christopher. (2000). Population-based time preferences for future health outcomes. *Medical Decision Making*, *20*(3), 263–270. https://doi.org/10.1177/0272989X0002000302

Greiner, Ben. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Gutierrez, Cédric, Sloof, Randolph, & Crilly, Donal. (2024). Time is not money! Temporal preferences for time investments and entry into entrepreneurship. *Organization Science*, *35*(2), 622–643. https://doi.org/10.1287/orsc.2023.1681

Halevy, Yoram. (2015). Time consistency: Stationarity and time invariance. *Econometrica*, *83*(1), 335–352. https://doi.org/10.3982/ECTA10872

Hardisty, David J., & Weber, Elke U. (2009). Discounting future green: Money versus the environment. *Journal of Experimental Psychology: General*, *138*(3), 329–340. https://doi.org/10.1037/a0016433

Harris, Christine R. (2012). Feelings of dread and intertemporal choice. *Journal of Behavioral Decision Making*, *25*(1), 13–28. https://doi.org/10.1002/bdm.709

Hsee, Christopher K., & Rottenstreich, Yuval. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23–30. https://doi.org/10.1037/0096-3445.133.1.23

Ida, Takanori. (2014). A quasi-hyperbolic discounting approach to smoking behavior. *Health Economics Review*, *4*(1), 5. https://doi.org/10.1186/s13561-014-0005-7

Ikeda, Shinsuke, Kang, Myong-Il, & Ohtake, Fumio. (2010). Hyperbolic discounting, the sign effect, and the body mass index. *Journal of Health Economics*, *29*(2), 268–284. https://doi.org/10.1016/j.jhealeco.2010.01.002

Imai, Taisuke, Rutter, Tom A., & Camerer, Colin F. (2021). Meta-analysis of present-bias estimation using convex time budgets. *Economic Journal*, *131*(636), 1788–1814. https://doi.org/10.1093/ej/ueaa115

Imas, Alex, Kuhn, Michael A., & Mironova, Vera. (2022). Waiting to choose: The role of deliberation in intertemporal choice. *American Economic Journal: Microeconomics*, *14*(3), 414–440. https://doi.org/10.1257/mic.20180233

Inukai, Keigo, Shimodaira, Yuta, & Shiozawa, Kohei. (2024). Investigation of the convex time budget experiment by parameter recovery simulation. *Journal of Behavioral and Experimental Fi-*

nance, *43*, 100962. https://doi.org/10.1016/j.jbef.2024.100962

Ishii, Keiko, Eisen, Charis, & Hitokoto, Hidefumi. (2017). The effects of social status and culture on delay discounting. *Japanese Psychological Research*, *59*(3), 230–237. https://doi.org/10.1111/jpr.12154

Kahneman, Daniel, & Tversky, Amos. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kim, Bokyung, Sung, Young Shin, & McClure, Samuel M. (2012). The neural basis of cultural differences in delay discounting. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 650–656. https://doi.org/10.1098/rstb.2011.0292

Kölle, Felix, & Wenner, Lukas. (2023). Is generosity time-inconsistent? Present bias across individual and social contexts. *The Review of Economics and Statistics*, *105*(3), 683–699. https://doi.org/10.1162/rest_a_01049

Krause, Jan, & Schmidt, Ulrich. (2023). *Hyperopic loss aversion* (SSRN Scholarly Paper No. 4651320). Social Science Research Network. https://doi.org/10.2139/ssrn.4651320

Laibson, David. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478. https://doi.org/10.1162/003355397555253

List, John A. (2025). *Experimental economics: Theory and practice*. University of Chicago Press.

Loewenstein, George. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, *97*(387), 666–684. https://doi.org/10.2307/2232929

MacKeigan, Linda D., Larson, Lon N., Draugalis, JoLaine R., Bootman, J. Lyle, & Burns, Lawton R. (1993). Time preference for health gains versus health losses. *PharmacoEconomics*, *3*(5), 374–386. https://doi.org/10.2165/00019053-199303050-00005

Mansell, Wade Sean, Li, Ye, & Hardisty, David. (2024). Assessing patience and predictivity validity for mixed sign intertemporal choices. *Judgment and Decision Making*, *19*, e30. https://doi.org/10.1017/jdm.2024.30

Matousek, Jindrich, Havranek, Tomas, & Irsova, Zuzana. (2022). Individual discount rates: A meta-analysis of experimental evidence. *Experimental Economics*, *25*(1), 318–358. https://doi.org/10.1007/s10683-021-09716-9

