Revisiting mental accounting classic paradigms: Replication of the problems reviewed in Thaler (1999)

Mengfei Li
ORCID: 0000-0001-6460-555X
The University of Hong Kong, Hong Kong SAR
lmf0312@hku.hk / lmf0312@connect.hku.hk

^Gilad Feldman
ORCID: 0000-0003-2812-6599
Department of Psychology, University of Hong Kong, Hong Kong SAR
gfeldman@hku.hk / giladfel@gmail.com

^Corresponding author

Author bios:

Mengfei Li was an undergraduate student at the University of Hong Kong during the academic year 2021-2022.

Gilad Feldman is an assistant professor with the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

Declaration of Conflict of Interest:

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial disclosure/funding:

This project has been supported by the Teaching Development Grant from the University of Hong Kong awarded to Gilad Feldman.

Authorship declaration:

Mengfei Li conducted the project as part of her thesis in psychology.

Gilad was the thesis advisor and guided Mengfei Li. Gilad supervised each step in the project, conducted the pre-registrations, ran data collection, and edited the manuscript for submission.

Corresponding author

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; gfeldman@hku.hk ; 0000-0003-2812-6599

Rights:

Important links and information

Citation of the target research article:

> Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making, 12*(3), 183-206.

Link to the target research article.

In-principle Acceptance and open-review

https://rr.peercommunityin.org/articles/rec?id=164

Contributor Roles Taxonomy

Role	Mengfei Li	Gilad Feldman
Conceptualization	X	X
Pre-registration	X	
Data curation		X
Formal analysis	X	
Funding acquisition		X
Investigation	X	
Pre-registration peer review / verification		X
Data analysis peer review / verification		X
Methodology	X	
Project administration		X
Resources		
Software	X	
Supervision		X
Validation		X
Visualization	X	
Writing-original draft	X	
Writing-review and editing		X

Abstract

Mental accounting, the internal categorization system individuals adopt to manage their financial activities, makes people prone to irrational decision-making. In Replication Registered Report with an American online Amazon Mechanical Turk sample using CloudResearch (N = 1007), we attempted a replication of 17 problems reviewed in Thaler (1999) on the topic of mental accounting. We concluded a mostly successful replication: Out of the 17 problems, we found empirical support for 12, mixed empirical support for 3, and no empirical support for 2. Extending the replication, we provided an initial test of four predictions that were described in Thaler's (1999) paper, for which we found empirical support for 2, mixed support for 1, and no support for 1. Systematic replications and extensions of review articles using a single data collection are a promising direction in revisiting seminal findings, mapping and examining untested assumptions and predictions, comparing different designs and effects, and identifying possible links, gaps, and future directions. Materials, dataset, and analysis code were made available on the OSF: https://osf.io/v7fbj/

Keywords: Mental accounting, bias, judgment and decision making, registered report, replication

PCIRR-Study Design Table

Question	Hypothesis	Sampling plan	Analysis plan	Interpretation given different outcomes
Do people engage in mental accounting activities?	For all the problems, participants -on average - follow the patterns of the original findings.	Aiming for a high statistical power, the current study aimed to recruit a sample size of 1000, larger than the required sample size suggested by power analysis.	We followed the data analysis plan conducted in the original studies and added additional analysis when needed.	Support the findings of the original studies reviewed by Thaler (1999), confirming our hypothesis Raise doubts on the reproducibility of the original results if we fail to replicate.
Are there links between and a consistency among the different mental accounting behaviors?	There is a high intercorrelation among the mental accounting problems.		Examined intercorrelations between mental accounting problems.	High intercorrelation suggests a high consistency among the different mental accounting paradigms, confirming our hypothesis. Low intercorrelation suggests a lack of consistency among the paradigms.

Note. Requested as part of the PCI-RR submissions

Revisiting mental accounting classic paradigms:

Replication of the problems reviewed in Thaler (1999)

Background

Mental accounting is an internal control system that individuals use to evaluate, manage, and monitor their financial activities (Thaler, 1999). By utilizing this set of cognitive operations, people aim to simplify their financial decision-making process. As mental accounting violates fundamental neo-classical economic principles and can influence consumer choice, Thaler (1999) reviewed a decade of relevant research to summarize and to emphasize the importance of the topic.

Thaler (1999) focused on three most noticeable components of the mental accounting paradigms. Firstly, Thaler (1999) pointed out that mental accounting describes how people perceive and experience outcomes. It explains how people make and evaluate their financial decisions. Secondly, grouping expenses into categories is another defining feature of mental accounting (Zhang & Sussman, 2018). The mental accounting system demonstrates how different activities are assigned into specific separate accounts. For example, Heath and Soll (1996) suggested that expense must first be "booked" and then "posted" into proper account with reference to the similarity and categorization. Thirdly, mental accounting concerns how choices are grouped together and how frequently people evaluate the mental accounts. Individuals and households can balance accounts on a daily, monthly, or a yearly basis and can define the accounts either narrowly or broadly (Thaler, 1999). Mental accounting is comparable to financial accounting that businesses conduct to monitor expenditures (Jha-Dang, 2006).

We report a very close replication of the work reviewed in Thaler (1999), with the following goals (closeness evaluation based on the LeBel et al., 2018 criteria). Our first goal was

to conduct close replications of the classic mental accounting problems reviewed by Thaler (1999) by an external independent lab (KNAW, 2018; Peels, 2019). Our second goal was to examine several predictions made by Thaler regarding mental accounting behaviors that the review did not cover empirical tests for.

We begin by introducing the literature on mental accounting and the chosen article for replication - Thaler (1999). We then highlight the motivations for the current replication study and provide an overview of our replications problems.

Mental accounting

Mental accounting has long been a heated topic in the field of behavioral economics, psychology, and judgment and decision making. The earliest empirical evidence on mental accounting behaviors dates back to Tversky and Kahneman's (1981) famous theater-ticket experiment (one of our replication problems). Tversky and Kahneman proposed that mental accounting is a form of decision framing by which people formulate (psychological) accounts to evaluate events and options (as cited in Henderson & Peterson, 1992). Their findings suggested that in people's minds, losses tend to be labeled into different categories, violating the long-standing economic notion of fungibility (Thaler, 1999). Ever since, the concept of mental accounting has been used to understand a wide range of decision-making behaviors, such as gambling, risk taking, and investment (Thaler & Johnson, 1990). Although these subsequent mental accounting studies differ in specific objectives and orientations, nearly all research has touched upon gains and losses, and indicated the violation of fundamental economic norms (Henderson & Peterson, 1992). In Thaler (1985) and our target article-Thaler (1999), the mental accounting phenomenon was further elaborated and summarized into a broad theory on consumer choices (Muehlbacher & Kirchler, 2019). In 2011, Soman and Ahn reviewed substantial mental

accounting research focusing on the framing effects. More recently, Zhang and Sussman (2018)'s review paper again outlined the categorization process of mental accounting, and they summarized it as a way for people to "group expenses into categories, assign funds to these categories, determine budgets, and perform elements of cost–benefit analyses." (p.65). The two recent review papers cited very similar research to Thaler (1999), such as Heath and Soll (1996), Tversky and Kahneman (1981), Thaler (1980), and Thaler and Johnson (1990). This further exemplifies the necessity in revisiting these classic findings and testing the reproducibility, robustness, and generalizability of these influential and pioneering works, to substantiate and strengthen the empirical foundations of the theoretical framework of mental accounting. We aimed to examine the evidence reviewed in Thaler (1999) targeting different subsets of the mental accounting framework.

Choice of article for replication: Thaler (1999)

We chose the Thaler (1999) article based on three factors: extensive academic impact, need for systematic direct replications, and the potential in methodological improvements.

As of February 2022, the time of writing, there were 4567 Google Scholar citations of the review article and many important follow-up theoretical and empirical research. The review has had an immense impact on scholarly research in the area of behavioral economics, judgment and decision-making, and consumer psychology. The research covered in Thaler (1999) has also been highly influential. We summarized the citations impact of each of the problems covered in the article in Table 1.

To the best of our knowledge, there are currently no published systematic attempts for direct replications of the mental accounting findings reviewed in Thaler (1999), and there are no published independent direct pre-registered well-powered replications of Thaler's own work.

We also recognized the potential for improving on both transparency and methods. For example, among the problems we aimed to replicate, several did not report basic methodological details like sample size. The statistical analysis strategies were also often not reported in detail. These suggest the need to revisit these problems to reproduce their materials, deduce and improve on their methods, and reassess and update their findings to current day.

In sum, we aimed to revisit the classic mental accounting phenomenon to examine the reproducibility and replicability of the findings with replications by an external independent team. Following the recent growing recognition of the importance of reproducibility and replicability in psychological science (Open Science Framework & Lakens, 2012), we embarked on a well-powered pre-registered very close replication of the work reviewed by Thaler (1999).

Original hypotheses and findings in target article

Thaler (1999) reviewed a decade of research demonstrating the mental accounting phenomenon and we aimed to replicate most of the problems he summarized. We provided a summary of the original studies and their hypotheses in Table 1. We noted that for some of the problems, the original hypotheses were not explicitly stated, and so we deduced our version of the underlying hypotheses. Please see Tables 9-12 for a summary of the findings of the original studies. We provided further details of the problems in the supplementary materials section "Instructions and experimental material".

Table 1

Problems reviewed in Thaler (1999): Citations, descriptions, and hypotheses

Problem	Google scholar citations	Description and Explanation	Hypothesis

Problem	Google scholar citations	Description and Explanation	Hypothesis
1	Based on **Tversky and Kahneman, 1986** (608)	**(Risk Taking)** Manipulation with two conditions testing diminishing sensitivity towards gain and loss. People perceive outcomes based on the value function in the prospect theory.	**H1:** People are risk-averse for gains and risk-seeking for losses.
2	**Tversky and Kahneman, 1981** (23905)	**(Time investment versus price reference point)** Manipulation with two conditions testing reference points: Spending 20 minutes to save $5 out of $15 versus $5 out of $125	**H2:** Reference points shift evaluations of value. People are more likely to spend 20 minutes to save $5 out of $15 than to save $5 out of $125
3	**Tversky and Kahneman, 1981** (23905)	**(Theater Play Ticket)** Manipulation with two conditions testing the impact of an existing account on decisions. The sunk cost effect arises when the decision is referred to an existing account set up by a related act.	**H3:** Not explicitly reported
4	**Thaler, 1985** (8237)	**(Events and Happiness)** Four pairs of scenarios testing the hedonic framing. People tend to frame outcomes in ways that make them the happiest.	**H4:** People follow four principles (a) segregate gains, (b) integrate loss, (c) cancel losses against larger gains, and (d) segregate "silver linings".
5	**Thaler and Johnson, 1990** (3194)	**(Same Day or Two Weeks Apart)** Three pairs of events testing the temporal spacing of hedonic editing. People tend to simplify and encode multiple outcomes in a hedonically optimal manner.	**H5:** The hedonic editing hypothesis argues that subjects choose to have the events occur "apart" when segregation is preferred, and "together" when integration is hedonically optimal.
6	**Thaler and Johnson, 1990** (3194)	**(Emotional Impact of Losing $9)** Manipulation with two conditions testing the effect of a prior loss. People do not integrate subsequent losses with the prior loss when faced with a two-stage gamble.	**H6:** Not explicitly reported
7	**Thaler, 1985** (8237)	**(Location and Price)** Manipulation with two conditions testing transaction utility. Consumption decisions are influenced by people's perceived value of the "deal".	**H7:** Transaction utility influences willingness to pay.

Problem	Google scholar citations	Description and Explanation	Hypothesis
8	**Thaler, 1985** (8237)	**(Selling Ticket)** Manipulation with three conditions testing the determinants of the reference point. Fairness is the dominant factor in determining reference price.	**H8:** Not explicitly reported
9[a]	**Shafir and Thaler, 1998** (5)	**(Wine Bottle)** Manipulation with two conditions testing the value of wine. People hold mixed perceptions of the value of items when the consumption and purchase is temporally separated.	**H9:** Not explicitly reported
10	**Shafir and Thaler, 1998** (5)	**(Feelings about Purchase)** Three statements examining "investment" purchases. People can avoid the feeling of spending when the purchase is perceived as investment.	**H10:** Not explicitly reported
11	**Heath and Soll, 1996** (1035)	**(Previous Events and New Payment)** Manipulation with two conditions testing the underconsumption of a typical target. People set budgets for different accounts and recomputed the remaining budgets periodically. They will resist further expenses if a particular budget is depleted.	**H11a:** The budget-setting process promotes greater underconsumption in the $50 condition than the $20 condition. **H11b:**The expense-tracking process promotes greater underconsumption for more typical purchases.
12	**Leclerc et al., 1995** (706)	**(The Performance)** Manipulation with two conditions testing whether the value of time is influenced by price-related characteristics of the decision situation. The value of time is influenced by contextual effects.	**H12:** Subjects are willing to pay more money to avoid waiting the same amount of time for a higher-priced food or service than for a lower-priced product.
13-15	**Thaler, 1999** (4637)	**(Choices)** A gain (Q13) vs. loss (Q14-15) scenario examines prior outcomes and risky choices. When gambles are bracketed together, the outcome of the prior gamble can influence subsequent choices.	**H13:** "House Money" effect- Prior gain stimulates risk seeking **H14-15**: Weaker to no effects for prior loss unless the gamble offers a chance to break even

Problem	Google scholar citations	Description and Explanation	Hypothesis
16	**Samuelson, 1963** (983)	**(Coin Flip Bet)** A scenario testing how bracketing the gambles affects the attractiveness of individual bets.	**H16**: People shift between single gambles and long-term repeating gambles.
17	**Thaler, 1999** (4637)	**(Division Investment)** Two scenarios examining the myopic loss aversion effect. Narrow framing inhibits risk-taking.	**H17:** Not explicitly reported

Note. The Google scholar citations were of April 2022. [a]For Shafir and Thaler (1998), it was later published titled "Invest Now, Drink Later, Spend Never: The Mental Accounting of Delayed Consumption" and there were 229 Google Scholar citations.

Extensions - Prediction extensions

We extended the replication by also adding a test of four predictions that Thaler (1999) reflected on but did not review empirical evidence that directly tested these predictions. We summarized our extensions in Table 2.

Overview of replication and extension

Thaler's (1999) review paper covered a long list of classic mental accounting problems, and we focused on 17 of those. For each of the replication problems, we followed the original experimental design with minor adjustments needed to update those to current times to make those suitable for our target sample. We summarized the minor changes we made in Table 7. We then added four additional experiments to examine predictions Thaler made that were not reviewed with supporting empirical evidence.

Table 2

Extension: Summary of predictions made by Thaler (1999) with no reviewed supporting evidence

Extensions	Description and Explanation	Predictions/Hypothesis
Problem 18	**Thaler, 1980** **Game in Bad Weather** Two scenarios testing the effect of sunk costs. Payment for a good increases the likelihood of its usage	If the family pays for the tickets, they will go anyway. If the tickets are given to them, they will stay home.
Problem 19	**Thaler, 1980** **Membership and Tennis Elbow** One scenario examining the effect of sunk costs. Paying for the right to use a service increases the likelihood of utilization.	The person will continue to play in pain.
Problem 20[a]	**Thaler, 1999** **Price and Decision** Two statements testing how sunk costs affect subsequent decisions.	The more one pays for the shoes, the more times one will try to wear them. Eventually one stops wearing the shoes, but will not throw them away. The more one pays for the shoes, the longer before the shoes are thrown away.
Problem 21[b]	**Thaler, 1999** **Annual membership** Manipulations with three conditions testing expenses framing. People tend to ignore small, routine expenses.	Membership phrased as "merely 27 cents a day" will be more attractive.

Note. The papers listed are the sources of the predictions and none of the predictions have been tested directly to the best of our knowledge. [a] For Problem 20, we aimed to examine how much participants identify with Thaler's prediction. [b] The pennies-a-day effect in Problem 21 has been investigated in the marketing field (e.g. Gourville, 1998).

Pre-registration and open-science

We provided all materials, data, and code on: https://osf.io/v7fbj/. This project received Peer Community in Registered Report Stage 1 in-principle acceptance (https://osf.io/d6cjk/; https://rr.peercommunityin.org/articles/rec?id=164) after which we created a frozen pre-registration version of the entire Stage 1 packet (https://osf.io/xu7jb/) and proceeded to data collection. All measures, manipulations, exclusions conducted for this investigation are reported, and data collection was completed before analyses.

Method

Power analysis

To ensure that the current replication sample has sufficient power, we calculated effect sizes and power based on the statistics reported in the original studies. For the replication studies, Rstudio was implemented to perform power analysis, where alpha (two-sided)=0.05 and power=0.95 were used. The largest required sample size was 321 participants, indicated by the power analysis of Problem 15. Therefore, we concluded that the minimum required sample size for a power of 0.95 and alpha of 0.05 was 321 participants. We provided more information regarding these calculations in Section "Power analysis of original study effect to assess required sample for replication" in the supplementary.

