**Revisiting mental accounting classic paradigms:**

**Replication Registered Report 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/orpublication 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; ORCID: 0000-0003-2812-6599

## Rights:

CC BY or equivalent license is applied to the AAM arising from this submission. ([clarification](https://bit.ly/rrs-primer))

**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. [https://doi.org/10.1002/(SICI)1099-0771(199909)12:3%3C183::AID-BDM318%3E3.0.CO;2-F](https://doi.org/10.1002/%28SICI%291099-0771%28199909%2912%3A3%3C183%3A%3AAID-BDM318%3E3.0.CO;2-F)

##

## 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, may result in suboptimal decisions or decision-making not aligned with one’s own goals. In a Registered Report with an online U.S. sample recruited from Amazon Mechanical Turk using CloudResearch, we conducted a replication of 17 classic problems reviewed in Thaler (1999) (*N* = ~500 per problem; overall: *N* = 1007). We concluded a mostly successful replication: out of the 17 problems, we found empirical support for 11, mixed empirical support for 3, and no empirical support for 3. Extending the replication, we provided an initial test of four untested predictions described in Thaler (1999), of which we found empirical support for 2, mixed support for 1, and no support for 1. Materials, data, and analysis code are available on: <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 the predictions made and findings reviewed in Thaler (1999) replicate? | For all the problems, participants -on average - follow the pattern of the original findings. | We aimed to recruit a sample size of 1000. Participants were randomized into 9 out of 18 survey blocks, resulting in ~500 participants answering each question. | When the empirical test was clear in the original studies, we followed their data analysis. We added additional analyses when needed. | When we could calculate original studies’ effect sizes, we used the LeBel et al. (2019) criteria to compare replication effects to original’s. Else and in the extensions, we examine signal (p < alpha) and effect direction alignment with hypothesis. |
|  |  |  |  |  |

# Revisiting mental accounting classic paradigms: Replication Registered Report 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, yet mental accounting may lead to decisions that seem to violate fundamental neo-classical economic principles. To summarize this literature, Thaler (1999) wrote a seminal review summarizing over a decade of observations and empirical findings that relate to mental accounting.

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 expenses 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 studies reviewed in Thaler (1999) (replication closeness evaluation based on the LeBel et al., 2018, criteria). Our first goal was to conduct independent close well-powered replications of the classic effects reviewed by Thaler (1999). Our second goal was to empirically test several predictions made in Thaler (1999) that the review did not provide empirical tests for.

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

## 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). In that study, participants were asked whether they would be willing to pay $10 for a ticket following a loss, and the authors contrasted two conditions which manipulated whether the participants had lost a previously purchased ticket for the same show or lost an equivalent $10 bill. The results showed that people were less willing to purchase the ticket after losing a ticket compared to after losing an equivalent cash amount.

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). People categorize funds into different mental accounts designed for different purposes. Participants likely perceive the funds required to "repurchase" the ticket as drawn from the mental account for ticket expenditures, which had already been used in the initial purchase. In contrast, the cash loss was not assigned to a discrete mental account. This distinction violates 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 in Thaler (1999), our target article, the mental accounting phenomenon was further elaborated and expanded into a broader theory of decision-making and choice (Muehlbacher & Kirchler, 2019). In 2011, Soman and Ahn reviewed substantial mental accounting research focusing on 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). A very similar set of papers (e.g., Heath & Soll, 1996; Thaler, 1980; Thaler & Johnson, 1990; Tversky & Kahneman, 1981) served as the basis for Thaler (1999) and the two recent review papers. 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 therefore aimed to revisit the evidence reviewed in Thaler (1999) and re-examine the different subsets of the mental accounting framework.

## Choice of article for replication: Thaler (1999)

We chose the Thaler (1999) article based on three factors: its extensive academic impact, the need for systematic direct replications of many studies covering a single domain, and the potential in methodological improvements in classic older studies.

As of May 2025, the time of writing, there were 5912 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, with the research covered in Thaler (1999) becoming highly influential. We summarized the citations impact of each of the problems covered in the article in Table 1. Thaler received the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel in the year 2017, recognizing mental accounting among Thaler’s most influential work - “Thaler developed the theory of mental accounting, explaining how people simplify financial decision-making by creating separate accounts in their minds, focusing on the narrow impact of each individual decision rather than its overall effect”.

We also recognized the potential for updating and improving both the transparency and the methods used in some of the problems reviewed by Thaler (1999). For example, among the problems we aimed to replicate, several did not report basic methodological details like sample size or descriptives. Statistical analyses were also often not conducted or not reported in sufficient detail for reproducibility. These challenges suggest the need to revisit these problems to reproduce their materials, deduce and improve on their methods, and reassess and update their findings to meet current day best practices.