McClelland, Jessica, Dalton, Bethan, Kekic, Maria, Bartholdy, Savani, Campbell, Iain C., & Schmidt, Ulrike. (2016). A systematic review of temporal discounting in eating disorders and obesity: Behavioural and neuroimaging findings. *Neuroscience and Biobehavioral Reviews*, *71*, 506–528. https://doi.org/10.1016/j.neubiorev.2016.09.024

Meier, Stephan, & Sprenger, Charles. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, *2*(1), 193–210. https://doi.org/10.1257/app.2.1.193

Meissner, Thomas, & Pfeiffer, Philipp. (2022). Measuring preferences over the temporal resolution of consumption uncertainty. *Journal of Economic Theory*, *200*, 105379. https://doi.org/10.1016/j.jet.2021.105379

Moffatt, Peter G., & Peters, Simon A. (2001). Testing for the presence of a tremble in economic experiments. *Experimental Economics*, *4*(3), 221–228. https://doi.org/10.1023/A:1013265203635

Myerson, Joel, Baumann, Ana A., & Green, Leonard. (2017). Individual differences in delay discounting: Differences are quantitative with gains, but qualitative with losses. *Journal of Behavioral Decision Making*, *30*(2), 359–372. https://doi.org/10.1002/bdm.1947

O'Donoghue, Ted, & Rabin, Matthew. (2015). Present bias: Lessons learned and to be learned. *American Economic Review*, *105*(5), 273–279. https://doi.org/10.1257/aer.p20151085

Odum, Amy L., Baumann, Ana A. L., & Rimington, Delores D. (2006). Discounting of delayed hypothetical money and food: Effects of amount. *Behavioural Processes*, *73*(3), 278–284. https://doi.org/10.1016/j.beproc.2006.06.008

Odum, Amy L., Madden, Gregory J., & Bickel, Warren K. (2002). Discounting of delayed health gains and losses by current, never- and ex-smokers of cigarettes. *Nicotine & Tobacco Research*, *4*(3), 295–303. https://doi.org/10.1080/14622200210141257

Reed, Derek D., & Martens, Brian K. (2011). Temporal discounting predicts student responsiveness to exchange delays in a classroom token system. *Journal of Applied Behavior Analysis*, *44*(1), 1–18. https://doi.org/10.1901/jaba.2011.44-1

Reuben, Ernesto, Sapienza, Paola, & Zingales, Luigi. (2010). Time discounting for primary and monetary rewards. *Economics Letters*, *106*(2), 125–127. https://doi.org/10.1016/j.econlet.2009.10.020

Sawada, Yasuyuki, & Kuroishi, Yusuke. (2015). How to Strengthen Social Capital in Disaster Affected Communities? The Case of the Great East Japan Earthquake. In *Disaster Risks, Social Preferences, and Policy Effects: Field Experiments in Selected ASEAN and East Asian Countries* (pp. 163–199). ERIA.

Shiba, Shotaro, & Shimizu, Kazumi. (2020). Does time inconsistency differ between gain and loss? An intra-personal comparison using a non-parametric elicitation method. *Theory and Decision*, *88*(3), 431–452. https://doi.org/10.1007/s11238-019-09728-1

Sutter, Matthias, Kocher, Martin G., Glätzle-Rützler, Daniela, & Trautmann, Stefan T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, *103*(1), 510–531. https://doi.org/10.1257/aer.103.1.510

Takahashi, Taiki, Hadzibeganovic, Tarik, Cannas, Sergio A., Makino, Takaki, Fukui, Hiroki, & Kitayama, Shinobu. (2010). Cultural Neuroeconomics of Intertemporal Choice. *Journal of Behavioral Economics and Finance*, *3*, 133–135. https://doi.org/10.11167/jbef.3.133

Tanaka, Tomomi, Camerer, Colin F., & Nguyen, Quang. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, *100*(1), 557–571. https://doi.org/10.1257/aer.100.1.557

Thaler, Richard. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207. https://doi.org/10.1016/0165-1765(81)90067-7

Thaler, Richard, & Johnson, Eric J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, *36*(6), 643–660. https://doi.org/10.1287/mnsc.36.6.643

Tsukayama, Eli, & Duckworth, Angela Lee. (2010). Domain-specific temporal discounting and temptation. *Judgment and Decision Making*, *5*(2), 72–82. https://doi.org/10.1017/S1930297500000930

Vuchinich, Rudy E., & Simpson, Cathy A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology*, *6*(3), 292–305. https://doi.org/10.1037/1064-1297.6.3.292

Weber, Elke U., Shafir, Sharoni, & Blais, Ann-Renée. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430–445. https://doi.org/10.1037/0033-295X.111.2.430

Weller, Rosalyn E., Cook, Edwin W., Avsar, Kathy B., & Cox, James E. (2008). Obese women show greater delay discounting than healthy-weight women. *Appetite*, *51*(3), 563–569. https://doi.org/10.1016/j.appet.2008.04.010

Yamamoto, Shohei, & Navarro-Martinez, Daniel. (2022). The endowment effect in the future: How time shapes buying and selling prices. *Judgment and Decision Making*, *17*(5), 988–1014. https://doi.org/10.1017/S1930297500009293

Yoon, Haewon, & Chapman, Gretchen B. (2016). A closer look at the yardstick: A new discount rate measure with precision and range. *Journal of Behavioral Decision Making*, *29*(5), 470–480. https://doi.org/10.1002/bdm.1890

Table 1. The summary table of demographic and behavioral variables, by condition, Main Experiment.