Given the possibility that the original effects are overestimated, and taking into account the issues of multiple comparisons and potential exclusions, we aimed to recruit 500 participants. Given reviewer's feedback, we decided to make a change in our implementation so that each participant will be randomized into 9 of the 18 Qualtrics blocks, aiming to cut survey time by half. The implication is that the actual sample for each of the Problems would be on average about half of what we previously intended. To compensate for that, we doubled our overall sample to 1000. A sensitivity analysis indicates that we would be powered to detect effects of f = 0.17 (groups = 3, df = 1) and d = 0.29/0.36 (between, 250/166 in each condition) (both 95% power, alpha = 5%, one-tail), which are effects much weaker than any of the supported effects in the reviewed studies.

Participants

We recruited native English speakers who were born, raised, and located in the US on Amazon Mechanical Turk using the CloudResearch/Turkprime platform (Litman et al., 2017). A total of 1007 participants completed the study (M_{age} = 43.28 years, SD=12.61 years, 471 females, 526 males, 3 others, and 7 rather not disclose). In fact, 1073 subjects began the survey but 66 did not proceed beyond the consent and verifications. We summarized the sample differences between the current replication and the original studies in Table 3 .

Based on our extensive experience of running similar judgment and decision-making replications on MTurk, to ensure high-quality data collection, we employed the following CloudResearch options: Duplicate IP Block. Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, Block Low Quality Participants, etc. We also employed the Qualtrics fraud and spam prevention measures: reCAPTCHA, prevent multiple submission, prevent ballotstuffing, bot detection, security scan monitor, relevantID, etc.

Assignment pay is based on the federal wage of 7.25USD/hour, per minute. We first pretested survey duration with 30 participants to test time run estimate and adjusted pay based on the duration. The data of the 30 participants was not analyzed other than to assess survey completion duration and needed pay adjustments.

Table 3

Summary of samples in the original studies and our replication

Factors	Sample size	Characteristics	Medium (location) and Compensation
The current replication	1007 (471 females, 526 males, 3 others, and 7 rather not disclose)	US American (Median age=40.00 years, Average age=43.28 years, Standard deviation age=12.61 years, age range=20-80 years)	Computer (online)
Problem 1	254	126 for Gain condition and 128 for Loss condition	Unreported
Problem 2	181	93 for $15 Calculator Condition and 88 for $125 Calculator Condition	Unreported
Problem 3	383	183 for Lost a $10 bill condition and 200 for Lost the ticket condition	Unreported
Problem 4	87	Undergraduate students in a statistical class at Cornell University	(In person)
Problem 5	65	/	Unreported
Problem 6	168	Cornell MBA students, 87 for Condition A and 81 for Condition B	Unreported
Problem 7	Unreported	Regular beer drinkers in an executive development program	(In person)
Problem 8	85	First-year MBA students, 31 for Free condition, 28 for Paid $5 condition, 26 for Paid $10 condition	Unreported
Problem 9	173	Subscribers to a wine newsletter, *Liquid Assets*, and are highly knowledgeable wine consumers with substantial home cellars, 97 for Giving away condition and 76 for Drinking condition	Unreported
Problem 10	Unreported	Subscribers to a wine newsletter, *Liquid Assets.*	Unreported
Problem 11	66	MBA students, split evenly across conditions	(In person), Pizza and beer
Problem 12	67 (37 male and 30 female)	MBA students	Unreported
Problem 13	Unreported	MBA students	(In person) Played for real money

Factors	Sample size	Characteristics	Medium (location) and Compensation
Problem 14	Unreported	MBA students	(In person) Played for real money
Problem 15	Unreported	MBA students	(In person) Played for real money
Problem 16	1	An economist colleague	(In person)
Problem 17	26	A CEO and 25 executives from one firm, each of whom was responsible for managing a separate division	(In person)

Design and procedure

We summarized the experimental designs in Table 4. We mapped the designs used in the problems, which included one-sample, between-subject, and within-subject experimental designs. We set up all the Problems using Qualtrics. Adopting a formatting method that can best reduce participants' cognitive load, we had a total of 18 Qualtrics blocks. Participants were randomly assigned to complete 9 of the 18 blocks, in order to address reviewer's feedback to decrease the length of the survey and the burden on participants. The display of Problems and conditions was counterbalanced using the randomizer "evenly present" function in Qualtrics. Problems were presented in random order and participants were randomly and evenly assigned into different conditions. We previously tested this method in many other replications and extensions conducted by our team, for example, a similar replication of an influential review paper by Kahneman and Tversky (1972) (Wan & Feldman, 2021). Our findings from projects using a similar design (e.g., Chen et al., 2023; Yeung & Feldman, 2022) suggest that combining several experiments in a single data collection in random order does not impact likelihood of replication success, and allows for important additional insights. This seems especially powerful in addressing concerns about the target sample (naivety, attentiveness, etc.) when some studies replicate successfully whereas others do not, as well as in the potential in drawing inferences

about the links between the different studies and consistency in participants' responding to similar paradigms.

We provided further details in the section "Materials and scales used in the replication + extension problems" in the supplementary materials.

Table 4

Replication and extension experimental design

Problem 1: Between	**IV: Gain condition** Choices between sure/uncertain *gain*		**IV: Loss condition** Choices between sure/uncertain *loss*
	DV: Risk taking preference (choice)		
Problem 2: Between	**IV: $15 Calculator Condition** Jacket is $125; Calculator is $15		**IV: $125 Calculator Condition** Jacket is $15; Calculator is $125
	DV:Willingness to travel to another store (choice)		
Problem 3: Between	**IV: "Lost a bill" Condition** Lost a $10 bill as you enter the theater		**IV: "Lost the ticket" Condition** Lost the $10 ticket as you enter the theater
	DV: Willingness to buy (another) ticket (choice)		
Problem 4: Multiple experiments, one sample proportions	IV: Hedonic framing		
	DV: Whether perceived as emotionally equivalent (choice) Specific DV items: After reading the scenario, participants choose who was happier/more upset.		
Problem 5: Within	IV: Temporal spacing		
	DV: Whether perceived as emotionally equivalent (choice) Specific DV items: After reading the scenario, participants are to choose who was happier/more unhappy.		
Problem 6: Between	IV: Incremental impact of loss Manipulation: Different prior outcomes		
	DV: Emotional Impact of Losing $9 (choice) Specific DV items: After reading the scenario, participants are to choose which event hurts more.		
Problem 7: Between	**IV: Hotel condition** The soda is sold at a fancy resort hotel.		**IV: Grocery store condition** The soda is sold at a small, run-down grocery store.
	DV: Price willing to pay (continuous)		
Problem 8: Mixed: Between-subject design (Free vs. $5	**IV: Free ticket condition** The tickets were given for free by a friend.	**IV: Paid $5 condition** The tickets were bought at $5.	**IV: Paid $10 condition** The tickets were bought at $10.

vs. $10) Within: Friend vs. Stranger. Within: Market price $5 vs. $10	DV: Price willing to sell (continuous) Specific DV items: Participants are to indicate their selling price when the customer is a friend/stranger when the going price is $5/$10.	
Problem 9: Between	**IV: Drinking condition** Participants are to imagine drinking a bottle of the wine with dinner	**IV: Giving away condition** Participants are to imagine giving one bottle of the wine to a friend as a gift
	DV: Feeling of the cost Specific DV items: Participants are to choose which statement best captures their feeling of the cost.	
Problem 10: Within	IV: Purchase of Bordeaux futures at $400	
	DV: Feeling about purchase (ordinal) Specific DV items: Participants are to choose which statement best captures their feeling at the purchase.	
Problem 11^{a}: Mixed: Between-subject design ($50 vs. $20) Within: Dinner vs. Ticket vs. Flu Within: Spent/given	**IV: $50 condition** Spent $50 on the previous event	**IV: $20 condition** Spent $20 on the previous event
	DV: Willingness to buy a $25 theater ticket later in the week (choice)	
Problem 12^{b}: Between	**IV: $15 condition** The tickets will cost $15 each	**IV: $40 condition** The tickets will cost $40 each
	DV: Price willing to pay to avoid waiting (continuous)	
Problem 13: Within	IV: Won $30 scenario	
	DV: Risk taking preference Specific DV items: Imagine winning $30, participants are to choose between uncertain gain/loss or no further gain/loss.	
Problem 14: Within	IV: Lost $30 scenario A	
	DV: Risk taking preference Specific DV items: Imagine losing $30, participants are to choose between uncertain gain/loss or no further gain/loss.	
Problem 15: Within	IV: Lost $30 scenario B	
	DV: Risk taking preference Specific DV items: Imagine losing $30, participants are to choose between uncertain gain/loss or a sure gain.	
Problem 16: Within	IV: Coin Flip Bet	
	DV: Willingness to take the bet Specific DV items: Decision under a single coin flip/100 coin flips	
Problem 17: Within	IV: Division Investment	
	DV: Willingness to undertake the investment Specific DV items: Decision under a single project/ a portfolio of 25 projects	

Problem 18: Within (extension)	IV: The cost of the ticket		
	DV: Willingness to go to the game Specific DV items: Decision between go to the game and stay home when the ticket is bought/given		
Problem 19: Within (extension)	IV: Membership at tennis club		
	DV: Willingness to play Specific DV items: Decision between stop playing and continue playing		
Problem 20: Within (extension)	IV: Shoe purchase scenario		
	DV: Sunk cost effects Specific DV items: Participants are to indicate how accurately the statements apply to them.		
Problem 21: Between (extension)	**IV: Day expression condition**	**IV: Year expression condition**	**IV: Both expressions condition**
	DV: Attractiveness of the membership plan		

Note. Please see Tables 9 and 10 for the options of each Problem. [a] For problem 11, Thaler (1999) did not summarize the study design comprehensively, and we also found the method of the original article difficult to understand. Therefore, we only replicated part of Study 2 in Heath and Soll (1996). [b] For Problem 12, it is possible that Thaler (1999) wrongly reported the second condition, as our understanding is that the ticket price should be $40, whereas Thaler (1999) wrote $45. For our replication, we followed our understanding of the original version.

Participants first read a consent form and indicated their willingness to participate, and then answered several verification questions. They are then randomly assigned to answer Problems in 9 of the 18 Qualtric blocks. At the end of the survey, participants answered funneling questions and provided their demographic information before being directed to the debriefing.

Manipulations

We provided full details of the manipulations between the conditions and the experimental designs in the supplementary materials section “Materials and scales used in the replication + extension problems”. Problems 1, 2, 3, 6, 7, 8, 9, 11, 12, and 21, involved between-subjects manipulations, and participants were randomly assigned to conditions separately in each of those. The order of the problems was also randomized. Please see Table 4 for a summary of all problems and manipulations.

Measures

Replications: 17 problems testing mental accounting paradigms

We summarized the measures and data analysis strategery for all replication problems in Table 5.

Table 5

Measures and data analysis strategery for replication problems

		Data analysis strategery	
Problem	**Measure**	**In the original**	**Deduced additional analysis**
1	Choose between two choices (displayed in random order)	Calculated the cumulative percent frequency for each choice	Chi-square
2	Answer the Yes/No question (options displayed in random order)	Calculated the cumulative percent frequency for the Yes option	Chi-square
3	Answer the Yes/No question (options displayed in random order)	Cumulative percent frequency for each choice	Chi-square
4	Four pairs of scenarios are presented in random order. Choose among three choices.	Calculated the cumulative percent frequency for all three choices.	Proportion tests
5	Three pairs of scenarios are presented in random order. Choose among three choices.	Calculated the cumulative percent frequency for all three choices.	Proportion tests McNemar paired-samples tests: A-B A-C
6	Five pairs of events displayed in random order. Choose among three choices.	Calculated the cumulative percent frequency for all three choices.	Proportion tests
7	Report what price they will tell the friend.	Calculated the median for the two conditions	Independent samples t-test

		Data analysis strategery	
Problem	**Measure**	**In the original**	**Deduced additional analysis**
8[a]	Report what price they will ask under different condition	Calculated percent of subjects giving common answers (0, 5, 10, Other).	Mixed ANOVA: 3 between: free vs. paid $5 vs. paid $10 2 within: friend vs. stranger 2 within: market worth $5 vs. $10
9	Choose among five statements (displayed in random order).	Calculated the cumulative percent frequency for all five statements.	Chi-square
10	Indicate which statement more accurately captures their feelings on a 5-point Likert scale (four statements displayed in random order).	Calculated the mean score of the statements.	Repeated measures ANOVA
11	Five scenarios are presented in random order. Answer the Yes/No question.	Calculated the correlation between underconsumption and typicality.	Mixed ANOVA: 2 between: $20 low-cost vs. $50 high-cost 2 within: given vs. spent 2 within: dinner vs. ticket
12	Report how much they would be willing to pay to avoid waiting.	Calculated the mean score for each condition and conducted independent sample t-tests.	Subtract the price of the ticket , exclude data below 0, and conduct independent samples t-test
13	Choose between two choices (displayed in random order).	Calculated the cumulative percent frequency for each choice	Baseline against 14 and 15
14	Choose between two choices (displayed in random order).	Calculated the cumulative percent frequency for each choice	McNemar paired-samples tests 13-14
15	Choose between two choices (displayed in random order).	Calculated the cumulative percent frequency for each choice	McNemar paired-samples tests 13-15
16	Two scenarios are presented in random order. Answer the Yes/No question	No data analysis was performed	McNemar paired-samples tests
17	Two scenarios are presented in random order. Answer the Yes/No question	No data analysis was performed	McNemar paired-samples tests

Note: [a] For Problem 8, there was no explanation provided regarding the classification of "common answers", so we can only assume that any value other than 0, 5, and 10 were counted as "Other".

Extensions: Testing predictions by Thaler with no reviewed supporting evidence

We added four problems that were not originally tested in the review article, and we summarized the measures and data analysis strategy for these extensions in Table 6.

Table 6

Measures and data analysis strategery for prediction extension problems

Problem	Measure	Data analysis strategery
18	Choose between two choices (displayed in random order)	McNemar paired-samples tests
19	Choose between two choices (displayed in random order)	Proportions test
20	Report how accurately the two statements express their feelings on a 5-point Likert scale.	One-sample t-test against the midpoint
21	Rate the attractiveness of the membership plan on a 0-100 scale.	Independent samples t-test Paired sample t-test

Deviations from the original studies

Our replication deviations from the original's studies include participants' characteristics, delivery mode, and the extensions. We summarized sample deviations in Table 3 and technical deviations in Table 7.

Table 7

Deviations for specific problems

Problem number	Deviation details	Reason for change
Replication Problem 1	We adopted the wording Thaler used in his work. For the Gain condition, the original second option was framed as "*A 50 % chance to gain $200 and a 50% chance to lose $0*". We changed it to "*A 50 % chance to gain $200 and a 50% chance to gain $0*" in our current replication.	We tried to be as close as possible to Thaler. While comparing the option with the loss condition, we suspected it as a typo.
Replication Problem 4	The description of the problem is slightly simplified and we randomized the sequence of the scenarios.	Minor wording changes for clarity. Randomization to eliminate order bias.
Replication Problem 5, 6	The description of the problem and the options are revised and simplified. We also randomized the sequence of the scenarios.	Minor wording changes for clarity. Randomization to eliminate order bias.
Replication Problem 7	The original study used beer in the two conditions and we changed it into soda.	Some of our targeted participants may not drink.
Replication Problem 9	Added "Imagine that you enjoy drinking wine" at the beginning of the scenario. We also randomized the sequence of the statements.	Our targeted population would mostly come from the working/lower-middle class and might not enjoy drinking wine. Randomization to eliminate order bias.
Replication Problem 10	1. Added "Imagine that you enjoy drinking wine" at the beginning of the scenario. 2. Added another option "*I cannot understand this question*" 3. Changed the Likert scale to a 1 (*not accurate at all*) to 5 (*very accurate*) rating 4. Randomized the sequence of the statements.	1. Our targeted population would mostly come from the working/lower-middle class and might not enjoy drinking wine. 2. Our pretest showed that this scenario might be too vague and difficult to comprehend for our targeted participants so we added another statement to check for understanding.By adding this option, we ensure that participants will not just choose a random option when they cannot understand the question. 3. To reduce cognitive load 4. Randomization to eliminate order bias.

Problem number	Deviation details	Reason for change
Replication Problem 11	The instruction of the problem is simplified.	Minor wording changes for clarity and understandability.
Replication Problem 12	The original study used student tickets at the student window in the scenarios. We changed it into discounted tickets and discount windows. The question is revised.	Our targeted population would have a wide age range from 18 to 80 so many of them might not be students. Our pretest showed that the original framing of the question caused misunderstandings. We revised for greater clarity.
Replication Problem 13, 14, 15	Added "Imagine that…"at the beginning of the scenarios.	To facilitate perspective-taking
Extension Problem 21	Thaler (1999) used "local public radio station" in his prediction while we changed it into "music online streaming service".	The original scenario does not apply to 2022 so we change it to update to current time.