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 (e.g., Nosek et al., 2022; Zwaan et al., 2018), we embarked on a well-powered pre-registered very close replication of the work reviewed by Thaler (1999).

We note that when we embarked on this Registered Report there were no published systematic attempts for direct replications of the mental accounting findings reviewed in Thaler (1999), and there were no published independent direct pre-registered well-powered replications of Thaler’s own work. During our work on revising Stage 2, a multi-country group (Priolo et al., 2023) released the findings from their multi-country collaboration in which they revisited many mental accounting experiments, with results very similar to the ones we report below.

## Thaler (1999): Hypotheses and findings

Thaler's (1999) review synthesized over a decade of research on the mental accounting phenomenon and we aimed to focus on 17 classic problems he covered. We provided a summary of the original studies and their hypotheses in Table 1 and their findings in Tables 9, 10, 11, and 16.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

For each of the replication problems, we followed the original experimental design with minor adjustments to make those suitable for our target sample (see “deviations from original studies” section). We then added four additional experiments to examine predictions Thaler made that were not reviewed with supporting empirical evidence. We provided a full description of all the problems in the supplementary materials section “Materials and scales used”.

###### Table 1*Problems reviewed in Thaler (1999): Citations, descriptions, and hypotheses*

| **Problem** | **Google scholar citations** | **Description and hypothesis** |  **Explanation** |
| --- | --- | --- | --- |
| 1 | **Tversky and Kahneman, 1986**(790) | **(Risk Taking)** Manipulation with two conditions testing diminishing sensitivity towards gain and loss.**H1:** People are risk-averse for gains and risk-seeking for losses. | People perceive outcomes according to the value function in the prospect theory. Changes in wealth (gains and losses), rather than wealth levels, are effect carriers of values. |
| 2 | **Tversky and Kahneman, 1981**(28623) | **(Time Investment versus Price Reference Point)**Manipulation with two conditions testing reference points.**H2:** People are more likely to spend 20 minutes to save $5 out of $15 than to save $5 out of $125.  | People may frame outcomes in terms of a topical account, where the consequences of possible choices are related to a reference level that is determined by the decision’s context. And these reference points can shift evaluations of value. |
| 3 | **Tversky and Kahneman, 1981**(28623) | **(Theater Play Ticket)**Manipulation with two conditions testing the impact of an existing account on decisions.**H3:** Not explicitly reported**Reconstructed hypothesis:** People are more willing to buy a ticket when they have lost an equivalent amount of cash than when they lost their ticket (different mental accounts). | People may evaluate decisions in a more inclusive account when the outcomes of the act can influence the balance in an account that was previously established by a related act. In general, the sunk cost effect occurs when the decision is referred to an existing account with a negative current balance. |
| 4 | **Thaler, 1985**(9860) | **(Events and Happiness)**Four pairs of scenarios testing the hedonic framing.**H4:** People follow four principles (a) segregate gains, (b) integrate loss, (c) cancel smaller losses against larger gains, and (d) segregate “silver linings” (small gains) from larger losses | People tend to frame outcomes or code combinations of events in ways that make them the happiest. |
| 5 | **Thaler and Johnson, 1990**(3673) | **(Same Day or Two Weeks Apart)**Three pairs of events testing the temporal spacing of hedonic editing.**H5:** People choose to have the events occur “apart” when segregation is preferred, and “together” when integration is hedonically optimal.  | People tend to simplify and encode multiple outcomes in a hedonically optimal manner. The temporal separation will facilitate cognitive segregation, whereas the temporal proximity will facilitate cognitive integration.  |
| 6 | **Thaler and Johnson, 1990**(3673) | **(Emotional Impact of Losing $9)**Manipulation with two conditions testing the effect of a prior loss. **H6:** Not explicitly reported**Reconstructed hypothesis:** People integrate subsequent losses with prior gains but not with prior losses. Also, people are not very sensitive to the exact value of the prior losses when they are within the same magnitude as subsequent losses. | People only follow the hedonic editing rules for part of the time (the quasi-hedonic editing hypothesis).  |
| 7 | **Thaler, 1985**(9860) | **(Location and Price)**Manipulation with two conditions testing the impact of context reference points on willingness to pay.**H7:** People are willing to pay more for the same product if bought from a resort than if bought from a grocery store.  | Consumption decisions are influenced by people’s reference points as set by the context (resort is perceived as a higher reference price than that of a grocery store). |
| 8 | **Thaler, 1985**(9860) | **(Selling Ticket)**Manipulation with three conditions testing the determinants of the reference points. **H8:** Not explicitly reported**Reconstructed hypothesis:** People request a price equal to cost when selling to a friend and a price equal to market price when selling to a stranger, unless their cost exceeds market price.  | Fairness, which largely depends on the cost to the seller, is the dominant factor in determining reference price. |
| 9a  | **Shafir and Thaler, 1998**(8) | **(Wine Bottle)**Manipulation with two conditions testing the value of wine.**H9:** Not explicitly reported**Reconstructed hypothesis:** People exhibit a lack of consensus regarding the perception of the cost.  | People hold mixed perceptions of the value of items when the consumption and purchase are temporally separated.  |
| 10 | **Shafir and Thaler, 1998**(8) | **(Feelings about Purchase)**Three statements examining “investment” purchases. **H10:** Not explicitly reported**Reconstructed hypothesis:** People are more agreeable with the view that the wine purchase is an investment. | For purchases to be consumed in the distant future, people may perceive the expense as an “investment”, thereby avoiding the feeling of spending.  |
| 11 | **Heath and Soll, 1996**b(1211) [Study 2] | **(Previous Events and New Payment)**Manipulation with two conditions testing the underconsumption of a typical target in three contexts, two unrelated, and one related.**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 related purchases. | People set budgets for different accounts and recomputed the remaining budgets periodically. They will decrease further expenses for related activities but less so for unrelated activities.  |
| 12 | **Leclerc et al., 1995**(865) | **(The Performance)**Manipulation with two conditions testing whether the value of time is influenced by price-related characteristics of a decision situation.**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.  | The value of time is influenced by contextual effects postulated by the prospect theory. |
| 13-15 | **Thaler, 1999**(5912) | **(Choices)**A gain (Q13) vs. loss (Q14-15) scenario examining prior outcomes and risky choices. **H13:** The “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 | When gambles are bracketed together, the outcome of the prior gamble can influence subsequent choices.  |
| 16 | **Samuelson, 1963**(1102) | **(Coin Flip Bet)**A scenario testing how bracketing gambles affects the attractiveness of individual bets. **H16**: People shift between single gambles and long-term repeating gambles. | People’s preference follows a piecewise linear version of the value function in the prospect theory. One bet yields negative expected utility, while repetitive bets yield positive expected utility. |
| 17 | **Thaler, 1999** (5912) | **(Division Investment)**Two scenarios examining the myopic loss aversion effect. **H17:** Not explicitly reported**Reconstructed hypothesis:** People are less willing to undertake a single risky investment, but more willing to undertake a portfolio of 25 investments. | Narrow framing inhibits risk-taking, but this can be avoided by aggregation across time or across different divisions.  |