Variable	MG Mean (SD) (n=60)	ML Mean (SD) (n=75)	EG Mean (SD) (n=73)	EL Mean (SD) (n=70)	Kruskal-Wallis p-value
Age	22.62 (2.50)	22.54 (3.45)	22.56 (2.75)	22.54 (2.69)	0.87
Female	0.48 (0.50)	0.43 (0.50)	0.42 (0.50)	0.51 (0.50)	0.74
Grad	0.30 (0.46)	0.28 (0.45)	0.29 (0.46)	0.36 (0.48)	0.85
BMI	20.48 (2.25)	20.50 (2.31)	20.35 (2.49)	19.91 (2.14)	0.31
GPA	2.96 (0.73)	2.97 (0.66)	2.94 (0.69)	3.01 (0.58)	1.0
Saving_rate	28.83 (20.79)	30.33 (20.47)	28.75 (22.58)	28.66 (22.95)	0.84
Matrix_score	1.97 (1.04)	1.96 (1.06)	2.23 (1.25)	2.00 (1.01)	0.43
Rotation_score	0.82 (1.02)	1.05 (1.24)	0.81 (1.06)	0.81 (1.05)	0.72

Notes: 'Saving_rate' refers to the proportion of the monthly income that participants saved (in the Within-Subject Experiment, of the monthly salary). 'Matrix_score' represents the scores derived from four matrix reasoning items. 'Rotation_score' captures the scores from four assessments involving three-dimensional rotation tasks. n is the number of participants in each condition; variables with missing values have fewer observations. Kruskal–Wallis p-values are computed without the correction for ties.

Table 2. The summary statistics of descriptive measures, by condition, Main Experiment.

	(1) Money-Gain mean/sd	(2) Money-Loss mean/sd	(3) Effort-Gain mean/sd	(4) Effort-Loss mean/sd
Impatience	0.41	0.40	0.43	0.43
	(0.04)	(0.09)	(0.21)	(0.19)
PB	0.01	0.01	−0.05	−0.01
	(0.05)	(0.08)	(0.23)	(0.19)
TI	0.02	0.03	0.14	0.13
	(0.05)	(0.08)	(0.19)	(0.15)
Smooth	0.12	0.17	0.43	0.48
	(0.16)	(0.24)	(0.29)	(0.32)
N	60	75	73	70

Table 3. Quasi-hyperbolic parameter estimates, Main Experiment (individual level).

Median (interquartile range) by condition; benchmark tests are sign-rank tests (p-values; direction of the median deviation in brackets where rejected; no direction is shown for $\tau = 0$, since $\tau \geq 0$ by construction).

	Money-Gain (n=60)	Money-Loss (n=75)	Effort-Gain (n=73)	Effort-Loss (n=70)
δ	1.000 (0.970–1.000)	1.001 (0.994–1.008)	1.000 (0.912–1.335)	1.001 (0.737–1.508)
β	1.000 (0.999–1.000)	0.999 (0.997–1.001)	1.018 (0.961–1.432)	0.999 (0.840–1.338)
α	0.002 (−1.698–0.002)	−0.002 (−0.004–0.191)	0.142 (0.002–0.901)	−0.788 (−1.647–−0.021)
$\tau = \lvert 1-\beta \rvert$	0.000 (0.000–0.009)	0.001 (0.001–0.015)	0.249 (0.029–0.805)	0.201 (0.055–0.762)
Benchmark tests (sign-rank p)				
$\delta = 1$	0.086	0.196	0.134	0.375
$\beta = 1$	0.006 [<1]	0.007 [<1]	0.007 [>1]	0.470
$\alpha = 0$	0.073	0.895	<0.001 [>0]	<0.001 [<0]
$\tau = 0$	<0.001	<0.001	<0.001	<0.001

Notes: Between-condition comparisons (outcome-dependence and sign-dependence) are reported in the text with values, directions, and p-values; the complete test battery under both test families is in Online Appendix OA.4. The corresponding estimates for the Within-Subject Experiment are in Table OA.1.2 (Online Appendix OA.1).