Evaluation criteria for replication findings

We aimed to compare the replication effects with the effects in the original studies using the criteria set by LeBel et al. (2019) (see Section "Replication Evaluation" in the supplementary).

Replication closeness evaluation

We provided details on the classification of the replications using the criteria by LeBel et al. (2018) in Table 8 below (also see Section *"*Replication Evaluation" in the supplementary). We summarized the replication as a "very close" replication.

Table 8

Classification of the replication, based on LeBel et al. (2018)

Design facet	Replication	Details of deviation
Effect/hypothesis	Same	
IV construct	Same	
DV construct	Same	
IV operationalization	Same	
DV operationalization	Same	
Population (e.g. age)	Different	The current replication collected data from MTurk.
IV stimuli	Similar	Scenarios were slightly adjusted to update to current time and the targeted population.
DV stimuli	Similar	Problem 1 was changed for suspicious typo and we added another statement in Problem 10.
Procedural details	Different	To account for the order effect, we randomized the order of scenarios/statements in the problems.
Physical settings	Different	The current replication was conducted online via Qualtrics.
Contextual variables	Different	
Replication classification	Very close replication	Based on the above analysis, we summarized our replications as a “very close” replication of the original studies.

Data analyses plans: Outliers and exclusions

We categorized values more extreme than 3 standard deviations around the mean as outliers for Problems with numeric answers (Leys et al., 2019). Please refer to the supplementary Section “Exclusion criteria” for detailed data exclusion method.

Results

Replications and extensions

We conducted our analyses on the full sample. In Tables 9-12, we summarized the descriptives for all the problems, alongside the findings from the original studies to allow for an easy comparison. We summarized the statistical tests in Tables 13-17.

Overall, we concluded the replication as mostly successful. Six out of the nine problems that employed a between-subjects design, and six out of the eight problems that had a within-subjects design showed consistent results with the findings reported in the target article.

Table 9

Comparison and descriptive statistics for Problems with options and a between-subjects design

Problem	Conditions and Options	Original		Replication	
		Percentage	*N*	Count	Percentage
1	**Gain:**				
	A sure gain of $100	72%	250	209	84%
	A 50 % chance to gain $200 and a 50% chance to gain $0	28%		41	16%
	Loss:				
	A sure loss of $100	36%	254	112	44%
	A 50% chance to lose $200 and a 50% chance to lose $0	64%		142	56%
2	**$15 Calculator:**				
	Make the trip	68%	253	29	11%
	Not making the trip	[32%]		224	89%
	$125 Calculator:				
	Make the trip	29%	252	17	7%
	Not making the trip	[71%]		235	93%
3	**Lost a $10 bill:**				
	Buy the ticket	88%	252	228	90%
	Not buying the ticket	12%		24	10%
	Lost the ticket:				
	Buy another ticket	46%	251	194	77%
	Not buying another ticket	54%		57	23%
6^a (between)	**First group of questions:**				
	1. (A) You lose $9. (B) You lose $9 after having gained $30.		253		
	A hurts more	70%		208	82%
	B hurts more	9%		32	13%
	No difference	21%		13	5%
	2. (A) You lose $9. (B) You lose $9 after having lost $30.				
	A hurts more	13%	253	19	8%
	B hurts more	55%		218	86%
	No difference	31%		16	6%
	3. (A) You lose $9. (B) You lose $9 after having lost $250.				
	A hurts more	39%	253	32	13%
	B hurts more	38%		209	83%
	No difference	23%		12	5%
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $1000.				
	A hurts more	50%	253	37	15%
	B hurts more	33%		201	79%
	No difference	17%		15	6%
	5. (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000.		253		
	A hurts more	51%		51	20%
	B hurts more	38%		184	73%

Problem	Conditions and Options	Original		Replication	
		Percentage	*N*	Count	Percentage
	No difference	21%		18	7%
	Second group of questions:				
	1. (A) You lose $9. (B) You lose $9 after suffering a loss of $9.				
	A hurts more	7%	252	9	4%
	B hurts more	64%		209	83%
	No difference	28%		34	13%
	2. (A) You lose $9. (B) You lose $9 after suffering a loss of $18.				
	A hurts more	11%	252	9	4%
	B hurts more	65%		231	92%
	No difference	23%		12	5%
	3. (A) You lose $9. (B) You lose $9 after suffering a loss of $36.				
	A hurts more	12%	252	11	4%
	B hurts more	62%		228	90%
	No difference	26%		13	5%
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $45.				
	A hurts more	14%	252	10	4%
	B hurts more	65%		230	91%
	No difference	21%		12	5%
	5. (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36.				
	A hurts more	7%	252	16	6%
	B hurts more	68%		224	89%
	No difference	25%		12	5%
9^b	**Giving away:**				
	$0	30%	254	75	30%
	$20	17%		52	20%
	$20 plus interest	9%		14	6%
	$75	30%		53	21%
	A $55 saving	14%		60	24%
	Drinking:				
	$0	30%	251	57	23%
	$20	18%		54	22%
	$20 plus interest	7%		16	6%
	$75	20%		64	25%
	A $55 saving	25%		60	24%
11^c	**$50 high cost condition:** **Spent $50 on dinner. Would you buy a $25 theater ticket later in the week?**		254		
	Buy the ticket	/		101	40%
	Not buying the ticket	/		153	60%

Problem	Conditions and Options	Original		Replication	
		Percentage	*N*	Count	Percentage
	Given a $50 dinner. Would you buy a $25 theater ticket-later in the week?		254		
	Buy the ticket	/		211	83%
	Not buying the ticket	/		43	17%
	Spent $50 on a sports ticket. Would you purchase a $25 theater ticket later in the week?		254		
	Buy the ticket	/		70	28%
	Not buying the ticket	/		184	72%
	Given a $50 sports ticket. Would you purchase a $25 theater ticket later in the week?		254		
	Buy the ticket	/		198	78%
	Not buying the ticket	/		56	22%
	Spent $50 on an inoculation. Would you purchase a $25 theater ticket later in the week?		254		
	Buy the ticket	/		75	30%
	Not buying the ticket	/		179	70%
	$20 low cost condition:				
	Spent $20 on dinner. Would you buy a $25 theater ticket later in the week?		252		
	Buy the ticket	/		138	55%
	Not buying the ticket	/		114	45%
	Given a $20 dinner. Would you buy a $25 theater ticket-later in the week?		252		
	Buy the ticket	/		213	85%
	Not buying the ticket	/		39	15%
	Spent $20 on a sports ticket. Would you purchase a $25 theater ticket later in the week?		252		
	Buy the ticket	/		103	41%
	Not buying the ticket	/		149	59%
	Given a $20 sports ticket. Would you purchase a $25 theater ticket later in the week?		252		
	Buy the ticket	/		201	80%
	Not buying the ticket	/		51	20%
	Spent $20 on an inoculation. Would you purchase a $25 theater ticket later in the week?		252		
	Buy the ticket	/		109	43%
	Not buying the ticket	/		143	57%

Note. N represents sample size. [a] The statistical results reported in Problem 6-Condition A-5 added up to 110% rather than 100%, suggesting a possible reporting mistake in the original article. [b] In Problem 9, the mean for the "I don't understand" option was only 1.16, indicating that participants had a good understanding of the materials on average. [c] For Problem 11, Heath and Soll (1996) revealed that a larger proportion of people are more likely to underconsume in the $50 high-cost than in the $20 low-cost condition ($t(26)=2.17$, $p<.05$ by paired *t*-test). The proportion of subjects who underconsume the target is highly correlated with typicality for both $50 high-cost ($r(25)=.80$, $p<.01$) and $20 low-cost conditions ($r(25)=.67$, $p<.01$). Yet we are unsure about the paired t-test reported in Problem 11 as the experiment seems to adopt a between-subject design

Table 10

Comparison and descriptive statistics for Problems with options and a within-subjects design

Problem	Sub Questions and Options	Original		Replication	
		Percentage	*N*	Count	Percentage
4	**1. Two wins: $50, and $25 versus One win: $75. Who was happier?**				
	Two wins is happier	64%	505	178	35%
	One win is happier	18%		62	12%
	No difference	17%		265	52%
	2. Two mistakes: $100, and $50 versus One mistake: $150. Who was more upset?				
	Two mistakes is more upset	76%	505	213	42%
	One mistake is more upset	16%		69	14%
	No difference	8%		223	44%
	3. Two events: Win $100, and loss $80 versus One event: Win $20. Who was happier?				
	Two events is happier	25%	505	61	12%
	One event is happier	70%		386	76%
	No difference	5%		58	11%
	4. Two events: Loss $200, and win $25 versus One event: Loss $175. Who was more upset?				
	Two events is more upset	22%	505	63	12%
	One event is more upset	73%		316	63%
	No difference	6%		126	25%
5	**A. Two events: (1) win $25 (2) win $50. Who is happier?**				
	Happier on the same day	25%	495	200	40%
	Happier two weeks apart	63%		174	35%
	No difference	12%		121	24%
	B. Two events: (1) $100 must be paid (2) $50 must be paid. Who is more unhappy?				
	More unhappy on the same day	57%	495	303	61%
	More unhappy two weeks apart	34%		96	19%
	No difference	9%		96	19%
	C. Two events: (1) a $20 parking ticket (2) a $25 bill. Who is more unhappy?				
	More unhappy on the same day	75%	495	278	56%
	More unhappy two weeks apart	17%		115	23%
	No difference	7%		102	21%
13	**Imagine that you have just won $30.**				
	A 50% chance to gain $9 and a 50% chance to lose $9.	70%	504	143	28%
	No further gain or loss.	30%		361	72%

Problem	Sub Questions and Options	Original		Replication	
		Percentage	*N*	Count	Percentage
14	**Imagine that you have just lost $30.**				
	A 50% chance to gain $9 and a 50% chance to lose $9.	40%	504	85	17%
	No further gain or loss.	60%		419	83%
15	**Imagine that you have just lost $30.**				
	A 33% chance to gain $30 and a 67% chance to gain nothing.	60%	504	119	24%
	A sure $10.	40%		385	76%
16[a]	**A single coin flip, heads you win $200, tail you lose $100. Would you take the bet?**	/	506		
	Take the bet.			129	25%
	Not taking the bet	/		377	75%
	A package bet of 100 coin flips, each coin flip you either win $200 or lose $100. Would you take the bet?	/	506		
	Take the bet.			247	49%
	Not taking the bet	/		259	51%
17[b]	**A project:** 50% chance to gain $2 million, 50% chance to lose $1 million.		504		
	Undertake the project	/		152	30%
	Not undertaking the project	/		352	70%
	A portfolio of 25 of investments: Each has a 50% chance of gaining $2 million and 50% chance of losing $1 million.		504		
	Undertake the investments	12%		236	47%
	Not undertaking the investments	/		268	53%
18	**Paid $40 for tickets:**				32%
	Go to the game	/	502	160	
	Stay home	/		342	68%
	Tickets given by friends:				
	Go to the game	/	502	68	14%
	Stay home	/		434	86%
19	**Imagine that you joined a tennis club and paid a $300 yearly membership fee.**		502		
	Stop playing	/		380	76%
	Continue to play	/		122	24%

Note. N represents sample size. [a] In Problem 16, the economist answered No for the single coin flip, and Yes for playing the bet 100 times. [b] In Problem 17, 3 of the 25 executives accepted the single investment, and the CEO accepted the portfolio of 25 of these investments.

Table 11

Comparison and descriptive statistics for Problems with scale/text entry and a between-subjects design

Problem	Condition (if applicable)	Original findings	Replication		
			N	Mean	Standard deviation
7[a]	Hotel (fancy) purchase condition	*Median*=$2.65	254	7.09	19.85
	Grocery (run-down) purchase condition	*Median*=$1.50	254	4.17	3.98
8	Free condition-Market value $5-Friend	68% answer 0, 26% answer 5, 3% answer 10, and 3% answer Other	166	2.37	3.49
	Free condition-Market value $5-Stranger	6% answer 0, 77% answer 5, 10% answer 10, and 6% answer Other	166	6.04	5.24
	Free condition-Market value $10-Friend	65% answer 0, 26% answer 5, 6% answer 10, and 3% answer Other	166	4.26	5.43
	Free condition-Market value $10-Stranger	6% answer 0, 16% answer 5, 58% answer 10, and 19% answer Other	166	10.32	6.79
	Paid 5 condition-Market value $5-Friend	14% answer 0, 79% answer 5, 0% answer 10, and 7% answer Other	169	3.72	2.47
	Paid 5 condition-Market value $5-Stranger	0% answer 0, 79% answer 5, 7% answer 10, and 14% answer Other	169	6.68	3.49
	Paid 5 condition-Market value $10-Friend	7% answer 0, 79% answer 5, 4% answer 10, and 9% answer Other	169	6.12	4.20
	Paid 5 condition-Market value $10-Stranger	0% answer 0, 14% answer 5, 57% answer 10, and 29% answer Other	169	11.51	5.59
	Paid 10 condition-Market value $5-Friend	0% answer 0, 69% answer 5, 23% answer 10, and 8% answer Other	162	5.01	2.68
	Paid 10 condition-Market value $5-Stranger	0% answer 0, 42% answer 5, 46% answer 10, and 12% answer Other	162	8.19	3.58
	Paid 10 condition-Market value $10-Friend	0% answer 0, 15% answer 5, 69% answer 10, and 15% answer Other	162	7.55	3.82

Problem	Condition (if applicable)	Original findings	Replication		
			N	Mean	Standard deviation
	Paid 10 condition-Market value $10-Stranger	0% answer 0, 0% answer 5, 73% answer 10, and 27% answer Other	162	11.43	4.01
12	$15 condition	/	235	8.14	11.32
	$40 condition	People are willing to pay twice as much to avoid waiting for the $40 ticket than for the $15 ticket ($\underline{X}$=$7.20 vs. $\underline{X}$=$3.60, t=1.92(39), p=.06).	222	10.34	7.66
21	Day expression	/	167	44.53	32.32
	Year expression	/	166	26.04	28.05
	Both expressions-Day	/	170	45.16	31.71
	Both expressions-Year	/	170	35.72	29.05

Note. [a] A caveat needed to be noted for Problem 7, as there are many major outliers in the Hotel condition.

Table 12

Comparison and descriptive statistics for Problems with scale and a within-subjects design

Problem	Condition (if applicable)	Original findings	Replication		
			N	Mean	Standard deviation
10[a]	I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway.	*Mean*=3.31	502	2.98	1.45
	I feel like I made a $400 investment which I will gradually consume after a period of years.	*Mean*= 1.94	502	3.56	1.30
	I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered.	*Mean*=2.88	502	3.08	1.36
	I cannot understand this question.	/	502	1.16	0.65
20	The more you paid for the shoes, the more times you will try to wear them.	/	507	3.10	1.41
	Eventually you stop wearing the shoes, but you do not throw them away. The more you paid for the shoes, the longer they sit in the back of your closet before you throw them away.	/	507	3.45	1.33

Note. N represents sample size.

[a] For Problem 10, the original study used the Likert Scale with a 1 *(strongly agree) to* 5 *(strongly disagree)* rating and the replication used a 1 (*not accurate at all*) to 5 (*very accurate*) rating.

Table 13

Summary of all chi-square results

Problem	**χ^2**	**χ^2 continuity correction**	**Difference in 2 proportions** [95% CI]	**Odds ratio** [95% CI]	**Phi-coefficient**	**Cramer's V**
1	85.03 (df=1, p <.001)	83.33 (df=1, p <.001)	-0.40 [-0.47; -0.32]	0.15 [0.10; 0.23]	0.41	0.41
2	3.39 (df=1, p = .066)	2.85 (df=1, p = 0.092)	0.05 [-0.00; 0.10]	1.79 [0.96; 3.35]	0.08	0.08
3	16.18 (df=1, p < .001)	15.22 (df=1, p <.001)	0.13 [0.07; 0.20]	2.79 [1.67; 4.67]	0.18	0.18
9	3.64 (df=4, p = .457)	3.64 (df=4, p = .457)	/	/	/	0.08

Note. df indicates degree of freedom and CI indicates Confidence Interval.