The Google scholar citations were noted in May 2025.
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 266 Google Scholar citations.

b Thaler (1999) referred to Heath and Soll (1996) example of mental accounting of “sports game” and “parking ticket” as different accounts. Sample items and statistics were not provided for that example in Heath and Soll (1996) , and so we tested it using the example that was provided regarding sports and theater tickets as related accounts, and flu inoculation, dinner, and theater as separate accounts.

## Extensions: Prediction extensions

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

###### Table 2*Extension: Summary of predictions made by Thaler (1999) with no reviewed supporting evidence*

|  |  |  |
| --- | --- | --- |
| **Extensions** | **Description and Predictions/Hypothesis** |  |
| Problem 18  | **Thaler, 1980****Game in Bad Weather**Two scenarios testing the sunk cost effect.**H18:** If the family pays for the tickets, they will tend to go despite the weather. If the tickets are given to them, they will tend to stay home. |  |
| Problem 19 | **Thaler, 1980****Membership and Tennis Elbow**Onescenario examining the sunk cost effect.**H19:** Purchasing membership will lead to continued play despite pain.  | Paying for the right to use a service increases the likelihood of utilization. |
| Problem 20a  | **Thaler, 1999****Price and Decision**Two statements testing how sunk costs affect subsequent decisions. **H20:** The more one paid for the shoes, the more times one will try to wear them. Eventually one stops trying. But the more one paid, the longer one will keep the shoes before throwing them away. | The effect of sunk costs on subsequent decisions is not persistent. |
| Problem 21b  | **Thaler, 1999****Annual membership**Manipulations with three conditions testing expenses framing.**H21:** Membership phrased as “merely 27 cents a day” will be more attractive.  | People tend to ignore small, routine expenses.  |

*Note*. The papers listed are the sources of the predictions yet 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 the Open Science Framework (OSF): <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. [To be updated after Stage 2 endorsement:] It has then gone through peer review and officially endorsed by Peer Community in Registered Reports ([Endorsement citation]; [Endorsement link]). All measures, manipulations, exclusions conducted for this investigation are reported, and data collection was completed before analyses. This Registered Report was written using the Registered Report template by Feldman (2023). The current study is part of a mass replications project that received ethical approval from the University of Hong Kong Research Ethics Committee (#EA210265).