Table 4. The mean differences of the four descriptive measures.

(a) Within-Subject Experiment

	Sign Dependence		Outcome Dependence		Augenblick Comparison
	(1)	(2)	(3)	(4)	(5)
	MG - ML	EG - EL	MG - EG	ML - EL	MG - EL
Impatience	0.02***	−0.02**	−0.01	−0.06***	−0.04**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
PB	0.00	0.01	0.01	0.03	0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
TI	−0.01	0.00	−0.10***	−0.08***	−0.10***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Smooth	−0.01	−0.04***	−0.21***	−0.24***	−0.25***
	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
N	178	152	165	165	165

(b) Main Experiment

	Sign Dependence		Outcome Dependence		Augenblick Comparison
	(1)	(2)	(3)	(4)	(5)
	MG - ML	EG - EL	MG - EG	ML - EL	MG - EL
Impatience	0.01	0.00	−0.03	−0.03	−0.02
	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)
PB	0.00	−0.04	0.06*	0.02	0.02
	(0.01)	(0.04)	(0.03)	(0.02)	(0.03)
TI	−0.01	0.01	−0.12***	−0.09***	−0.10***
	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
Smooth	−0.05	−0.05	−0.31***	−0.31***	−0.36***
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
N	135	143	133	145	130

Notes: Standard errors in parentheses; *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. In panel (a), N in Columns (1) and (2) counts observations (one gain and one loss observation per participant); elsewhere N counts participants.

Table 5. Relationship between δ and β and self-reported everyday behaviors in the Within-Subject Experiment (top) and the Main Experiment (bottom).

(a) Within-Subject Experiment

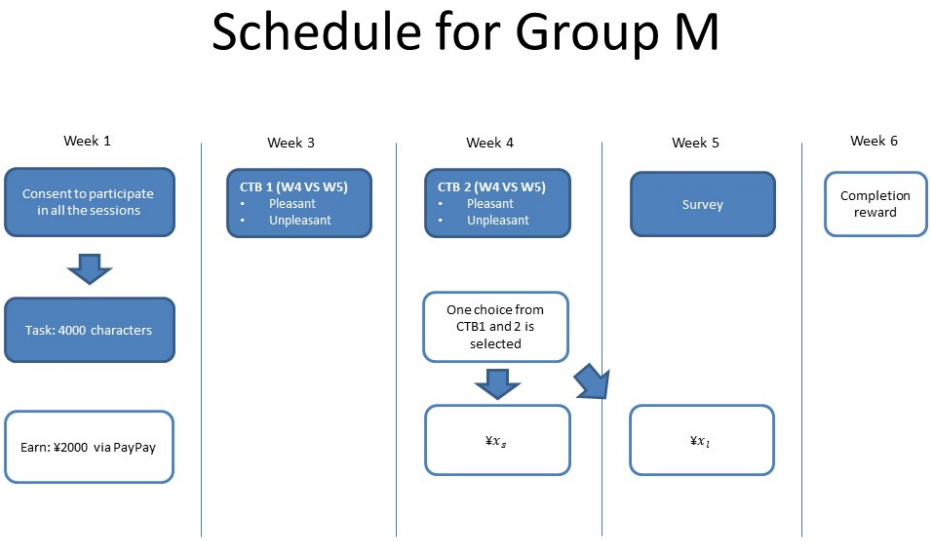
Outcome	Term	Money-gain	Money-loss	Effort-gain	Effort-loss	Model
BMI	δ	−3.9928	−6.9616	−0.2789***	−0.1539	OLS
		(5.0700)	(6.7684)	(0.1019)	(0.0945)	
	β	−2.6347	+1.1590	+0.0615	+0.0993*	
		(1.9158)	(1.6863)	(0.0465)	(0.0520)	
	n	86	86	74	74	
	Adjusted R-squared	−0.0216	−0.0152	0.1345	0.1478	
GPA	δ	−1.4618	+0.1432	+0.0648***	+0.0598*	OLS
		(1.2637)	(0.8875)	(0.0230)	(0.0301)	
	β	−0.6375	+0.0518	−0.0089	−0.0020	
		(0.7223)	(0.2235)	(0.0102)	(0.0109)	
	n	82	82	75	75	
	Adjusted R-squared	0.0615	0.0408	−0.0021	−0.0157	
Binge eating	δ	−1.8174	−1.3068	+0.0616	−0.0386	OLS
		(2.5077)	(1.3585)	(0.0547)	(0.0715)	
	β	−1.1917	−0.8790***	−0.0192	−0.0429	
		(1.3017)	(0.2067)	(0.0170)	(0.0263)	
	n	87	87	76	76	
	Adjusted R-squared	0.0160	0.0185	0.1888	0.2312	