Table 14

Summary of χ^2 Goodness of Fit in all proportion tests

Problem	Conditions and sub questions	χ^2	df	p
4	1. Two wins: \$50, and \$25 versus One win: \$75. Who was happier?	123.24	2	< .001
	2. Two mistakes: \$100, and \$50 versus One mistake: \$150 Who was more upset?	88.22	2	< .001
	3. Two events: Win \$100, and loss \$80 versus One event: Win \$20 Who was happier?	422.21	2	< .001
	4. Two events: Loss \$200, and win \$25 versus One event: Loss \$175 Who was more upset?	206.10	2	< .001
5	A. Two events: (1) win \$25 (2) win \$50 Who is happier?	19.65	2	< .001
	B. Two events: (1) \$100 must be paid (2) \$50 must be paid. Who is more unhappy?	173.13	2	< .001
	C. Two events: (1) a \$20 parking ticket (2) a \$25 bill Who is more unhappy?	116.59	2	< .001
6	First group of questions: 1. (A) You lose \$9. (B) You lose \$9 after having gained \$30.	274.16	2	< .001
	2. (A) You lose \$9. (B) You lose \$9 after having lost \$30.	317.84	2	< .001
	3. (A) You lose \$9. (B) You lose \$9 after having lost \$250.	278.81	2	< .001
	4. (A) You lose \$9. (B) You lose \$9 after suffering a loss of \$1000.	244.96	2	< .001
	5. (A) You lose \$9 after suffering a loss of \$30. (B) You lose \$9 after suffering a loss of \$1000.	183.14	2	< .001
	Second group of questions: 1. (A) You lose \$9. (B) You lose \$9 after suffering a loss of \$9.	282.74	2	< .001
	2. (A) You lose \$9. (B) You lose \$9 after suffering a loss of \$18.	385.93	2	< .001
	3. (A) You lose \$9. (B) You lose \$9 after suffering a loss of \$36.	370.31	2	< .001
	4. (A) You lose \$9. (B) You lose \$9 after suffering a loss of \$45.	380.67	2	< .001
	5. (A) You lose \$9 after suffering a loss of \$9. (B) You lose \$9 after suffering a loss of \$36.	350.10	2	< .001
19	Imagine that you joined a tennis club and paid a \$300 yearly membership fee.	132.60	1	< .001

Note. df indicates degree of freedom

Table 15

Summary of all McNemar paired-samples tests

Problem	Comparisons	χ^2	df	p	Log odds ratio exact	Interpretation
5	Comparing A to B A: (1) win $25 (2) win $50 Who is happier? B: (1) $100 must be paid (2) $50 must be paid. Who is more unhappy?	46.74	3	< .001	/	A-B Combined: Signal, same direction
	Comparing A to C A: (1) win $25 (2) win $50 Who is happier? C: (1) a $20 parking ticket (2) a $25 bill Who is more unhappy?	38.78	3	< .001	/	A-C Combined: Signal, same direction
14	Problem 13 vs. 14	22.73	1	< .001	0.83 (p < .001)	Signal, same direction
15	Problem 13 vs. 15	3.27	1	0.070	0.27 (p = 0.083)	No signal, same direction
16	A single bet vs. 100 bets	84.90	1	< .001	-1.81 (p < .001)	Signal, same direction
17	A project vs 25 investments	45.82	1	< .001	-1.22 (p < .001)	Signal, same direction
18	Paid $40 vs. Given by friends	86.37	1	< .001	3.46 (p < .001)	Signal, same direction

Note. df indicates degree of freedom. Study 5 compared same day to two weeks apart, higher same day for negative than for positive.

Table 16

Summary of all t-tests results

Problem (test type)	Statistic	df	*p*	Mean difference	SE difference	Effect size (Cohen's *d*)	95% CI	Interpretation
7 Independent samples t-test	Student's *t*=2.30	506	.011	2.92	1.27	0.20	[0.03, 0.38]	Signal, same direction
	Welch's *t*=2.30	273.35	.011	2.92	1.27	0.20	/	
12 Independent samples t-test	Student's *t*=2.42	455	.016	2.20	0.91	0.23	[0.04, 0.41]	Signal, same direction weaker effect
	Welch's *t*=2.45	412.97	.015	2.20	0.90	0.23	/	
20 One-sample t-test	Statement 1: Student's *t*=1.64	506.00	.051	0.10	/	0.07	[-0.01, 0.16]	Combined: Signal, same direction
	Statement 2: Student's *t*=7.53	506.00	< .001	0.45	/	0.33	[0.24, 0.42]	
21 Independent samples t-test	Student's *t*=5.57	331	< .001	18.48	3.32	0.61	[0.39, 0.83]	Combined: Signal, same direction
	Welch's *t*=5.57	325.07	< .001	18.48	3.32	0.61	/	
21 Paired sample t-test	Student's *t*=3.82	169	< .001	9.44	2.47	0.29	[0.14, 0.45]	

Note. df indicates degree of freedom, SE indicates standard error, and CI indicates confidence interval.

Table 17

Summary of all ANOVA results

Problem 8-Mixed ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	*F*	*p*	η^2
(Friend vs. Stranger)	8718.31	1	8718.31	594.75	< .001	0.16
(Friend vs. Stranger) ∗ (Free ticket vs. $5 vs. $10)	146.59	2	73.29	5.00	0.007	0.00
Residual	7241.43	494	14.66			
(Market worth $5 vs. $10)	5072.36	1	5072.36	870.07	< .001	0.09
(Market worth $5 vs. $10) ∗ (Free ticket vs. $5 vs. $10)	47.91	2	23.95	4.11	0.017	0.00
Residual	2879.92	494	5.83			
(Friend vs. Stranger) ∗ (Market worth $5 vs. $10)	418.89	1	418.89	205.04	< .001	0.01
(Friend vs. Stranger) ∗ (Market worth $5 vs. $10) ∗ (Free ticket vs. $5 vs. $10)	79.70	2	39.85	19.50	< .001	0.00
Residual	1009.22	494	2.04			

Between Subjects Effects

	Sum of Squares	df	Mean Square	*F*	*p*	η^2
(Free ticket vs. $5 vs. $10)	1736.17	2	868.08	15.69	< .001	0.03
Residual	27323.41	494	55.31			

Problem 10-Repeated measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	*F*	*p*	η^2
Feeling	98.08	2	49.04	25.26	< .001	0.03
Residual	1945.26	1002	1.94			

Problem 11-Mixed ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	*F*	*p*	η^2
(Given/Spent)	83.36	1	83.36	382.59	< .001	0.17
(Given/Spent) ✻ (high vs. low cost)	1.98	1	1.98	9.11	0.003	0.00
Residual	109.81	504	0.22			
(Dinner/Ticket)	4.09	1	4.09	55.52	< .001	0.01
(Dinner/Ticket) ✻ (high vs. low cost)	0.01	1	0.01	0.08	0.783	0.00
Residual	37.15	504	0.07			
(Given/Spent) ✻ (Dinner/Ticket)	0.83	1	0.83	16.49	< .001	0.00
(Given/Spent) ✻ (Dinner/Ticket) ✻ (high vs. low cost)	0.01	1	0.01	0.26	0.610	0.00
Residual	25.41	504	0.05			

Between Subjects Effects

	Sum of Squares	df	Mean Square	*F*	*p*	η^2
(high vs. low cost)	3.15	1	3.15	7.37	0.007	0.01
Residual	215.53	504	0.43			

Note. Type 3 Sums of Squares. df represents degree of freedom.

Exploratory analysis

In the pre-registration, we planned to conduct exploratory analyses if we failed to

replicate the original findings. Problems 9, 10, and 21, all replicated successfully, so there was no need. We conducted additional exploratory analyses for Problems 4, 7, 14, and 15 to try and probe possible factors affecting the outcomes.

For Problem 4, we conducted further proportion tests to compare the two "indifferent options", and summarized the results in Table 18. For Problem 7, considering that the Levene's test was significant, we re-ran the independent samples t-test after excluding outliers (answers that are 3 standard deviations above the mean), and summarized statistics and statistical tests in Table 19. For Problems 14 and 15, we added a McNemar paired-samples test and summarized the results in Table 20.

Table 18

Descriptives and statistical results for Problem 4

					95% Confidence Interval	
Options	**N**	**Count**	**Percentage**	**p**	**Lower**	**Upper**
Two wins: $50, and $25 versus One win: $75. Who was happier?	240					
Two wins is happier		178	74%	<.001	0.68	0.80
One win is happier		62	26%	<.001	0.20	0.32
Two mistakes: $100, and $50 versus One mistake: $150. Who was more upset?	282					
Two mistakes is more upset		213	76%	<.001	0.70	0.80
One mistake is more upset		69	24%	<.001	0.20	0.30
Two events: Win $100, and loss $80 versus One event: Win $20. Who was happier?	447					
Two events is happier		61	14%	<.001	0.11	0.17
One event is happier		386	86%	<.001	0.83	0.89
Two events: Loss $200, and win $25 versus One event: Loss $175. Who was more upset?	379					
Two events is more upset		63	17%	<.001	0.13	0.21
One event is more upset		316	83%	<.001	0.79	0.87

Note. N represents sample size.

Table 19

Descriptives and statistical results Problem 7

Descriptives						
Condition	N	Mean	Standard Deviation			
Hotel condition	252	5.56	4.32			
Grocery condition	248	3.67	2.20			
Independent samples t-test						
Statistic	df	*p*	Mean difference	SE difference	Effect size (Cohen's *d*)	95% CI
Student's *t*=6.15	498	< .001	1.89	0.31	0.55	[0.37; 0.73]
Welch's *t*=6.18	374.55	< .001	1.89	0.31	0.55	/

Note. N represents sample size. df indicates degree of freedom, SE indicates standard error, and CI indicates confidence interval.

Table 20

McNemar test results for comparing Problem 14 and 15

	Value	**df**	***p***
χ^2	8.38	1	0.004
Log odds ratio exact	-0.50		0.005

Note. df indicates degree of freedom.

Moreover, we aimed to examine the intercorrelations among the mental accounting problems. We conducted a pooled analysis for all Problems meeting the following criteria: 1) adopted a within-subject design, 2) had choice questions, and 3) were directly related to mental accounting. We coded the value as 1 when the answer seemed affected by mental accounting, and coded 0 otherwise. The correlations table among the coding of the problems is summarized in Table 21. This exploratory analysis was an innovative and preliminary attempt to study the connections among different subsets of the mental accounting framework. The results indicated that further explorations hold some promise.

Table 21

Correlations among mental accounting problems

Variables	4_1	4_2	4_3	4_4	15	16_2	17_2	18_1
Problem 4_1	–							
Problem 4_2	0.36***	-						
Problem 4_3	0.14**	0.20***	-					
Problem 4_4	0.21***	0.29***	0.26***	-				
Problem 15	0.02	-0.02	-0.11	-0.01	-			
Problem 16_2	0.14*	0.12	0.12	0.08	0.09	-		
Problem 17_2	0.08	0.02	0.02	-0.13	0.06	0.55***	-	
Problem 18_1	-0.07	-0.05	-0.02	0.07	0.14*	0.01	0.15*	-

Note. *indicates p <0.5, **indicates p <0.1, ***indicates p <0.001

Comparing replication to original findings

We planned to evaluate the replication effect based on LeBel et al. 's (2019) framework in the pre-registration. However, given that the current replication adopted many different statistical analysis approaches from the original studies, a direct use of LeBel et al. 's (2019) outcome interpretation criteria was not easily applicable. Therefore, we only compared the direction and relative magnitude of the mental accounting effects in some of the problems where it seemed to be meaningful.

Table 22

Comparison of replication and original findings

Problem (design)	Condition/Sub questions	Original *Cohen's h, [95%CI]*	Replication *Cohen's h, [95%CI]*	Interpretation
1 (between)	Gain condition	0.46 [0.28, 0.63]	0.75 [0.62, 0.87]	consistent, stronger
	Loss condition	-0.28 [-0.46 -0.11]	-0.12 [-0.24, 0.00]	consistent, weaker
	Combined (gain vs. loss)		$\chi^2 = 85.03$, $p < .001$ V = 0.41	**combined: signal, same direction**
2 (between)	$15 Calculator condition	0.37 [0.17, 0.57]	-0.89 [-1.02, -0.77]	inconsistent, opposite
	$125 Calculator	-0.43 [-0.64, -0.22]	-1.04 [-1.16, -0.91]	consistent, stronger
	Combined ($12 vs. $125)		$\chi^2 = 3.39$, $p = .066$ V = 0.08	no signal, same direction
3 (between)	Lost a $10 bill condition	0.86 [0.72, 1.01]	0.93 [0.80, 1.05]	consistent
	Lost the ticket condition	-0.08 [-0.22, 0.06]	0.57 [0.45, 0.69]	inconsistent, signal
	Combined (Lost $10 vs. Lost ticket)		$\chi^2 = 16.18$, $p < .001$ V = 0.18	**combined: signal, same direction**
4 (within)	4_1 Segregate gains -Two wins is happier	0.64 [0.43, 0.85]	0.04 [-0.04, 0.13]	inconsistent, no signal
	4_2 Integrate loss -Two mistakes is more upset	0.89 [0.68, 1.10]	0.19 [0.10, 0.27]	**signal, weaker**
				combined: consistent
	4_3 Cancel losses against larger gains -One event is happier	0.76 [0.55, 0.97]	0.89 [0.81; 0.98]	**consistent, stronger**
	4_4 Segregate "silver linings" -One event is more upset	0.81 [0.60, 1.02]	0.61 [0.52; 0.70]	**consistent, weaker**
				combined: consistent
5 (within)	5A Prefer segregation -Happier two weeks apart	0.61 [0.37, 0.85]	0.04 [-0.05, 0.13]	inconsistent, no signal
	5B Prefer integration -More unhappy two weeks apart	0.02 [-0.22, 0.26]	-0.32 [-0.41, -0.23]	inconsistent signal
	5C Prefer integration -More unhappy two weeks apart	-0.37 [-0.62, -0.13]	-0.22 [-0.31, -0.14]	**consistent, weaker**
				combined A-B & A-C: consistent

Problem (design)	**Condition/Sub questions**	**Original** *Cohen's h, [95%CI]*	**Replication** *Cohen's h, [95%CI]*	**Interpretation**
6 (between)	First group of questions: 1. (A) You lose $9. (B) You lose $9 after having gained $30. -$9 hurts more in A	0.76 [0.55, 0.97]	1.04 [0.92, 1.16]	consistent, stronger
	2. (A) You lose $9. (B) You lose $9 after having lost $30 -$9 hurts more in A	-0.49 [-0.70, -0.28]	-0.65 [-0.77, -0.53]	consistent, stronger
	3. (A) You lose $9. (B) You lose $9 after having lost $250. -$9 hurts more in A	0.13 [-0.09, 0.34]	-0.49 [-0.61, -0.36]	inconsistent, opposite
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $1000. -$9 hurts more in A	0.35 [0.14, 0.56]	-0.43 [-0.55, -0.31]	inconsistent, opposite
	5. (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000. -$9 hurts more in A	0.37 [0.16, 0.58]	-0.30 [-0.42, -0.17]	inconsistent, opposite
	Second group of questions: 1. (A) You lose $9. (B) You lose $9 after suffering a loss of $9. -$9 hurts more in A	-0.69 [-0.91, -0.47]	-0.82 [-0.94, -0.70]	consistent, stronger
	2. (A) You lose $9. (B) You lose $9 after suffering a loss of $18. -$9 hurts more in A	-0.55 [-0.77 -0.33]	-0.82 [-0.94, -0.70]	consistent, stronger
	3. (A) You lose $9. (B) You lose $9 after suffering a loss of $36. -$9 hurts more in A	-0.52 [-0.73, -0.30]	-0.82 [-0.94, -0.70]	consistent, stronger
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $45. -$9 hurts more in A	-0.46 [-0.67, -0.24]	-0.82 [-0.94, -0.70]	consistent, stronger

Problem (design)	**Condition/Sub questions**	**Original** *Cohen's h, [95%CI]*	**Replication** *Cohen's h, [95%CI]*	**Interpretation**
	5. (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36. -$9 hurts more in B	0.72 [0.50, 0.93]	1.24 [1.12, 1.37]	consistent, stronger
13 (within)	A 50% chance to gain $9 and a 50% chance to lose $9.	0.41	-0.46 [-0.54, -0.37]	inconsistent
14 (within)	No further gain or loss.	0.20	0.72 [0.63, 0.81]	consistent, stronger
15 (within)	A 33% chance to gain $30 and a 67% chance to gain nothing.	0.20	-0.55 [-0.63, -0.46]	inconsistent

Note. Please refer to the supplementary for the calculations of effect sizes and 95% confidence intervals

Discussion

We carried out a well-powered pre-registered replication of the classical mental accounting effects reviewed by Thaler (1999). More than twenty years after the publication of Thaler's review paper, we were able to find support for 12 out of the 17 replication Problems, indicating high reliability of the mental accounting phenomenon. Specifically, we found consistent results for Problems 1, 3, 4, 7, 8, 9, 10, 12, 14, 15, 16, and 17. Yet the results for Problems 5, 6, and 11 were mixed, and the results of Problems 2 and 13 were inconsistent with the original findings. In the following section we evaluate the consistencies and inconsistencies between the current replication and the original studies, and review the results of the extensions. We then discuss the limitations and promising future directions.

Replication

Problems with consistent results

The results of Problems 1, 3, 4, 5, 7, 8, 9, 10, 12, 14, 15, 16, and 17 were in line with the original studies.