# Method

## Power analysis

To ensure that the replication would have sufficient power, we first calculated the effect sizes of the original studies based on the statistics reported. Then, power analysis was conducted with a setting of alpha (two-sided) = 0.05 and power = 0.95. The calculation of effect size and power were with the help of a guide by Jané et al. (2024) and R (Version 4.3.1; R Core team, 2021) using packages "MBESS"and "pwr" (Champely, 2020; Kelley, 2023). The largest required sample size was 321 participants, indicated by the power analysis of Problem 15. We provided more information regarding these calculations in the subsection “Power analysis of original study effect to assess required sample for replication” in the supplementary materials and Rmarkdown code provided in the OSF folder.

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 feedback in Stage 1 regarding the possibility of participant fatigue and the long survey duration, we made a change to our implementation so that each participant was randomized into 9 of the 18 Qualtrics blocks, cutting survey time on average by half. To compensate for that, we doubled our overall target sample to 1000. A sensitivity analysis indicated that a sample of 500 allows the detection of 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 platform (Litman et al., 2017). A total of 1007 participants completed the study (*M*age = 43.28, *SD* = 12.61; 471 females, 526 males, 3 others, and 7 rather not disclose). As participants were randomized to complete 9 out of 18 Qualtrics blocks, there were approximately 500 participants for each problem. We note that 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.

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 first pretested survey duration with 30 participants to test time run estimate and adjusted pay based on the duration. As pre-registered, the data of the 30 participants was included in the final data analysis but was not analyzed independently other than to assess survey completion duration and needed pay adjustments.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Design

We summarized the experimental designs in Table 4. We mapped the designs used in the problems, which included one-sample, between-subject, within-subject, and mixed experimental designs. We set up all the problems using Qualtrics. We had a total of 18 Qualtrics blocks, and empirically related problems were grouped in a single Qualtrics block so that the same participants answered all related problems - 1) Problems 13, 14, and 15, and 2) Problems 18 and 19.

In order to address reviewer’s feedback in Stage 1 to decrease the length of the survey and the cognitive burden on participants, we randomly assigned participants to complete 9 of the 18 blocks. The display of problems and conditions was counterbalanced using the randomizer “evenly present” function in Qualtrics.

We previously tested including many problems in a single data collection using a unified design in many other replications and extensions conducted by our team, and our experience has shown that combining several experiments in a single unified data collection in random order does not seem to impact likelihood of replication success. For example, we successfully ran a similar design in our replications of studies reviewed in influential review papers by Kahneman and Tversky (1972) (Mayiwar et al., 2024), Read et al. (1999) (Wong & Feldman, 2025), Heath et al. (1999) (Au & Feldman, 2020), and Tversky and Kahneman (1971) (Hong & Feldman, 2025). This design seems especially powerful in addressing concerns about the target sample (naivety, attentiveness, etc.) when some studies replicate successfully whereas others do not.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |
|  |  |  |
|  |
|  |  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |  |
|  |
|  |  |  |  |
|  |
|  |  |  |
|  |
|  |  |
|  |
|  |  |  |
|  |
|  |  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |  |  |
|  |

## Procedure

Participants first read a consent form and indicated their willingness to participate, and then answered several verification questions. Three questions assessed their eligibility, understanding, and agreement with study terms, which they had to answer with a “yes” and the required responses in order to proceed to the study. The three questions also served as attention checks, with a randomized display order of the options - 1) “Are you able to pay close attention to the details provided and carefully answer questions that follow?” (yes/no/not sure), 2) “Do you understand the study outline and are willing to participate in a survey with comprehension checks?” (yes/no/not sure), and 3) “Are you a native English speaker born, raised, and currently located in the US?” (yes/no). Failing any of the three questions meant that the participants did not indicate consent and therefore could not embark on the study. Upon completion of these steps, participants proceeded to begin the survey.

Participants were 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.