(b) Main Experiment

Outcome	Term	Money-gain	Money-loss	Effort-gain	Effort-loss	Model
BMI	δ	−13.6086	−0.9289	−0.5778***	+0.2254	OLS
		(8.9770)	(0.9030)	(0.1428)	(0.1423)	
	β	−4.1046*	+0.0039	+0.0140	−0.0318	
		(2.1010)	(0.9649)	(0.0530)	(0.0456)	
	n	59	71	69	69	
	Adjusted R-squared	−0.0082	0.1518	0.0704	−0.0298	
GPA	δ	−1.7748	+0.2302	+0.1287**	+0.0244	OLS
		(2.7538)	(0.1417)	(0.0627)	(0.0476)	
	β	−0.3974	+0.0338	−0.0179	+0.0100	
		(0.6645)	(0.3695)	(0.0182)	(0.0081)	
	n	58	67	69	65	
	Adjusted R-squared	−0.0779	0.0270	0.0718	0.1773	
Binge eating	δ	+9.6082*	−0.3259***	+0.1008	+0.0393	OLS
		(5.4126)	(0.0894)	(0.0945)	(0.0656)	
	β	+2.2042*	−0.2387	−0.0325	+0.0288**	
		(1.3146)	(0.2005)	(0.0348)	(0.0123)	
	n	60	74	72	70	
	Adjusted R-squared	0.0396	0.2434	0.1173	0.0969	

Notes: Each cell reports the coefficient with its heteroskedasticity-robust standard error in parentheses beneath. Every regression includes the same control variables: age, age squared, a female indicator and a graduate-student indicator. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The behavioral variables were obtained from the survey conducted in Week 5. 'Binge eating' is the average score for binge eating measured by the Shorter PROMIS Questionnaire (Christo et al., 2003). The full outputs for all ten behaviors, including the control-variable coefficients, are in Online Appendix OA.10.

A: Money treatment



B: Effort treatment

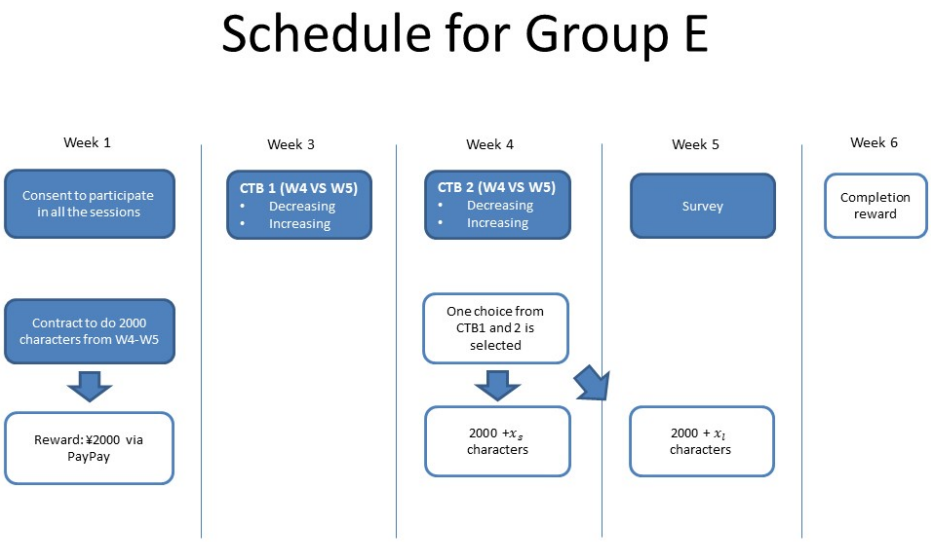


Figure 1. Experimental schedule for each treatment.