Among the Problems, Problems 1, 14, and 15 explored people's risk-taking attitudes. In Problem 1, a vast majority of the participants demonstrated a risk-averse tendency towards gains, with the effect stronger than the original. In contrast, only a small majority of participants displayed risk-seeking preferences toward loss. Problems 14 and 15 delved further to test the impact of a previous loss on subsequent risk-taking behavior, and the results were in the expected direction. The earlier loss could not induce risk-seeking in both Problems. However, when given the opportunity to break even, more people were willing to take the risk.

Problem 3 revealed clear support for the well-established sunk cost effect. It is observed that 90% of the participants were willing to buy the ticket if they had lost a $10 bill. In contrast, if the participants had lost the same $10 ticket, they became less inclined to buy a second ticket (77%). Though the replication effect was of a weaker magnitude compared to the original, it can be seen that the sunk cost effect emerged when the two activities were in the same mental account (Tversky & Kahneman, 1981).

The hedonic editing effect was investigated in Problem 4, and the participants responded in the way predicted by the hypothesis. However, the tendencies to separate gains, integrate losses, and segregate "silver linings" were all weaker than in the original studies. Additionally, much more participants perceived the options as indifferent in the current replication.

Though with slightly different focuses, Problems 7, 8, 9, 10, and 12 all supported how mental accounting allows people to value things in flexible and fluid ways (Shafir & Thaler, 1998). When buying, people were willing to pay higher prices for the same soda in a fancy resort hotel than in a grocery store *(Problem 7)*. Also, their willingness to pay was substantially higher for higher-priced tickets *(Problem 12)*. When selling, people asked for higher prices when 1) their costs were higher; 2) the market price was higher; 3) the buyer was a stranger instead of a friend *(Problem 8)*. These together verified the powerful effect of reference points and the determinant factors Thaler (1985) proposed. Problems 9 and 10 went deeper to examine people's perceptions of value when the purchase and consumption is temporarily separated (Shafir & Thaler, 1998). For a purchase to be consumed in the near future, as in the original, there was a lack of consensus of the item value *(Problem 9)*. Conversely, for a purchase to be consumed in the distant future, people would consider it as an investment *(Problem 10)*. People tend to evaluate the value of time and items in relative terms and are sensitive to price anchors (Seymour & McClure, 2008).

Problems 16 and 17 were powerful illustrations of the myopic loss aversion effect. Participants were more willing to take risks when there was a package of 100 bets (Problem 16), or a portfolio of 25 investments (Problem 17). When the risky episodes are bracketed together, people do not evaluate the events in isolation.

Though the magnitudes of the effects were different, it is still safe to conclude that the current project successfully replicated the above Problems. The replication success could be attributable to the methodological similarities. The materials adopted remained largely the same and changes were only made to enhance clarity. Also, it is unlikely that the differences in participant recruitment will have an impact on the results.

Problems with mixed support

Evidence regarding Problems 5, 6, and 11 was mixed.

Problems 5 and 6 offered new findings that were not entirely consistent with the original's findings and the hedonic editing hypothesis. Per each of the questions, participants did not prefer to spread out gains or to integrate losses in Problem 5, though if we take a wider view on the comparisons between gains and losses, they seemed to be in the same direction as in the original. So, it is possible that this simply represents a shifting of the preferences for this specific sample, compared to the original's.

Further zooming in to the loss integration principle, Problem 6 yielded surprising results. Out of a total of 10 questions, we found support for 7 of them. In agreement with the original findings, people actively integrated the loss of $9 into prior gains but not into prior losses. However, unlike the original, regardless of the magnitude of the previous loss, people seemed more loss averse after the loss.

To summarize, the hedonic editing hypothesis was only partially supported under the particular methodology and context. Together with Problem 4, these ambivalent results call for a more precise notion to fully capture the complexity of the hedonic editing effect. Replication, therefore, is an important method to set limits on certain effects.

Problem 11 focused on the mental-budgeting effect. As expected, the budgeting process led to greater underconsumption for all three activities when the previous expenditure was higher. Yet the process did not stimulate greater underconsumption for more typical activities. This once again confirms the inherent complexity of the mental accounting framework.

Problems with inconsistent results

Results from Problems 2 and 13 were in conflict with the original research findings.

Problem 2 examined people's perceptions of the value of time. The majority of the participants were unwilling to drive 20 minutes to save $5, regardless of the price. An important fact must be addressed to interpret the results. As pointed out by one of the participants in the feedback section, it may cost more than $5 to drive 20 minutes to the other store with the increasing cost of driving. Therefore, the inconsistency may be due to participants' awareness of driving costs rather than a lack of mental accounting effect. It is also possible that over time since it was conducted, the value of $5 in relation to transport costs has changed dramatically enough to shift participants' preferences entirely. It could also be due to our sample's demographics compared to the original's. Future replications should take this into consideration and make justifiable changes.

Problem 13 was a question on risk attitudes, with findings that differed from the original claim, with the effect in a completely opposite direction. The prior gain failed to trigger risk-seeking behavior as anticipated. In fact, the inconsistent result is in congruence with the long-held

debate regarding the direction of the impact (Merkle et al., 2021). According to Merkle and colleagues (2021), both risk-seeking and risk-averse behaviors after gains are justifiable. They argued that people can be motivated to be risk-seeking by the house money effect and the hedonic editing hypothesis, or be motivated to avoid risk by the prospect theory. Further research is needed to come up with a more unified explanation for this, and Imas (2016) and Merkle et al. (2022) have already made promising contributions by suggesting the realization effect.

Extension

Beyond the replication Problems, we also ran four extensions to examine the predictions Thaler made. Among them, we found empirical support for Problem 18 and 21, mixed support for Problem 20, and no support for Problem 19.

Problems 18, 19, and 20 all targeted the sunk cost effect. In Problem 18, more participants chose to go to the game when the ticket was bought by themselves. In Problem 19, the large majority of the participants will not continue playing after developing tennis elbow despite the expensive $300 membership fee. In Problem 20, participants agreed that they tend to keep the uncomfortable shoes longer when the price is higher. Taking into account the replication success of Problem 3, these together revealed that the sunk cost effect might be context-based.

The pennies-a-day effect was validated in Problem 21. Within and across conditions, the "merely 27 cents a day" plan was rated as more appealing than the "100 US$ a year" plan. The price frames appeared to affect the comparability of the offers, where expressing the price on a per-day basis helps to lower participants' price sensitivity (Chioveanu & Zhou, 2013).

Replications of an entire literature as reflect by a review article

This replication project differs from a typical replication registered report, in that rather than focusing on a single empirical study or article we targeted a review article which covered an entire body of literature on multiple related phenomena with empirical demonstrations from multiple seminal articles. Rather than replicating each independently using different samples, we combined all the studies into a single data collection and mapped out all the effects. This is a useful approach for many reasons, given that the one sample allows us to focus on comparing empirical designs and effect sizes, and to try and map links between the different studies. We provided one such initial analysis focusing on a subset using a similar design, yet the dataset made available allows others to continue this work and explore further.

In addition, many review articles, especially when conducted by those whose studies it covers - like Thaler - often reports anecdotal evidence and untested assumptions and predictions, in hope of making the implicit more explicit, and motivating future directions to test those predictions. However, assumptions in review papers at times become institutionalized to the point of being taken for granted, and predictions made are not picked up and empirically tested. Replications of a review article can help tackle that, by systematically mapping claims that can be empirically tested, even if there was no empirical test associated with and testing those assumptions or predictions. We hope to see more replication efforts of review papers, taking a similar approach to the one we present here.

Limitations and directions for future research

Despite our best efforts to follow the original studies as closely as possible, our replication differed from the originals in several ways and we had to make many adjustments and analytical decisions. Many of the original studies only reported descriptive statistics and there

were ambiguities regarding the exact analysis used. Take Problem 10 as an example, Shafir and Thaler (1998) did not specify which type of t-test they employed to compare the options. As a result, we deduced a set of comprehensive statistical analyses that we thought would help better interpret the answers. The lacking analytical details of the original studies raises the importance of reproductions and replications of old studies to facilitate a more transparent sharing of methods, data, code, and the documentation needed to facilitate reproducible replicable future research (Alston & Rick, 2021). We tried our best to compare the original's findings to the replication's, yet given our reconstruction and adjustments to the data analysis, we caution regarding over interpreting the comparisons between the replication results and the original effects.

We recruited a much larger and more diverse sample than the original studies, yet our participants were exclusively from the US and from an online sample. With all participants coming from the same notably WEIRD country, there is doubt about the demographic representativeness of the project (Simons et al., 2017). Considering the different financial beliefs and habits in other societies, it would be beneficial to recruit participants from non-western countries to explore the cross-cultural reliability of the mental accounting phenomenon. In addition, we note that the data collection for this project was conducted during the covid-19 pandemic. Though we found support for most studies, our participants may show different risk-seeking behaviors compared to non-pandemic periods. As pointed out by Yue et al. (2020), households altered their risk preference and became more risk-averse due to the pandemic. Thus, the temporal specificity sets another constraint on generalizability.

Finally, from a broader perspective, previous research and the current project focused predominantly on different components in the mental accounting theory. In this project, we aimed to systematically revisit experiments testing different accounts of the mental accounting

framework reviewed by Thaler (1999). We focused on the empirical aspects of the singular problems, and did not go further to discuss implications for mental accounting theory as a whole, which was beyond the scope of this investigation. Therefore, the results of our replications for each of the problems should be interpreted separately and cautiously, and we encourage future research to go further and attempt broader theoretical integrations.

Conclusion

We examined the replicability of the mental accounting studies summarized by Thaler (1999). We successfully replicated 12 Problems, found mixed support for 3 Problems, and failed to find support for 2 Problems. Through the replication and extension, we examined the stability of the mental accounting effects and revealed that some effects were more complex than originally documented. We see much promise in further studies of the links among the different aspects of the mental accounting framework. We believe our reconstruction and reanalysis of classic experiments as well as our exploratory analyses could provide an inspiration and practical tools to stimulate further follow-up research to examine the mental accounting phenomenon as a whole.

References

Alston, J. M., & Rick, J. A. (2021). A beginner's guide to Conducting reproducible research. *Bulletin of the Ecological Society of America*, *102*(2), 1-14. https://doi.org/10.1002/bes2.1801

Chen, Y., Chee, X., & Feldman, G. (2023). Revisiting the Differential Centrality of Experiential and Material Purchases to the Self: Replication and extension of Carter and Gilovich (2012). *Collabra:Psychology*. Retrieved December 2022, from https://osf.io/v2w5h/.

Chioveanu, I., & Zhou, J. (2013). Price competition with consumer confusion. *Management Science*, *59*(11), 2450-2469. https://doi.org/10.1287/mnsc.2013.1716

Gourville, J. T. (1998). Pennies-a-Day: The Effect of Temporal Reframing on Transaction Evaluation. *The Journal of Consumer Research*, *24*(4), 395–403. https://doi.org/10.1086/209517

Heath, C., & Soll, J. (1996). Mental Budgeting and Consumer Decisions. *The Journal of Consumer Research, 23*(1), 40–52. https://doi.org/10.1086/209465

Henderson, P. W. & Peterson, R. A. (1992). Mental accounting and categorization. *Organizational Behavior and Human Decision Processes*, *51*(1), 92–117. https://doi.org/10.1016/0749-5978(92)90006-S

Imas, A. (2016). The realization effect: Risk-taking after realized versus paper losses. *American Economic Review*, *106*(8), 2086-2109. http://dx.doi.org/10.1257/aer.20140386

Jha-Dang, P. (2006). A Review of Psychological Research on Consumer Promotions and a New Perspective Based on Mental Accounting. *Vision (New Delhi, India)*, *10*(3), 35–43. https://doi.org/10.1177/097226290601000304

KNAW (2018). *Replication studies – Improving reproducibility in the empirical sciences*, Amsterdam, KNAW

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402. https://doi.org/10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3, 1-9. https://doi.org/10.15626/MP.2018.843

Leclerc, F., Schmitt, B. H., & Dube, L. (1995). Waiting Time and Decision Making: Is Time like Money? *The Journal of Consumer Research*, *22*(1), 110–119. https://doi.org/10.1086/209439

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *Revue Internationale de Psychologie Sociale, 32*(1). https://doi.org/10.5334/irsp.289

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, 49(2), 433-442. https://doi.org/10.3758/s13428-016-0727-z

Merkle, C., Müller-Dethard, J., & Weber, M. (2021). Closing a mental account: The realization effect for gains and losses. *Experimental Economics*, *24*(1), 303-329. https://doi.org/10.1007/s10683-020-09663-x

Open Science Framework, O., & Lakens, D. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science, 7*(6), 657-660. https://doi.org/10.1177/1745691612462588

Peels, R. (2019). Replicability and replication in the humanities. *Research Integrity and Peer*

Review, *4*(1), 1–12. https://doi.org/10.1186/s41073-018-0060-4

Samuelson, P. (1963). Risk and uncertainty: A fallacy of large numbers. *Scientia, 57*(98), 49-56.

Seymour, B., & McClure, S. M. (2008). Anchors, scales and the relative coding of value in the brain. *Current opinion in neurobiology*, *18*(2), 173-178. https://doi.org/10.1016/j.conb.2008.07.010

Shafir, E., & Thaler, R. H. (1998). 'Invest now, drink later, spend never: the mental accounting of advanced purchases'. *Unpublished working paper: University of Chicago*.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123-1128. https://doi.org/10.1177/1745691617708630

Soman, D., & Ahn, H.-K. (2011). Mental accounting and individual welfare. In G. Keren (Ed.), *Perspectives on framing* (pp. 65–92). Psychology Press. https://doi-org.ezproxy.library.sydney.edu.au/10.4324/9780203854167

Thaler, R. H.(1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing science*, *4*(3), 199-214.

Thaler, R. H., & Johnson, E. 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

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*(3), 183–206. https://doi.org/10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F

Trepel, C., Fox, C. R., & Poldrack, R. A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Brain Research. Cognitive Brain Research*, *23*(1), 34–50. https://doi.org/10.1016/j.cogbrainres.2005.01.016

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science (American Association for the Advancement of Science)*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business, 59*(4), 251-278.

Wan, K., & Feldman, G. (2021). Kahneman & Tversky (1972): Replication and Extension. Unpublished undergraduate thesis. DOI 10.17605/OSF.IO/QSWE3. Retrieved March 2022 from https://osf.io/r4h6s/

Yeung, S. K., & Feldman, G. (2022). Revisiting the Temporal Pattern of Regret in Action Versus Inaction: Replication of Gilovich and Medvec (1994) With Extensions Examining Responsibility. *Collabra: Psychology*, 8(1). doi: 10.1525/collabra.37122

Yue, P., Gizem Korkmaz, A., & Zhou, H. (2020). Household financial decision making amidst the COVID-19 pandemic. *Emerging Markets Finance and Trade*, *56*(10), 2363-2377. https://doi.org/10.1080/1540496X.2020.1784717

Zhang, C. Y., & Sussman, A. B. (2018). Perspectives on mental accounting: An exploration of budgeting and investing. *Financial Planning Review (Hoboken, N.J.)*, *1*(1-2), e1011–n/a. https://doi.org/10.1002/cfp2.1011

Revisiting mental accounting classic paradigms: Replication of the problems reviewed in Thaler (1999)

Supplementary

Contents

Open Science disclosures **3**

Data collection 3

Conditions reporting 3

Data exclusions 3

Variables reporting 3

Analysis of the original article **4**

Effect size and 95% confidence interval calculations of the original effects 4

Power analysis of original study effects to assess required sample for replication 4

Rationale for problem selection 4

Materials and scales used in the replication + extension experiment **5**

Instructions and experimental material 5

Problem 1 (Tversky & Kahneman, 1986, p.258) 5

Problem 2 (Tversky & Kahneman, 1981, p.457) 5

Problem 3 (Tversky & Kahneman, 1981, p.457) 5

Problem 4 (Thaler, 1985, p.202-204) 6

Problem 5 (Thaler & Johnson, 1990, p.649) 7

Problem 6 (Thaler & Johnson, 1990, p.651) 8

Problem 7 (Thaler, 1985, p.206) 9

Problem 8 (Thaler, 1985, p.205-206) 10

Problem 9 (Shafir & Thaler, 1998, p.697) 11

Problem 10 (Shafir & Thaler, 1998, p.699-700) 12

Problem 11 (Heath & Soll, 1996, p.48) 13

Problem 12 (Leclerc et al., 1995, p.113) 14

Problem 13 (Thaler, 1999, p.198) 14

Problem 14 (Thaler, 1999, p.198) 14

Problem 15 (Thaler, 1999, p.198) 14

Problem 16 (Samuelson, 1963, p.50-51) 15

Problem 17 (Thaler, 1999, p.201) 15
Extensions 16
Problem 18 (Thaler, 1980, p. 47) 16
Problem 19 (Thaler, 1980, p. 47) 16
Problem 20 (Thaler, 1999, p. 191) 16
Problem 21 (Thaler, 1999, p. 194) 17
Exclusion criteria 18
Generalised exclusion criteria 18
Comparisons and deviations **19**
Pre-exclusions versus post-exclusions 19
Pre-registration plan versus final report 32
Replication evaluation **34**
Replication closeness 34
Replication versus the original 35
Additional information about the study **36**
Data collection details 36
Detailed changes for the JAMOVI data analysis 36
Effect size and 95% confidence interval calculations of the current replication 37
References **37**

Open Science disclosures

Data collection

Data collection was completed before analysing the data.