###### Table 3*Original studies reviewed in Thaler (1999): Summary of samples*

| **Factors** | **Sample size** | **Characteristics** | **Medium and Compensation** |
| --- | --- | --- | --- |
| Current replication | Total: 1007; Problem: ~500  | US American (Median age=40.00 years, Average age=43.28 years, Standard deviation age=12.61 years, age range=20-80 years) | Online via computer |
| Problem 1 | 254 | 126 for Gain condition, 128 for Loss condition | Unreported |
| Problem 2 | 181 | 93 for $15 calculator condition, 88 for $125 calculator condition | Unreported |
| Problem 3 | 383 | 183 for “Lost a bill” condition, 200 for “Lost the ticket” condition | Unreported |
| Problem 4 | 87 | Undergraduate students in a statistical class at Cornell University  | In person |
|  |  |  |  |
| Problem 6 | 137168 | Cornell UG students: 137 for Condition A 1-3Cornell 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 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 | MBA students, 37 male and 30 female | Unreported |
| Problem 13 | Unreported | MBA students | In person |
| Problem 14 | Unreported | MBA students | In person |
| Problem 15 | Unreported | MBA students | In person |
| Problem 16 | 1 | An economist colleague | In person |
| Problem 17 | 26 | A CEO and 25 executives from one firm, each managing a separate division  | In person |

###### Table 4*Replications and extensions experimental designs*

| **Problem design** | **Independent variables** | **Dependent variables** |
| --- | --- | --- |
| Problem 1: between | **Gain condition:** Choices between sure/uncertain *gain***Loss condition:** Choices between sure/uncertain *loss* | Risk taking preference (choice)Risk taking versus risk averse |
|
| Problem 2: between | **$15 calculator condition:** Jacket is $125, calculator is $15**$125 calculator condition:** Jacket is $15, calculator is $125 | Willingness to travel to another store (choice)Yes versus no |
|
| Problem 3: between | **“Lost a bill” condition:** Lost a $10 bill as you enter the theater **“Lost the ticket” condition:** Lost the $10 ticket as you enter the theater  | Willingness to buy (another) ticket (choice)Yes versus no |
|
| Problem 4: one sample proportions | Hedonic framing | Emotionally equivalence (choice)Participants indicate who was happier/more upset. |
| Problem 5: within | Temporal spacing | Emotionally equivalence (choice)Participants indicate who was happier/more unhappy.  |
| Problem 6: between  | Incremental impact of lossManipulation: Different prior outcomes | Emotional Impact of Losing $9 (choice)Participants indicate which event hurts more. |
| Problem 7: between | **Hotel condition:** The soda is sold at a fancy resort hotel.**Grocery store condition**: The soda is sold at a small, run-down grocery store. | Price willing to pay (continuous) |
|
| Problem 8: mixed  | **Cost** (between): Free vs. $5 vs. $10**Buyer** (within): Friend vs. Stranger.**Market price** (within): $5 vs. $10 | Price willing to sell (continuous)Participants indicated their selling price when the customer is a friend/stranger when the going price is $5/$10.  |
|
|
| Problem 9: between | **Drinking condition:**Participants are to imagine drinking a bottle of the wine with dinner**Giving away condition:** Participants are to imagine giving one bottle of the wine to a friend as a gift | Feeling of the cost (choice)Participants indicate which statement best captures their feelings regarding the cost. |
|
| Problem 10: within | Purchase of Bordeaux futures at $400 | Feeling about the purchase (ordinal)Participants indicate which statement best captures their feelings of the purchase.  |
| Problem 11a: mixed: | **Between**: $50 vs. $20**Within** (5 conditions): Dinner-spent vs. Dinner-given vs.Ticket-spent vs. Ticket-given vs. Flu | Willingness to buy a $25 theater ticket later in the week (choice) |
|
| Problem 12b: between | **$15 condition:** The tickets will cost $15 each**$40 condition:** The tickets will cost $40 each | Price willing to pay to avoid waiting (continuous) |
|
| Problem 13: single(within: compared against 14, 15) | (“House money”)Won $30 scenario - loss potential: Risk-seeking vs. risk-averse  | Risk taking (choice)Participants imagine winning $30, then choose between uncertain gain/loss or no further gain/loss. |
| Problem 14: single(within: compared against 13, 15) | Lost $30 - loss potential: Risk-seeking vs. risk-averse  | Risk taking (choice)Participants imagine losing $30, then choose between uncertain gain/loss or no further gain/loss. |
| Problem 15: single(within: compared against 13, 14) | Lost $30 - no loss potential (gain): Risk-seeking vs. risk-averse  | Risk taking (choice)Imagine losing $30, participants then choose between uncertain gain or a sure gain. |
| Problem 16: within | Coin Flip Bet  | Willingness to take the bet (choice)Decision under a single coin flip/100 coin flips  |
| Problem 17: within | Division Investment | Willingness to undertake the investment (choice)Decision under a single project/ a portfolio of 25 projects |
| Problem 18 (extension): within  | The cost of the ticket  | Willingness to go to the game (choice)Decision between go to the game and stay home when the ticket is bought/given |
| Problem 19 (extension): within | Membership at tennis club | Willingness to play (choice)Decision between stop playing and continue playing |
| Problem 20 (extension): within | Shoe purchase scenario | Sunk cost effects (continuous)Participants are to indicate how accurately the statements apply to them. |
| Problem 21(extension): between | **Day expression condition:** Merely 27 cents a day**Year expression condition:** 100 US$ a year**Both expressions condition:**Merely 27 cents a day versus 100 US$ a year | Attractiveness of the membership plan (continuous) |
|
|