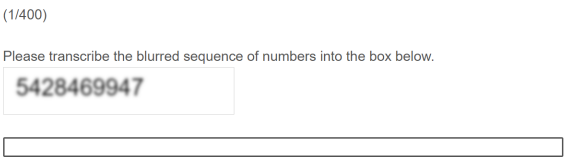


Figure 2. English translation of an example for the transcription tasks

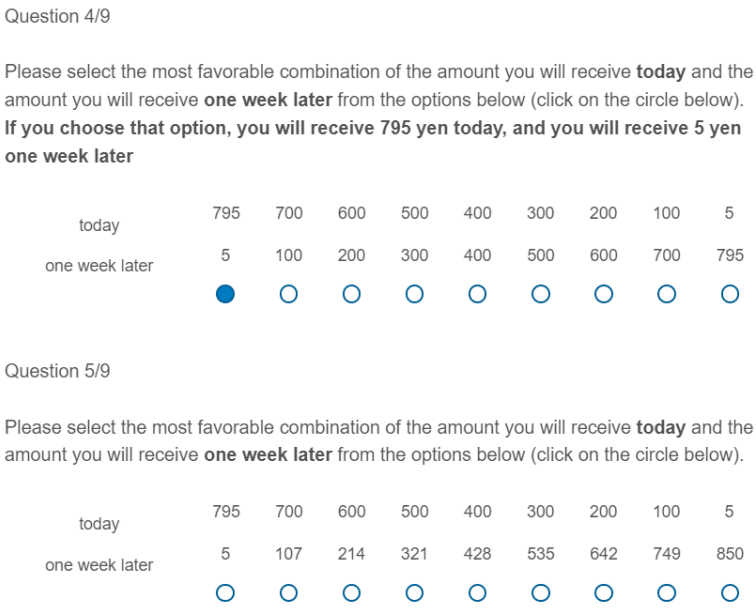
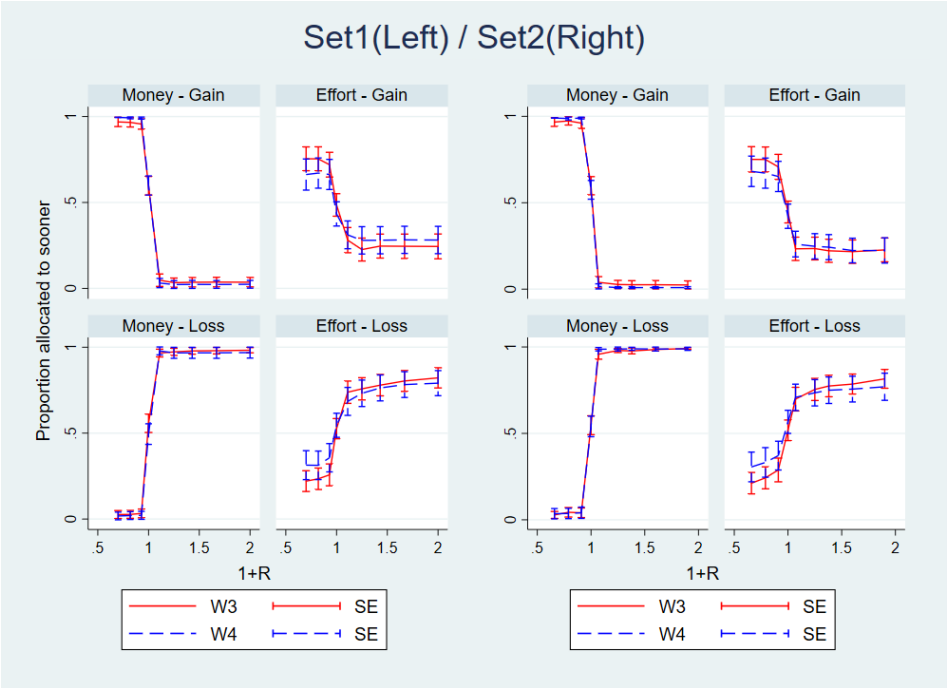


Figure 3. English translation of the screenshots of the allocation decisions in Week 4 (Within-Subject Experiment): monetary choices in the gain domain with R = 0 (top) and R = 0.07 (bottom).