Conditions reporting

All collected conditions are reported.

Data exclusions

Details are reported in the materials section of this document

Variables reporting

All variables collected for this study are reported and included in the provided data.

Analysis of the **original article**

Effect size and 95% confidence interval calculations of the original effects

See files Original studies_Thaler1999.Rmd/html/docx in the OSF folder.

Power analysis of original study effects to assess required sample for replication

Based on the effect sizes we calculated for the problems, power analysis was conducted to examine the required sample size for the current replication. We aimed for a 0.95 power and a 0.05 alpha error probability. The largest required sample size among all problems is 321.

See files Original studies_Thaler1999.Rmd/html/docx in the OSF folder.

Rationale for problem selection

In the targeted article, Thaler (1999) covered a wide array of mental accounting studies. In the current project we focused on problems that were simplified in design and were suitable for administration online with our target sample of the labour market.

An example for excluded studies is the study by Simonson (1990). In this study, Simonson assigned students to either 1) select among six snacks at each of the three class meetings held a week apart, or 2) select three snacks at the first class meeting to be consumed later every week. Such a study design cannot be adapted to online questions.

Materials and scales used in the **replication + extension experiment**

Instructions and experimental material

Replications

Problem 1 (Tversky & Kahneman, 1986, p.258)
Experimental condition 1.
Assume yourself richer by $300 than you are today. You are offered a choice between:

- A sure gain of $100
- A 50 % chance to gain $200 and a 50% chance to gain $0

Experimental condition 2.
Assume yourself richer by $500 than you are today. You are offered a choice between:

- A sure loss of $100
- A 50% chance to lose $200 and a 50% chance to lose $0

Problem 2 (Tversky & Kahneman, 1981, p.457)
Experimental condition 1.
Imagine that you are about to purchase a jacket for $125, and a calculator for $15. The calculator salesman informs you that the calculator you wish to buy is on sale for $10 at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?

- Yes
- No

Experimental condition 2.
Imagine that you are about to purchase a jacket for $15, and a calculator for $125. The calculator salesman informs you that the calculator you wish to buy is on sale for $120 at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?

- Yes
- No

Problem 3 (Tversky & Kahneman, 1981, p.457)
Experimental condition 1.
Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theater you discover that you have lost a $10 bill.
Would you still pay $10 for a ticket for the play?

- Yes
- No

Experimental condition 2
Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theater you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered.
Would you pay $10 for another ticket?

- Yes
- No

Problem 4 (Thaler, 1985, p.202-204)
Below you will find four pairs of scenarios, describing events in Mr. A's life and Mr. B's life.
You are asked to compare Mr. A to Mr. B.
If you think the two scenarios are emotionally equivalent, check "no difference."
In all cases the events are intended to be financially equivalent.

	Mr. A	Mr. B	No difference
Mr. A was given tickets to lotteries involving the World Series. He won $50 in one lottery and $25 in the other. Mr. B was given a ticket to a single, larger World Series lottery. He won $75. Who was **happier**?	O	O	O
Mr. A received a letter from the IRS saying that he made a minor arithmetical mistake on his tax return and owed $100. He received a similar letter the same day from his state income tax authority saying he owed $50. There were no other repercussions from either mistake. Mr. B received a letter from the IRS saying that he made a minor arithmetical mistake on his tax return and owed $150. There were no other repercussions from his mistake. Who was more **upset**?	O	O	O
Mr. A bought his first New York State lottery ticket and won $100. Also, in a freak accident, he damaged the rug in his apartment and had to pay the landlord $80. Mr. B bought his first New York State lottery ticket and won $20. Who was **happier**?	O	O	O
Mr. A's car was damaged in a parking lot. He had to spend $200 to repair the damage. The same day the car was damaged he won $25 in the office football pool. Mr. B's car was damaged in a parking lot. He had to spend $175 to repair the damage. Who was more **upset**?	O	O	O

Problem 5 (Thaler & Johnson, 1990, p.649)
Below you will find three pairs of events. In each case, the same events occur either on the same day (for Mr. A) or two weeks apart (for Mr. B).
You are asked to judge whether Mr. A or Mr. B is happier, or in the case of two negative events, who is more unhappy. If you think the alternatives are emotionally equivalent, select "no difference."
(Note: You are only asked to judge whether it is better to have the events separately or together).

	Mr. A (two events occur **on the same day**)	Mr. B (two events are **two weeks apart**)	No difference
The events are: (1) win $25 in an office lottery; (2) win $50 in an office lottery. Who is **happier**, Mr. A or Mr B.?	O	O	O
The events are: (1) receive a letter from the federal income tax authority saying that due to an arithmetical mistake $100 must be paid; (2) receive a letter from the state income tax authority saying that due to an arithmetical mistake $50 must be paid. Who is **more unhappy**, Mr. A or Mr B.?	O	O	O
The events are: (1) receive a $20 parking ticket; (2) receive a bill for $25 from the registrar because a form was filled in improperly. Who is **more unhappy**, Mr. A or Mr B.?	O	O	O

Problem 6 (Thaler & Johnson, 1990, p.651)
Experimental condition 1.
Consider the following two events: (A) you lose $x. (B) you lose $x after gaining/losing $y. We are interested in the emotional impact of the loss of $x in both cases.
Are you more upset about the loss of money when it occurs alone (A), or when it occurs directly after a prior gain/loss (B)?
Below are several questions of this type.In each case please compare the incremental effect of the event described. If you feel there is no difference you may check that, but please express a preference if you have one.
For each of the following pairs of events, please indicate which of the two hurts more:

	A hurts more than B	B hurts more than A	No difference
(A) You lose $9. (B) You lose $9 after having gained $30.	O	O	O
(A) You lose $9. (B) You lose $9 after having lost $30.	O	O	O
(A) You lose $9. (B) You lose $9 after having lost $250.	O	O	O
(A) You lose $9. (B) You lose $9 after suffering a loss of $1000.	O	O	O
(A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000.	O	O	O

Experimental condition 2.
Consider the following two events: (A) you lose $x. (B) you lose $x after gaining/losing $y. We are interested in the emotional impact of the loss of $x in both cases.
Are you more upset about the loss of money when it occurs alone (A), or when it occurs directly after a prior gain/loss (B)?
Below are several questions of this type.In each case please compare the incremental effect of the event described. If you feel there is no difference you may check that, but please express a preference if you have one.
For each of the following pairs of events, please indicate which of the two hurts more:

	A hurts more than B	B hurts more than A	No difference

(A) You lose $9. (B) You lose $9 after suffering a loss of $9.	O	O	O
(A) You lose $9. (B) You lose $9 after suffering a loss of $18.	O	O	O
(A) You lose $9. (B) You lose $9 after suffering a loss of $36.	O	O	O
(A) You lose $9. (B) You lose $9 after suffering a loss of $45.	O	O	O
(A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36.	O	O	O

Problem 7 (Thaler, 1985, p.206)
Experimental condition 1.
You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how much you would enjoy a nice cold bottle of your favorite brand of soda. A companion gets up to go make a phone call and offers to bring back a soda from the only nearby place where soda is sold at **a fancy resort hotel**. He says that the soda might be expensive and so asks how much you are willing to pay for the soda. He says that he will buy the soda if it costs as much or less than the price you state. But if it costs more than the price you state he will not buy it. You trust your friend, and there is no possibility of bargaining with the **bartender**.
What price do you tell him?
_____

Experimental condition 2.
You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how much you would enjoy a nice cold bottle of your favorite brand of soda. A companion gets up to go make a phone call and offers to bring back a soda from the only nearby place where soda is sold at a **small, run-down grocery store**. He says that the soda might be expensive and so asks how much you are willing to pay for the soda. He says that he will buy the soda if it costs as much or less than the price you state. But if it costs more than the price you state he will not buy it. You trust your friend, and there is no possibility of bargaining with the **store owner**.
What price do you tell him?
_____

Problem 8 (Thaler, 1985, p.205-206)
Experimental condition 1.
Imagine that you are going to a sold-out Cornell hockey playoff game, and you have an extra ticket to sell or give away. The price marked on the ticket is $5, but **you were given your tickets for free by a friend**. You get to the game early to make sure you get rid of the ticket. An informal survey of people selling tickets indicates that the **going price is $5**. You find someone who wants the ticket and takes out his wallet to pay you. He asks how much you want for the ticket.
Assume that there is no law against charging a price higher than that marked on the ticket. What price do you ask for if...
he is **a friend**

_____

he is **a stranger**

_____

What would you have said if instead you found the **going market price was $10**?

he is **a friend**

_____

he is **a stranger**

_____

Experimental condition 2.
Imagine that you are going to a sold-out Cornell hockey playoff game, and you have an extra ticket to sell or give away. The price marked on the ticket is **$5**, **which is what you paid for each ticket**. You get to the game early to make sure you get rid of the ticket. An informal survey of people selling tickets indicates that **the going price is $5**. You find someone who wants the ticket and takes out his wallet to pay you. He asks how much you want for the ticket.
Assume that there is no law against charging a price higher than that marked on the ticket. What price do you ask for if
he is **a friend**

_____

he is **a stranger**

_____

What would you have said if instead you found the **going market price was $10**?

he is **a friend**

_____

he is **a stranger**

_____

Experimental condition 3.
Imagine that you are going to a sold-out Cornell hockey playoff game, and you have an extra ticket to sell or give away. The price marked on the ticket is **$5, but you paid $10 each for your tickets when you bought them from another student**. You get to the game early to make sure you get rid of the ticket. An informal survey of people selling tickets indicates that the **going price is $5**. You find

someone who wants the ticket and takes out his wallet to pay you. He asks how much you want for the ticket.
Assume that there is no law against charging a price higher than that marked on the ticket. What price do you ask for if
he is **a friend**

_____

he is **a stranger**

_____

What would you have said if instead you found the **going market price was $10**?

he is **a friend**

_____

he is **a stranger**

_____

Problem 9 (Shafir & Thaler, 1998, p.697)
Experimental condition 1.
Imagine that you enjoy drinking wine. Suppose you bought a case of a good 1982 Bordeaux in the futures market for $20 a bottle. The wine now sells at auction for about $75 a bottle. You have decided to **drink a bottle of this wine with dinner**.
Which of the following best captures your feeling of the cost to you of drinking this bottle?

O Drinking the bottle does not feel like it costs me anything, I paid for the bottle already, many years ago, and probably do not remember exactly what I paid for it anyway.

O Drinking the bottle feels like it costs me $20, the amount I roughly remember paying for it.

O Drinking the bottle feels like it costs me $20, the amount I originally paid for it, plus whatever the interest would have been on the money I paid.

O Drinking the bottle feels like it costs me $75, the amount it would take to replace it.

O Drinking the bottle feels like I am saving $55, because I am able to drink a $75 bottle for which I only paid $20.

Experimental condition 2.
Imagine that you enjoy drinking wine. Suppose you bought a case of a good 1982 Bordeaux in the futures market for $20 a bottle. The wine now sells at auction for about $75 a bottle. You have decided to **give one bottle of this wine to a friend as a gift**.
Which of the following best captures your feeling of the cost to you of giving away this bottle?

O Giving away the bottle does not feel like it costs me anything, I paid for the bottle already, many years ago, and probably do not remember exactly what I paid for it anyway.

O Giving away the bottle feels like it costs me $20, the amount I roughly remember paying for it.

O Giving away the bottle feels like it costs me $20, the amount I originally paid for it, plus whatever the interest would have been on the money I paid.

O Giving away the bottle feels like it costs me $75, the amount it would take to replace it.

O Giving away the bottle feels like I am saving $55, because I am able to give a $75 gift for which I only paid $20.

Problem 10 (Shafir & Thaler, 1998, p.699-700)
Imagine that you enjoy drinking wine. Suppose you buy a case of Bordeaux futures at $400 a case. The wine will retail at about $500 a case when it is shipped. You do not intend to start drinking this wine for a decade.
At the time that you acquire this wine, which statement more accurately captures your feelings?

	1 Not accurate at all	2	3	4	5 Very accurate
I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway.	O	O	O	O	O
I feel like I made a $400 investment which I will gradually consume after a period of years.	O	O	O	O	O
I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered.	O	O	O	O	O
I cannot understand this question.	O	O	O	O	O

Problem 11 (Heath & Soll, 1996, p.48)
Experimental condition 1.
Each of the questions below asks you to imagine that a specific event took place at the beginning of the week and whether based on that you would make a purchase later in the week.

	Yes	No
You **go out to dinner with friends**. You **spend $50**. Would you buy a $25 theater ticket later in the week?	O	O
You are **given the dinner** above. Would you buy a $25 theater ticket-later in the week?	O	O
You **spend $50 on a sports ticket**. Would you purchase a $25 theater ticket later in the week?	O	O
You are **given the sports ticket** above. Would you purchase a $25 theater ticket later in the week?	O	O
You hear of a flu epidemic on the news. You **spend $50 for an inoculation**. Would you purchase a $25 theater ticket later in the week?	O	O

Experimental condition 2.
Each of the questions below asks you to imagine that a specific event took place at the beginning of the week and whether based on that you would make a purchase later in the week.

	Yes	No
You **go out to dinner with friends**. You **spend $20**. Would you buy a $25 theater ticket later in the week?	O	O
You are **given the dinner** above. Would you buy a $25 theater ticket-later in the week?	O	O
You **spend $20 on a sports ticket.** Would you purchase a $25 theater ticket later in the week?	O	O

You are **given the sports ticket** above. Would you purchase a $25 theater ticket later in the week?	O	O
You hear of a flu epidemic on the news. You **spend $20 for an inoculation**. Would you purchase a $25 theater ticket later in the week?	O	O

Problem 12 (Leclerc et al., 1995, p.113)
Experimental condition 1.
On the day of the performance, you decide to purchase a pair of tickets for a show at the performing arts center. You are sure that the show will not be sold out. However, because you know that you and your friend will arrive right before the performance begins, you have chosen to purchase the tickets earlier in the day. **Tickets will cost $15 each** if you buy them at the discounted window of the box office.
You have just arrived at the ticket counter when you realize that the discounted window will not open for 45 minutes. The window for regular tickets is already open, and you could purchase the same tickets there but at a higher price. In other words, you have to decide whether you should buy the tickets at the regular price now or whether you should wait for 45 minutes to buy the discounted tickets.
How much would you be willing to pay for **the regular ticket** to avoid waiting for 45 minutes?
_____

Experimental condition 2.
On the day of the performance, you decide to purchase a pair of tickets for a show at the performing arts center. You are sure that the show will not be sold out. However, because you know that you and your friend will arrive right before the performance begins, you have chosen to purchase the tickets earlier in the day. **Tickets will cost $40 each** if you buy them at the discounted window of the box office.
You have just arrived at the ticket counter when you realize that the discounted window will not open for 45 minutes. The window for regular tickets is already open, and you could purchase the same tickets there but at a higher price. In other words, you have to decide whether you should buy the tickets at the regular price now or whether you should wait for 45 minutes to buy the discounted tickets.
How much would you be willing to pay for **the regular ticket** to avoid waiting for 45 minutes?
_____

Problem 13 (Thaler, 1999, p.198)
Imagine that you have just **won $30**. Now choose between:
- o A 50% chance to gain $9 and a 50% chance to lose $9.
- o No further gain or loss.

Problem 14 (Thaler, 1999, p.198)
Imagine that you have just **lost $30**. Now choose between:
- o A 50% chance to gain $9 and a 50% chance to lose $9.
- o No further gain or loss.

Problem 15 (Thaler, 1999, p.198)
Imagine that you have just **lost $30**. Now choose between:

- A 33% chance to gain $30 and a 67% chance to gain nothing.
- A sure $10.

Problem 16 (Samuelson, 1963, p.50-51)
Please read the following scenario carefully and answer the questions that follow:

	Yes	No
A single coin flip, heads you win $200, tail you lose $100. Would you take the bet?	O	O
A package bet of 100 coin flips, each coin flip you either win $200 or lose $100. Would you take the bet?	O	O

Problem 17 (Thaler, 1999, p.201)
Imagine that you are managing a division

	Yes	No
Would you be willing to undertake a project for your division if the payoffs are as follows: 50% chance to gain $2 million, 50% chance to lose $1 million.	O	O
Would you be willing to undertake a portfolio of 25 of investments that **each** have a 50% chance of gaining $2 million and 50% chance of losing $1 million.	O	O

Extensions

Problem 18 (Thaler, 1980, p. 47)
Imagine that there is a basketball game to be played 60 miles from your home. Your family gave you a ticket. On the day of the game there is a snowstorm.