We detailed the options of each problem in Tables 9 and 10.

aFor Problem 11, Thaler (1999) did not summarize the study design comprehensively, and we also found the method of the original article difficult to understand.
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.

Problems 13, 14, and 15 were in a single Qualtrics block (grouped together, random order; within-subject design);
Problems 18 and 19 were in a single Qualtrics block (grouped together, random order; within-subject design)

## Manipulations and measures

We summarized all problems and manipulations in Table 4. We summarized the measures and data analysis strategery for all replication problems in Table 5.

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.

##

Our replication deviations from the target article’s studies include participants’ characteristics, delivery mode, and the extensions. We summarized our adjustments and deviations in Tables 7 and 8.

######

###### Table 5*Replication problems: Measures and data analysis strategery*

|  |  | **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 testsMcNemar 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 |
| 8a | 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 $102 within: friend vs. stranger2 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-cost2 within: given vs. spent2 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”.

###### 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-testPaired sample t-test |

###### Table 7*Deviations in replications compared to target studies*

| **Replication**  | **Deviation details**  | **Reasons for change** |
| --- | --- | --- |
| 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 aimed to be as close as possible to the target article. While comparing the option with the loss condition, we suspected it as a typo.  |
| Problem 4 | We slightly simplified the description of the problem and randomized the sequence of the scenarios.  | Minor wording changes for clarity.Randomization to eliminate order bias.  |
| Problem 5, 6  | We revised and simplified the description of the problem and the options. We also randomized the sequence of the scenarios.  | Minor wording changes for clarity.Randomization to eliminate order bias.  |
| Problem 7  | The original study used beer in the two conditions and we changed it to soda. | Some of our targeted participants may not drink.  |
| Problem 9  | Added “Imagine that you enjoy drinking wine” at the beginning of the scenario and randomized the sequence of the statements. | Our targeted population might not enjoy drinking wine.Randomization to eliminate order bias.  |
| 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 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 ensured that participants do not just choose a random option when they cannot understand the question. 3. To reduce cognitive load4. Randomization to eliminate order bias.  |
| Problem 11 | We simplified the instructions for the problem.  | Minor wording changes for clarity and understandability. |
| 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.  |
| Problem 13, 14, 15 | Added “Imagine that…”at the beginning of the scenarios. | To facilitate perspective-taking |
| Problem 21 (extension) | 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).

## 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.

## Exclusions

We did not exclude participants, following our pre-registered plan to focus on the main sample, and so in our analysis we included all the data of those who successfully completed the entire study.

###### 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 |  |
|  |  |  |
| IV stimuli | Similar | Scenarios were slightly adjusted to update to current time and the targeted population.  |
| DV stimuli | Similar | Problem 1 was changed for a suspicious typo and we added another statement in Problem 10. |
| Procedural details | Similar/Different | We combined all studies into a singular unified design, we randomized the order of scenarios/statements in the problems.  |
| Physical settings | Different | The current replication was conducted online via Qualtrics. |
| Population (e.g. age) | Different | We collected data online from MTurk using CloudResearch.  |
| 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.  |

#

# Results

We conducted our analyses on the full sample (pre-registered) using R (Version 4.3.2) and the JAMOVI “jmv” (Selker et al., 2022).

 In Tables 9-11, 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 12-15. We compared the original studies’ effects to ours in Table 16.