(a) Within-Subject Experiment



(b) Main Experiment

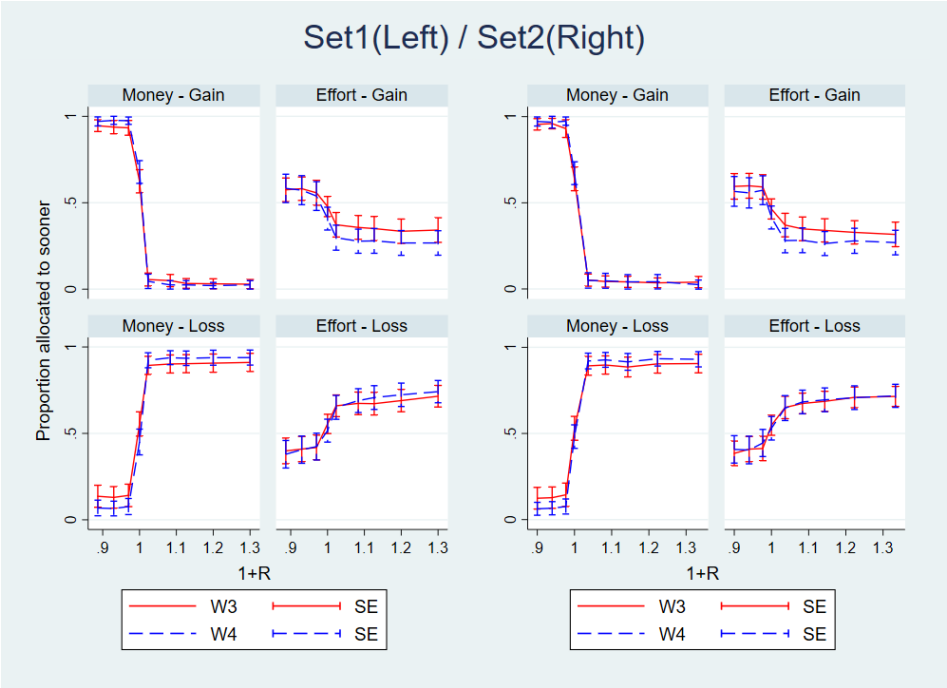
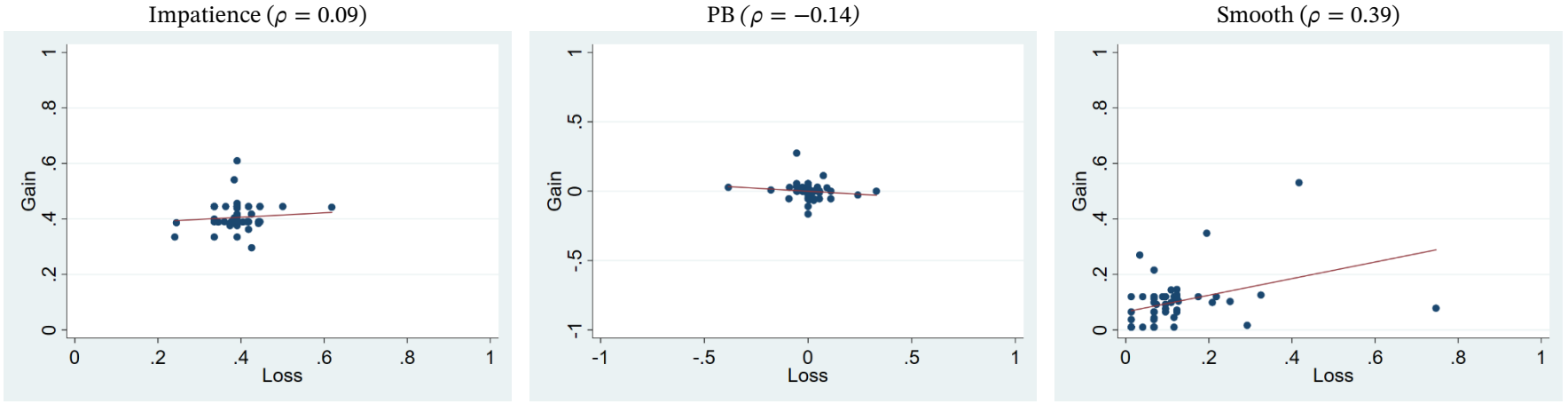


Figure 4. Proportions of allocations in the four conditions for Set 1 (left) and Set 2 (right) in the Within-Subject Experiment (top) and the Main Experiment (bottom).

NOTE: The solid lines indicate the proportion of the allocations in Week 3, and the dotted lines indicate the proportion of the allocations in Week 4. Each bar describes the standard error.