	Go to the game anyway	Stay home
Given the snowstorm, what would you choose to do if your family **paid $40 for your ticket**?	O	O
Given the snowstorm, what would you choose to do if the ticket was **given to your family for free**?	O	O

Problem 19 (Thaler, 1980, p. 47)
Imagine that you joined a tennis club and paid a $300 yearly membership fee. After two weeks of playing you develop a tennis elbow (a condition that causes pain around the outside of the elbow). Which of the following is more likely for you?

O Stop playing

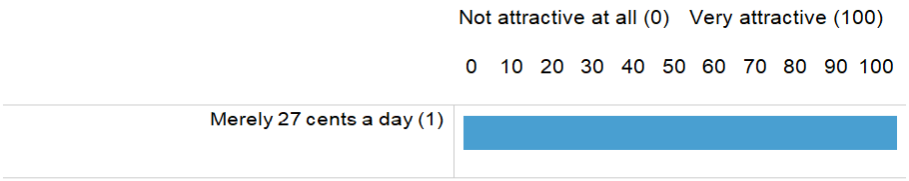
O Continue to play (in pain) saying 'I don't want to waste the $300!'

Problem 20 (Thaler, 1999, p. 191)
Suppose you buy a pair of shoes. They feel perfectly comfortable in the store, but the first day you wear them they hurt. A few days later you try them again, but they hurt even more than the first time. What happens now?

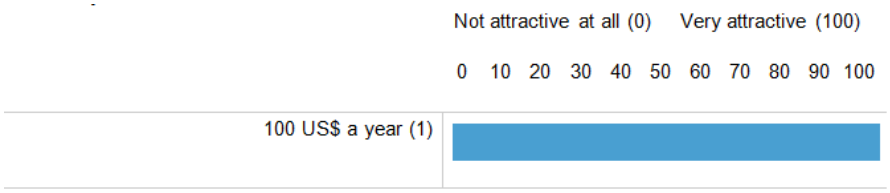
How accurately do the statements express your feelings?

	1 Not accurate at all	2	3	4	5 Very accurate
The more you paid for the shoes, the more times you will try to wear them.	O	O	O	O	O
Eventually you stop wearing the shoes, **but you do not throw them away**. The more you paid for the shoes, the longer they sit in the back of your closet before you throw them away.	O	O	O	O	O

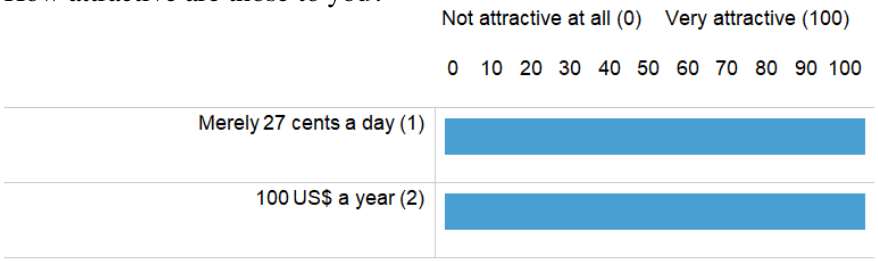
Problem 21 (Thaler, 1999, p. 194)
Experimental condition 1.
Below is an annual membership plan for an online music streaming subscription.
How attractive is this to you?



Experimental condition 2.
Below is an annual membership plan for an online music streaming subscription.
How attractive is this to you?



Experimental condition 3.
Below are two annual membership plans for an online music streaming subscription.
How attractive are those to you?



Following that, participants are to complete the funnelling section.
Five funnelling questions:

- *How serious were you in filling out this questionnaire?*
- *Have you ever seen the materials used in this study or similar before?If yes - please indicate where.*
- *What do you think the purpose of the study was? (one sentence)*
- *Have you ever seen the materials used in this study or similar before? If yes - please indicate where*
- *Help us improve for the next studies. Did you spot any errors? Anything missing or wrong? Something we should pay attention to in next runs? (briefly)*

Finally, participants were asked to fill in demographics and were debriefed. No filler items were included.

- *How old are you?*
- *Please indicate your gender*
- *Which country are you originally from? (country of birth)*
- *In which country are you currently residing?*
- *Please estimate your family's social class*
- *How would you generally rate your understanding of the English used in this study?*

- *How often do you drink alcohol?*
- *Do you like drinking wine?*

Exclusion criteria

Generalised exclusion criteria

In the actual data collection, we focused on our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we also determined further findings reports with exclusions. Please see the JAMOVI file *Exclusion_Data analysis_Thaler1999* for the results following exclusions.

General criteria:

1. Participants indicating a low proficiency of English (self-report < 5, on a 1-7 scale)
2. Participants who self-report not being serious about filling in the survey (self-report < 4, on a 1-5 scale).
3. Participants who have already seen or done the survey before.
4. Participants aged below 18. [there is a type-in answer box about participants' age in Qualtrics template with default age range 1-100]

Comparisons and deviations

Pre-exclusions versus post-exclusions

There were a total of 1007 participants for the current replication. Based on the general exclusion criteria, 21 participants were excluded. Table 1 summarised the sample characteristics before and after exclusion. Please see the below tables for the results of the post-exclusion sample.

Table 1

Comparison of sample characteristics before and after exclusion

	Full sample	**After Exclusion**
Sample size	1007	986
Gender	471 females, 526 males, 3 others, and 7 rather not disclose	458 females, 518 males, 3 others, and 7 rather not disclose
Median age (years)	40.00	41.00
Average age (years)	43.28	43.35
Age range (years)	20-80	20-80

Table 2

Descriptive statistics of all Problems with options

Problem	**Conditions and Options (if applicable)**	**Replication** N	Count	Percentage
1 (between)	**Gain:**	243		
	A sure gain of $100		203	84%
	A 50 % chance to gain $200 and a 50% chance to gain $0		40	16%
	Loss:	248		
	A sure loss of $100		107	43%
	A 50% chance to lose $200 and a 50% chance to lose $0		141	57%
2 (between)	**$15 Calculator:**	251		
	Make the trip		29	12%
	Not making the trip		222	88%
	$125 Calculator:	250		
	Make the trip		17	7%
	Not making the trip		233	93%
3 (between)	**Lost a $10 bill:**	249		
	Buy another ticket		225	90%
	Not buying another ticket		24	10%
	Lost the ticket:	248		
	Buy another ticket		193	78%
	Not buying another ticket		55	22%
4 (within)	**1. Two wins: $50, and $25 versus One win: $75. Who was happier?**	498		
	Two wins is happier		175	35%

Problem	Conditions and Options (if applicable)	Replication N	Count	Percentage
	One win is happier		60	12%
	No difference		263	53%
	2. Two mistakes: $100, and $50 versus One mistake: $150. Who was more upset?			
	Two mistakes is more upset	498	211	42%
	One mistake is more upset		66	13%
	No difference		221	44%
	3. Two events: Win $100, and loss $80 versus One event: Win $20. Who was happier?			
	Two events is happier	498	59	12%
	One event is happier		382	77%
	No difference		57	11%
	4. Two events: Loss $200, and win $25 versus One event: Loss $175. Who was more upset?			
	Two events is more upset	498	62	12%
	One event is more upset		311	62%
	No difference		125	25%
5 (within)	**A. Two events: (1) win $25 (2) win $50. Who is happier?**			
	Happier on the same day	481	194	40%
	Happier two weeks apart		167	35%
	No difference		120	25%
	B. Two events: (1) $100 must be paid (2) $50 must be paid. Who is more unhappy?			
	More unhappy on the same day	481	294	61%
	More unhappy two weeks apart		92	19%
	No difference		95	20%
	C. Two events: (1) a $20 parking ticket (2) a $25 bill. Who is more unhappy?			
	More unhappy on the same day	481	271	56%
	More unhappy two weeks apart		111	23%
	No difference		99	21%
6 (between)	**First group of questions:** **1. (A) You lose $9. (B) You lose $9 after having gained $30.**			
	A hurts more	250	207	83%
	B hurts more		31	12%
	No difference		12	5%
	2. (A) You lose $9. (B) You lose $9 after having lost $30.			
	A hurts more	250	19	8%
	B hurts more		216	86%
	No difference		15	6%

Problem	Conditions and Options (if applicable)	Replication N	Count	Percentage
	3. (A) You lose $9. (B) You lose $9 after having lost $250.			
	A hurts more	250	32	13%
	B hurts more		206	82%
	No difference		12	5%
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $1000.			
	A hurts more	250	36	14%
	B hurts more		199	80%
	No difference		15	6%
	5. (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000.			
	A hurts more	250	50	20%
	B hurts more		183	73%
	No difference		17	7%
	Second group of questions: **1. (A) You lose $9. (B) You lose $9 after suffering a loss of $9.**			
	A hurts more	249	9	4%
	B hurts more		207	83%
	No difference		33	13%
	2. (A) You lose $9. (B) You lose $9 after suffering a loss of $18.			
	A hurts more	249	9	4%
	B hurts more		228	92%
	No difference		12	5%
	3. (A) You lose $9. (B) You lose $9 after suffering a loss of $36.			
	A hurts more	249	11	4%
	B hurts more		225	90%
	No difference		13	5%
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $45.			
	A hurts more	249	10	4%
	B hurts more		227	91%
	No difference		12	5%
	5. (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36.			
	A hurts more	249	16	6%
	B hurts more		221	89%
	No difference		12	5%
9 (between)	**Giving away:**			
	$0	247	74	30%
	$20		50	20%

Problem	Conditions and Options (if applicable)	Replication N	Count	Percentage
	$20 plus interest		12	5%
	$75		52	21%
	A $55 saving		59	24%
	Drinking:			
	$0		56	23%
	$20		52	21%
	$20 plus interest	247	16	6%
	$75		64	26%
	A $55 saving		59	24%
11 (between)	**$50 high cost condition: Spend $50 on dinner. Would you buy a $25 theater ticket later in the week?**	250		
	Buy the ticket		99	40%
	Not buying the ticket		151	60%
	Given a $50 dinner. Would you buy a $25 theater ticket-later in the week?	250		
	Buy the ticket		209	84%
	Not buying the ticket		41	16%
	Spend $50 on a sports ticket. Would you purchase a $25 theater ticket later in the week?	250		
	Buy the ticket		68	27%
	Not buying the ticket		182	73%
	Given a $50 sports ticket. Would you purchase a $25 theater ticket later in the week?	250		
	Buy the ticket		197	79%
	Not buying the ticket		53	21%
	Spend $50 for an inoculation. Would you purchase a $25 theater ticket later in the week?	250		
	Buy the ticket		74	30%
	Not buying the ticket		176	70%
	$20 low cost condition: Spend $20 on dinner. Would you buy a $25 theater ticket later in the week?	246		
	Buy the ticket		134	54%
	Not buying the ticket		112	46%
	Given a $20 dinner. Would you buy a $25 theater ticket-later in the week?	246		
	Buy the ticket		209	85%
	Not buying the ticket		37	15%
	Spend $20 on a sports ticket. Would you purchase a $25 theater ticket later in the week?	246		
	Buy the ticket		100	41%
	Not buying the ticket		146	59%

Problem	Conditions and Options (if applicable)	Replication N	Count	Percentage
	Given a $20 sports ticket. Would you purchase a $25 theater ticket later in the week?	246		
	Buy the ticket		198	80%
	Not buying the ticket		48	20%
	Spend $20 for an inoculation. Would you purchase a $25 theater ticket later in the week?	246		
	Buy the ticket		105	43%
	Not buying the ticket		141	57%
13	**Imagine that you have just won $30.**			
	A 50% chance to gain $9 and a 50% chance to lose $9.	495	138	28%
	No further gain or loss.		357	72%
14	**Imagine that you have just lost $30.**			
	A 50% chance to gain $9 and a 50% chance to lose $9.	495	83	17%
	No further gain or loss.		412	83%
15	**Imagine that you have just lost $30.**			
	A 33% chance to gain $30 and a 67% chance to gain nothing.	495	116	23%
	A sure $10.		379	77%
16 (within)	**A single coin flip, heads you win $200, tail you lose $100. Would you take the bet?**	493		
	Take the bet.		124	25%
	Not taking the bet		369	75%
	A package bet of 100 coin flips, each coin flip you either win $200 or lose $100. Would you take the bet?	493		
	Take the bet.		239	48%
	Not taking the bet		254	52%
17 (within)	**A project:**			
	50% chance to gain $2 million, 50% chance to lose $1 million.	494		
	Undertake the project		147	30%
	Not undertaking the project		347	70%
	A portfolio of 25 of investments:			
	Each has a 50% chance of gaining $2 million and 50% chance of losing $1 million.	494		
	Undertake the investments		229	46%
	Not undertaking the investments		265	54%
18 (within)	**Paid $40 for tickets:**			
	Go to the game	488	156	32%
	Stay home		332	68%
	Tickets given by friends:			
	Go to the game	488	64	13%
	Stay home		424	87%

Problem	Conditions and Options (if applicable)	Replication		
		N	Count	Percentage
19	**Imagine that you joined a tennis club and paid a $300 yearly membership fee.**	488		
	Stop playing		373	76%
	Continue to play		115	24%

Note. N represents sample size.

Table 3

Descriptive statistics for Problems with scale and text entry

Problem	Condition (if applicable)	Replication		
		N	Mean	Standard deviation
7 (between)	Hotel condition	247	6.79	19.23
	Grocery condition	246	4.13	3.91
8 (between)	Free condition-Market value $5-Friend	161	2.26	3.42
	Free condition-Market value $5-Stranger	161	5.95	5.00
	Free condition-Market value $10-Friend	161	4.14	5.40
	Free condition-Market value $10-Stranger	161	10.17	6.76
	Paid 5 condition-Market value $5-Friend	165	3.75	2.46
	Paid 5 condition-Market value $5-Stranger	165	6.71	3.48
	Paid 5 condition-Market value $10-Friend	165	6.18	4.18
	Paid 5 condition-Market value $10-Stranger	165	11.58	5.58
	Paid 10 condition-Market value $5-Friend	160	4.98	2.66
	Paid 10 condition-Market value $5-Stranger	160	8.20	3.59
	Paid 10 condition-Market value $10-Friend	160	7.52	3.83
	Paid 10 condition-Market value $10-Stranger	160	11.44	4.03
10 (within)	I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway.	491	2.98	1.46
	I feel like I made a $400 investment which I will gradually consume after a period of years.	491	3.56	1.31
	I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered.	491	3.09	1.35
	I cannot understand this question.	491	1.16	0.66
12 (between)	$15 condition	228	7.72	7.76
	$40 condition	217	10.37	7.63
20 (within)	The more you paid for the shoes, the more times you will try to wear them.	494	3.10	1.41
	Eventually you stop wearing the shoes, but you do not throw them away. The more you paid for the shoes, the longer they sit in the back of your closet before you throw them away.	494	3.46	1.33
21 (between)	Day expression	163	44.80	32.11
	Year expression	164	26.20	28.14
	Both expressions-Day	167	45.66	31.62
	Both expressions-Year	167	36.25	29.02

Note. N represents sample size.

Table 4

Summary of all chi square results

Problem	χ^2	χ^2 **continuity correction**	**Difference in 2 proportions** [95% CI]	**Odds ratio** [95% CI]	**Phi -coefficient**	**Cramer's V**
1	86.05 (df=1, p <.001)	84.32 (df=1, p <.001)	-0.40 [-0.48; -0.33]	0.15 [0.10; 0.23]	0.42	0.42
2	3.39 (df=1, p=0.065)	2.85 (df=1, p=0.091)	0.05 [-0.00; 0.10]	1.79 [0.96; 3.35]	0.08	0.08
3	14.61 (df=1, p<.001)	13.69 (df=1, p<.001)	0.13 [0.06; 0.19]	2.67 [1.59; 4.48]	0.17	0.17
9	4.34 (df=4, p= 0.361)	4.34 (df=4, p= 0.361)	/	/	/	0.09

Note. df indicates degree of freedom and CI indicates Confidence Interval.