###### Table 9*Descriptive statistics for choice problems*

| **Problem** | **Conditions and Options**  | **Replication** | **Original**  | **Replication Interpretation** |
| --- | --- | --- | --- | --- |
| *n* | Count  | Percentage |  |  | Percentage |
| 1 | **Gain:**  |  |  |  |  | Successful |
| A sure gain of $100 | 250 | 209 | 84% | 72% |
| A 50 % chance to gain $200 and a 50% chance to gain $0 |  | 41 | 16% | 28% |
| **Loss:** |  |  |  |  |
| A sure loss of $100  | 254 | 112 | 44% | 36% |
|  | A 50% chance to lose $200 and a 50% chance to lose $0  |  | 142 | 56% | 64% |
| 2 | **$15 Calculator:** |  |  |  |  | *Unsuccessful, no signal*Original showed a reversal. Replication no difference and generally low rates of willingness to make the trip. |
|  | Make the trip | 253 | 29 | 11% | 68% |
|  | Not making the trip |  | 224 | 89% | 32% |
|  | **$125 Calculator:** |  |  |  |  |
|  | Make the trip  | 252 | 17 | 7% | 29% |
|  | Not making the trip |  | 235 | 93% | 71% |
| 3 |  |  |  |  |  | Successful |
| Buy the ticket | 252 | 228 | 90% | 88% |
| Not buying the ticket |  | 24 | 10% | 12% |
| **Lost the ticket:** |  |  |  |  |
|  | 251 | 194 | 77% | 46% |
| Not buying another ticket |  | 57 | 23% | 54% |
| 4 | **A. Two wins: $50, and $25 versus One win: $75.**  | 505 |  |  |  |  |
|  | Two wins is happier | 178 | 35% | 64% | Successful |
|  | One win is happier | 62 | 12% | 18% |  |
|  | No difference | 265 | 52% | 17% |  |
|  | **B. Two mistakes: $100, and $50 versus One mistake: $150.** |  |  |  |  |
|  | Two mistakes is more upset | 213 | 42% | 76% | Successful |
|  | One mistake is more upset | 69 | 14% | 16% |  |
|  | No difference  | 223 | 44% | 8% |  |
|  | **C. Two events: Win $100, and loss $80 versus One event: Win $20.**  |  |  |  |  |
|  | Two events is happier | 61 | 12% | 25% | Successful |
|  | One event is happier | 386 | 76% | 70% |  |
|  | No difference  | 58 | 11% | 5% |  |
|  | **D. Two events: Loss $200, and win $25 versus One event: Loss $175.**  |  |  |  |  |
|  | Two events is more upset | 63 | 12% | 22% | Successful |
|  | One event is more upset | 316 | 63% | 72% |  |
|  | No difference  | 126 | 25% | 6% |  |
| 5 | **A. Two events: (1) win $25 (2) win $50.**  | 495 |  |  |  |  |
|  | Happier on the same day | 200 | 40% | 25% | *Unsuccessful, opposite* |
|  | Happier two weeks apart  | 174 | 35% | 63% |  |
|  | No difference  | 121 | 24% | 12% |  |
|  | **B. Two events: (1) $100 must be paid (2) $50 must be paid.**  |  |  |  |  |
|  | More unhappy on the same day | 303 | 61% | 57% | Successful |
|  | More unhappy two weeks apart | 96 | 19% | 34% |  |
|  | No difference  | 96 | 19% | 9% |  |
|  | **C. Two events: (1) a $20 parking ticket (2) a $25 bill.**  |  |  |  |  |
|  | More unhappy on the same day  | 278 | 56% | 75% | Successful |
|  | More unhappy two weeks apart | 115 | 23% | 17% |  |
|  | No difference  | 102 |  | 21% | 7% |  |
|  |  |  |  |  |  |  |
|  |  | **1. (A) Lose $9. (B) Lose $9 after having gained $30.** |  |  |  |  | Successful |
|  |  | 253 | 208 | 82% | 70% / 84% |  |
|  |  |  | 32 | 13% | 9% / 10% |  |
|  |  |  | 13 | 5% | 21% / 6% |  |
|  | **2. (A) Lose $9. (B) Lose $9 after having lost $30.** |  |  |  |  | Successful |
|  |  | 253 | 19 | 8% | 13% / 22% |  |
|  |  |  | 218 | 86% | 55% / 75% |  |
|  |  |  | 16 | 6% | 31% / 3% |  |
|  | **3. (A) Lose $9. (B) Lose $9 after having lost $250.** |  |  |  |  | *Unsuccessful, opposite* |