A: Money treatment



B: Effort treatment

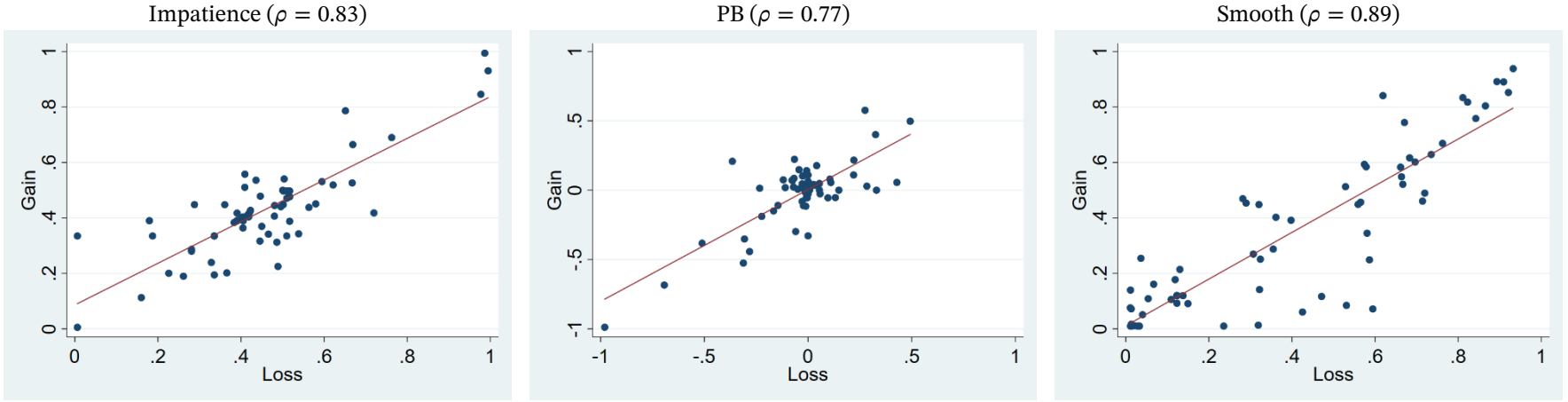


Figure 5. Indices of time preferences based on descriptive measures in money (top) and effort (bottom).

Note: The values in the gain domain are shown on the x-axis, and the values in the loss domain are shown on the y-axis. The dots represent observations. The red lines describe the fitted lines. ρ is the Pearson correlation between the gain- and loss-domain values.

Table B.1. The four parameters of time preferences by MLEs.

(a) Within-Subject Experiment

	Money			Effort		
	(1)	(2)	(3)	(4)	(5)	(6)
	Gain	Loss	All	Gain	Loss	All
δ	0.98***	1.00***	0.98***	1.16***	1.04***	1.05***
	(0.01)	(0.01)	(0.01)	(0.16)	(0.06)	(0.09)
Loss			0.02***			−0.01
			(0.01)			(0.04)
β	0.99***	1.01***	1.00***	1.15***	1.09***	1.10***
	(0.01)	(0.01)	(0.01)	(0.20)	(0.11)	(0.16)
Loss			−0.00			−0.03
			(0.01)			(0.05)
α	−0.10	0.05	−0.08*	−0.86**	0.21	−0.49***
	(0.08)	(0.03)	(0.05)	(0.35)	(0.13)	(0.12)
Loss			0.15***			0.72***
			(0.06)			(0.15)
σ	0.02**	0.01**	0.01***	0.14*	0.17**	0.12**
	(0.01)	(0.01)	(0.00)	(0.08)	(0.07)	(0.06)
Loss			0.00			0.06
			(0.01)			(0.04)
LogLikelihood	−2059.30	−2053.95	−4264.67	−5068.75	−5028.18	−10256.98
Cluster	89	89	89	76	76	76
N	3204	3204	6408	2736	2736	5472
p-value:$\delta = 1$	0.103	0.532	0.035	0.298	0.495	0.577
p-value:$\beta = 1$	0.404	0.368	0.856	0.468	0.393	0.517

(b) Main Experiment

	Money			Effort		
	(1)	(2)	(3)	(4)	(5)	(6)
	Gain	Loss	All	Gain	Loss	All
δ	0.99***	0.99***	0.98***	1.03***	1.10***	1.00***
	(0.00)	(0.01)	(0.00)	(0.03)	(0.05)	(0.05)
Loss			0.00			0.01
			(0.00)			(0.04)
β	1.01***	0.99***	0.99***	0.98***	0.99***	1.05***
	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.09)
Loss			0.00			−0.06
			(0.00)			(0.05)
α	−0.07***	0.12***	−0.06***	0.31***	0.14	0.11
	(0.01)	(0.01)	(0.01)	(0.04)	(0.12)	(0.12)
Loss			0.14***			0.05
			(0.02)			(0.07)
σ	0.02***	0.06***	0.02***	0.08***	0.37	0.12
	(0.00)	(0.01)	(0.01)	(0.02)	(0.24)	(0.08)
Loss			0.03**			0.01
			(0.01)			(0.05)
LogLikelihood	−1754.61	−2959.28	−4859.89	−5175.82	−5212.20	−10536.77
Cluster			135			143
N	2160	2700	4860	2628	2520	5148
p-value:$\delta = 1$	0.004	0.066	0.001	0.236	0.044	0.924
p-value:$\beta = 1$	0.249	0.255	0.369	0.579	0.812	0.556

Notes: Standard errors in parentheses are clustered at the individual level in all columns of panel (a) and in Columns (3) and (6) of panel (b), and heteroskedasticity-robust at the observation level in Columns (1), (2), (4), and (5) of panel (b); *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. In every model, each of the four parameters is estimated with a female indicator, a graduate-student indicator, and the matrix reasoning and three-dimensional rotation scores as covariates (coefficients not shown); Columns (3) and (6) additionally include the loss-domain indicator. The reported parameters are the estimates with all covariates set to zero (male undergraduates with both cognitive scores of zero and, in Columns (3) and (6), the gain domain); the Loss rows give the implied loss-minus-gain difference in the parameter level at the same covariate values.