Table 5

Summary of all χ^2 Goodness of Fit in proportion tests

Problem	Conditions and sub questions	χ^2	df	p
4	1. Two wins: $50, and $25 versus One win: $75. Who was happier?	124.86	2	< .001
	2. Two mistakes: $100, and $50 versus One mistake: $150 Who was more upset?	90.66	2	< .001
	3. Two events: Win $100, and loss $80 versus One event: Win $20 Who was happier?	421.60	2	< .001
	4. Two events: Loss $200, and win $25 versus One event: Loss $175 Who was more upset?	201.94	2	< .001
5	A. Two events: (1) win $25 (2) win $50 Who is happier?	17.49	2	< .001
	B. Two events: (1) $100 must be paid (2) $50 must be paid. Who is more unhappy?	167.18	2	< .001
	C. Two events: (1) a $20 parking ticket (2) a $25 bill Who is more unhappy?	115.03	2	< .001
6	First group of questions: 1. (A) You lose $9. (B) You lose $9 after having gained $30.	277.45	2	< .001
	2. (A) You lose $9. (B) You lose $9 after having lost $30.	316.90	2	< .001
	3. (A) You lose $9. (B) You lose $9 after having lost $250.	273.25	2	< .001
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $1000.	243.46	2	< .001
	5. (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000.	185.34	2	< .001
	Second group of questions: 1. (A) You lose $9. (B) You lose $9 after suffering a loss of $9.	281.35	2	< .001
	2. (A) You lose $9. (B) You lose $9 after suffering a loss of $18.	380.02	2	< .001
	3. (A) You lose $9. (B) You lose $9 after suffering a loss of $36.	364.43	2	< .001
	4. (A) You lose $9. (B) You lose $9 after suffering a loss of $45.	374.77	2	< .001
	5. (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36.	344.27	2	< .001
19	Imagine that you joined a tennis club and paid a	136.40	1	< .001

Problem	Conditions and sub questions	χ^2	df	p
	$300 yearly membership fee.			

Note. df indicates degree of freedom

Table 6

Summary of all McNemar paired-samples tests

Problem	Comparisons	χ^2	df	p	χ^2 continuity correction	Log odds ratio exact
5	Comparing A to B A-(1) win $25 (2) win $50 Who is happier? B-(1) $100 must be paid (2) $50 must be paid. Who is more unhappy?	45.05	3	<.001	45.05 (df=3, p<.001)	/
	Comparing A to C A- (1) win $25 (2) win $50 Who is happier? C-(1) a $20 parking ticket (2) a $25 bill Who is more unhappy?	38.24	3	<.001	38.24 (df=3, p<.001)	/
14	Problem 13 vs. 14	20.86	1	<.001	/	0.80 (p<.001)
15	Problem 13 vs. 15	2.81	1	0.093		0.26 (p=0.109)
16	A single bet vs. 100 bets	82.14	1	< .001	/	-1.79 (p <.001)
17	A project vs 25 investments	44.24	1	< .001	/	-1.21 (p <.001)
18	Paid $40 vs. Given	88.17	1	< .001		3.85 (p <.001)

Note. df indicates degree of freedom

Table 7

Summary of all t-tests results

Problem (test type)	**Statistic**	**df**	**p**	**Mean difference**	**SE difference**	**Effect size** (Cohen's d)	**95% CI**
7 Independent samples t-test	Student's t=2.12	491.00	0.034	2.66	1.25	0.19	[0.01, 0.37]
	Welch's t=2.13	266.39	0.034	2.66	1.25	0.19	/
12 Independent samples t-test	Student's t=3.64	443.00	<.001	2.66	0.73	0.35	[0.16, 0.53]
	Welch's t=3.64	442.52	<.001	2.66	0.73	0.35	/
20 One-sample t-test	Statement 1: Student's t=1.63	493.00	0.052	0.10	/	0.07	[-0.02, 0.16]
	Statement 2: Student's t=7.69	493.00	<.001	0.46	/	0.35	[0.25, 0.44]
21 Independent samples t-test	Student's t=5.57	325.00	< .001	18.60	3.34	0.62	[0.39, 0.84]
	Welch's t=5.57	319.01	< .001	18.60	3.34	0.62	/
21 Paired sample t-test	Student's t=3.75	166.00	< .001	9.41	2.51	0.29	[0.13, 0.44]

Note. df indicates degree of freedom, SE indicates standard error, and CI indicates confidence interval.

Table 8

Summary of all ANOVA results

Problem 8-Mixed ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2
Friend vs. Stranger	8586.38	1	8586.38	580.11	< .001	0.16
Friend vs. Stranger ✻ Q8-Condition	133.37	2	66.69	4.51	0.012	0.00
Residual	7149.00	483	14.80			
Market worth $5 vs. $10	4970.51	1	4970.51	886.10	< .001	0.09
Market worth $5 vs. $10 ✻ Q8-Condition	52.44	2	26.22	4.67	0.010	0.00
Residual	2709.35	483	5.61			
Friend vs. Stranger ✻ Market worth $5 vs. $10	405.07	1	405.07	198.94	< .001	0.01
Friend vs. Stranger ✻ Market worth $5 vs. $10 ✻ Q8-Condition	75.91	2	37.96	18.64	< .001	0.00
Residual	983.46	483	2.04			

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2
Q8-Condition	1876.67	2	938.34	17.13	< .001	0.04
Residual	26457.85	483	54.78			

Problem 10-Repeated measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2
Feeling	92.89	2	46.44	23.87	< .001	0.03
Residual	1906.44	980	1.95			

Problem 11-Mixed ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2
Dinner/Ticket	3.90	1	3.90	54.97	< .001	0.01
Dinner/Ticket ✻ Q11-Condition	0.00	1	0.00	0.05	0.820	0.00
Residual	35.09	494	0.07			
Given/Spent	85.34	1	85.34	392.31	< .001	0.18
Given/Spent ✻ Q11-Condition	1.98	1	1.98	9.10	0.003	0.00
Residual	107.46	494	0.22			
Dinner/Ticket ✻ Given/Spent	0.89	1	0.89	18.25	< .001	0.00
Dinner/Ticket ✻ Given/Spent ✻ Q11-Condition	0.01	1	0.01	0.19	0.659	0.00
Residual	24.10	494	0.05			

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2
Q11-Condition	3.05	1	3.05	7.21	0.008	0.01
Residual	209.11	494	0.42			

Note. Type 3 Sums of Squares. df represents degree of freedom.

Pre-registration plan versus final report

Components in pre-registration	**Location of 1) pre-registered decision/plan, and 2) rationale for decision/plan https://osf.io/brgc6**	**Were there deviations? What type? [no / minor / major]***	**If yes - describe details of deviation(s)**	**Rationale for deviation**	**How might the results be different if you had/had not deviated**	**Date/time of decision for deviation + stage**
Study design	p.21-25 , Method, "Design and procedure "	no	/	/	/	/
Measured variables	p.25-28 , Method, "Measures"	no	/	/	/	/
Exclusion criteria	p.57, Generalised exclusion criteria , "General criteria"	minor	The pre-registration stated that we would also exclude 1) Participants who failed to complete the survey (duration = 0, leave question blank), and 2) Participants not from the US in the post-exclusion analysis. We deleted the two criteria in the final report.	Given our data collection filters these subjects were not allowed to proceed to the survey and therefore did not count as participants in the first place.	The change would not influence the results.	After pre-registration
IV	p.23-25, Method, "Table 5"	no	/	/	/	/
DV	p.23-25, Method, "Table 5"	no	/	/	/	/

Components in pre-registration	Location of 1) pre-registered decision/plan, and 2) rationale for decision/plan https://osf.io/brgc6	Were there deviations? What type? [no / minor / major]*	If yes - describe details of deviation(s)	Rationale for deviation	How might the results be different if you had/had not deviated	Date/time of decision for deviation + stage
Data analysis	p.26-28, Method, "Table 6 and 8"	major	Problem 12: The pre-registration stated that we would subtract the price of the ticket and conduct independent samples t-test. In the final report, we also excluded data below 0 prior to the independent samples t-test. Problem 20: The pre-registration stated that we would conduct Paired samples t-test, while in the final report, we used One-sample t-test against the midpoint	Problem 12: The subtracted value represents willingness to pay. Therefore, data below 0 indicates that the participant did not understand the question correctly and needed to be excluded. Problem 20: One-sample t-test is more appropriate to evaluate whether participants agree with the statement. And we did not aim to compare the results between the two statement	The results would be incorrect.	After pre-registration

Notes. *Categories for deviations: Minor - Change probably did not affect results or interpretations; Major - Change likely affected results or interpretat

Replication evaluation

Replication closeness

Lebel et al. (2018):

Target similarity	Highly similar				Highly dissimilar
Category	Direct replication			Conceptual replication	
Design facet	**Exact replication**	**Very close replication**	**Close replication**	**Far replication**	**Very far replication**
Effect/hypothesis	Same/similar	Same/similar	Same/similar	Same/similar	Same/similar
IV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different
DV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different
IV operationalization	Same/similar	Same/similar	Same/similar	Different	
DV operationalization	Same/similar	Same/similar	Same/similar	Different	
Population (e.g. age)	Same/similar	Same/similar	Same/similar	Different	
IV stimuli	Same/similar	Same/similar	Different		
DV stimuli	Same/similar	Same/similar	Different		
Procedural details	Same/similar	Different			
Physical setting	Same/similar	Different			
Contextual variables	Different				

Figure 1. Criteria for evaluation of replications by LeBel et al. (2018).
A classification of relative methodological similarity of a replication study to an original study. "Same" ("different") indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable. "Everything controllable" indicates design facets over which a researcher has control. Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.).
"Similar" category was added to the Lebel et al. (2018) typology to refer to minor deviations or extensions aimed to adjust the study to the target sample that are not expected to have major implications on replication success. See Olsson-Collentine, van Assen, and Wicherts (2020) on meta analysis showing minor to no expected impact due to variations in sample population or setting.

Replication versus the original

LeBel et al. (2019) criteria:

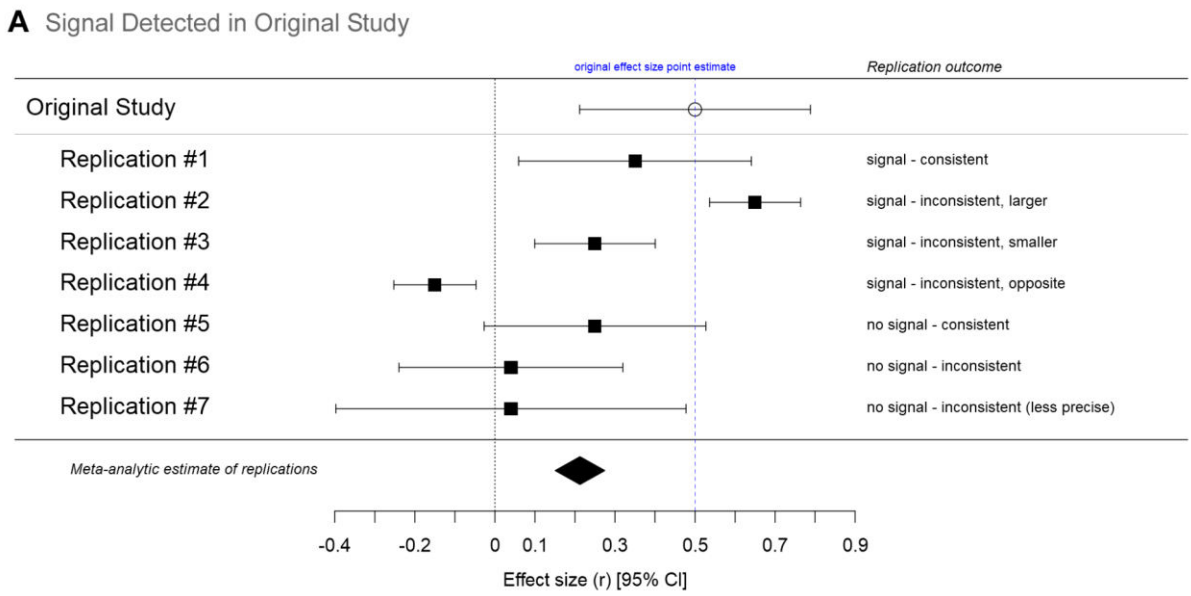


Figure 2. Interpretation criteria for evaluation of replications outcomes by LeBel et al. (2019), if the original study detected a signal. A simplified replication taxonomy for comparing replication effects confidence intervals to target article original effect sizes.

Please refer to Figure 3 when the original experiment has a null hypothesis, or failed to find support for one/some of their hypotheses.

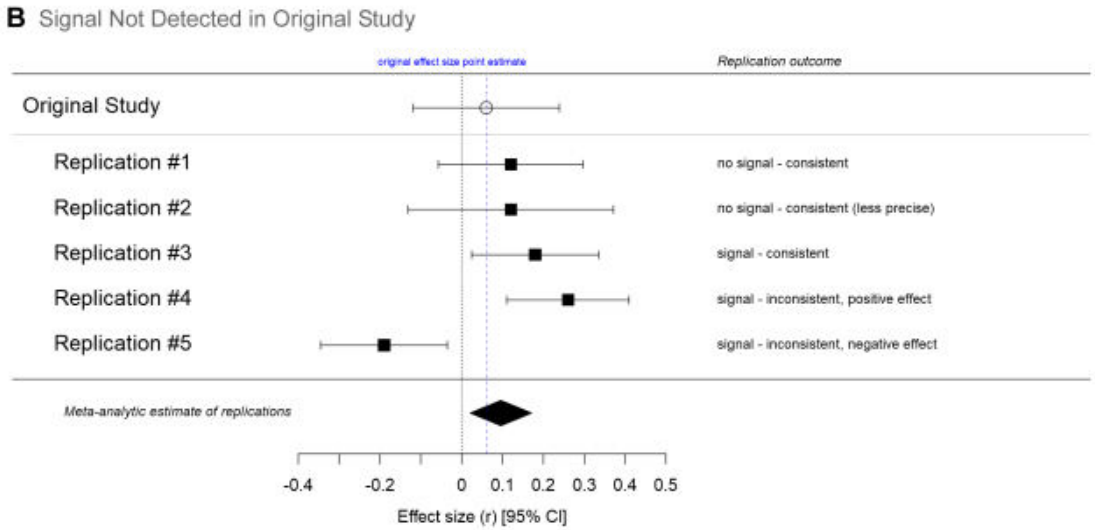


Figure 3. Interpretation criteria for evaluation of replications outcomes by (LeBel et al., 2019), if the original study failed to detect a signal.

Additional information about the study

Data collection details

In order to meet the RAMS inventory from "A Duty to Describe: Better the Devil You Know Than the Devil You Don't" (Brown et al., 2014), we provided the following additional information in the actual data collection.

1. Data collection date: 2022.5.20
2. Data collection procedures:
   We imposed the following settings in recruiting our participants:

1) Participants were paid $1.05 as a fixed participation reward.
2) The expected completion time was set at 8 minutes in advance.
3) We limited all workers' HIT Approval Rate to be between 95% and 100%.
4) We limited each worker's number of HITs approved to be between 5,000 and 100,000.
5) We blocked Suspicious Geocode Locations and Universal Exclude List Workers.
6) We blocked duplicate IP addresses and duplicate geolocation.
7) We enabled HyperBatch so that all eligible workers were able to participate in our HIT immediately after the survey was launched.
8) We restricted workers' location to be in the U.S.

Detailed changes for the JAMOVI data analysis

General changes:

1. Corrected the sequences of the analysis
2. Notes were rearranged with more details to reduce cognitive load
3. Minor word changes/corrections for greater clarity

Specific changes:

1. Age analysis: Added sample size
1. Problem 1, 2, 3, 9: Added percentages for within rows and within columns in the contingency tables
2. Problem 7, 12: Added plots for the independent samples t-test
3. Problem 6, 8, 10, 19: Added descriptives
4. Problem 11: In the repeated measures cells, we corrected the places of Q11-Combined-Ticketspend and Q11-Combined-Ticketgiven.

Effect size and 95% confidence interval calculations of the current replication

See files Replication and extension_Thaler1999.Rmd/html/docx in the OSF folder.

References

Brown, S. D., Furrow, D., Hill, D. F., Gable, J. C., Porter, L. P., & Jacobs, W. J. (2014). A Duty to Describe: Better the Devil You Know Than the Devil You Don't. *Perspectives on Psychological Science*, *9*(6), 626–640. https://doi.org/10.1177/1745691614551749

Heath, C., & Soll, J. (1996). Mental Budgeting and Consumer Decisions. *The Journal of Consumer Research, 23*(1), 40–52. https://doi.org/10.1086/209465

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402. https://doi.org/10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3. https://doi.org/10.15626/MP.2018.843

Leclerc, F., Schmitt, B. H., & Dube, L. (1995). Waiting Time and Decision Making: Is Time like Money? *The Journal of Consumer Research*, *22*(1), 110–119. https://doi.org/10.1086/209439

Olsson-Collentine, A., van Assen, M. A. L. M., & Wicherts, J. M. (2020). Heterogeneity in direct replications in psychology and its association with effect size. *Psychological Bulletin*. https://doi.org/10.1037/bul0000294

Samuelson, P. (1963). Risk and uncertainty: A fallacy of large numbers. *Scientia, 57*(98), 49-56.

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research, 27*(2), 150–162. https://doi.org/10.2307/3172842

Shafir, E., & Thaler, R. H. (1998). 'Invest now, drink later, spend never: the mental accounting of advanced purchases'. *Unpublished working paper: University of Chicago*.

Thaler, R. H.(1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing science*, *4*(3), 199-214.

Thaler, R. H., & Johnson, E. 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

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*(3), 183–206. https://doi.org/10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science (American Association for the Advancement of Science)*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business, 59*(4), 251-278.