|  |  | 253 | 32 | 13% | 39% / 54% |  |
|  |  |  | 209 | 83% | 38% / 37% |  |
|  | No difference |  | 12 | 5% | 23% / 9% |  |
|  | **4. (A) Lose $9. (B) Lose $9 after suffering a loss of $1000.** |  |  |  |  | *Unsuccessful, opposite* |
|  |  | 253 | 37 | 15% | 50% |  |
|  |  |  | 201 | 79% | 33% |  |
|  | No difference |  | 15 | 6% | 17% |  |
|  | **5. (A) Lose $9 after suffering a loss of $30. (B) Lose $9 after suffering a loss of $1000.** |  |  |  |  | *Unsuccessful, opposite* |
|  | A hurts more | 253 | 51 | 20% | 51% |  |
|  |  |  | 184 | 73% | 38% |  |
|  | No difference  |  | 18 | 7% | 21% |  |
| 6a |  |  |  |  |  |  |
|  | **1. (A) Lose $9. (B) Lose $9 after suffering a loss of $9.** |  |  |  |  | Successful |
|  |  | 252 | 9 | 4% | 7% |  |
|  |  |  | 209 | 83% | 64% |  |
|  | No difference  |  | 34 | 13% | 28% |  |
|  | **2. (A) Lose $9. (B) Lose $9 after suffering a loss of $18.** |  |  |  |  | Successful |
|  |  | 252 | 9 | 4% | 11% |  |
|  | B hurts more |  | 231 | 92% | 65% |  |
|  | No difference |  | 12 | 5% | 23% |  |
|  | **3. (A) Lose $9. (B) Lose $9 after suffering a loss of $36.** |  |  |  |  | Successful |
|  |  | 252 | 11 | 4% | 12% |  |
|  |  |  | 228 | 90% | 62% |  |
|  | No difference |  | 13 | 5% | 26% |  |
|  | **4. (A) Lose $9. (B) Lose $9 after suffering a loss of $45.**  |  |  |  |  | Successful |
|  |  | 252 | 10 | 4% | 14% |  |
|  |  |  | 230 | 91% | 65% |  |
|  | No difference |  | 12 | 5% | 21% |  |
|  | **5. (A) Lose $9 after suffering a loss of $9. (B) Lose $9 after suffering a loss of $36.** |  |  |  |  | Successful |
|  |  | 252 | 16 | 6% | 7% |  |
|  |  |  | 224 | 89% | 68% |  |
|  | No difference |  | 12 | 5% | 25% |  |
| 9b |  |  |  |  |  | Successful.$75 is considered the correct answer, and both original and replication show most participants did not choose the correct answer. |
| $0 | 254 | 75 | 30% | 30% |
| $20 |  | 52 | 20% | 17% |
| $20 plus interest  |  | 14 | 6% | 9% |
| $75 |  | 53 | 21% | 30% |
| A $55 saving |  | 60 | 24% | 14% |
| **Drinking:**  |  |  |  |  |
| $0 | 251 | 57 | 23% | 30% |
| $20 |  | 54 | 22% | 18% |
| $20 plus interest  |  | 16 | 6% | 7% |
| $75  |  | 64 | 25% | 20% |
| A $55 saving |  | 60 | 24% | 25% |
|  | **$50 high cost:**  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  | 254 | 101 | 40% | / |  |
|  |  |  | 153 | 60% | / |  |
|  |  **later in the week?** |  |  |  |  |  |
|  |  | 254 | 211 | 83% | / |  |
|  |  |  | 43 | 17% | / |  |
|  |  |  |  |  |  |  |
|  |  | 254 | 70 | 28% | / |  |
|  |  |  | 184 | 72% | / |  |
|  |  |  |  |  |  |  |
|  |  | 254 | 198 | 78% | / |  |
|  |  |  | 56 | 22% | / |  |
|  |  |  |  |  |  |  |
|  |  | 254 | 75 | 30% | / |  |
|  |  |  | 179 | 70% | / |  |
| 11c | **$20 low cost:** |  |  |  |  | Partially successful |
|  |  |  |  |  |  |  |
|  |  | 252 | 138 | 55% | / |  |
|  |  |  | 114 | 45% | / |  |
|  |   |  |  |  |  |  |
|  |   | 252 | 213 | 85% | / |  |
|  |   |  | 39 | 15% | / |  |
|  | **Spent $20 on a sports ticket. Would you purchase a $25 theater ticket later in the week?**  |  |  |  |  |  |
|  |   | 252 | 103 | 41% | / |  |
|  |   |  | 149 | 59% | / |  |
|  |   |  |  |  |  |  |
|  |   | 252 | 201 | 80% |  |  |
|  |  |  | 51 | 20% |  |  |
|  | **Spent $20 on an inoculation. Would you purchase a $25 theater ticket later in the week?**  |  |  |  |  |  |
|  |   | 252 | 109 | 43% | / |  |
|  |   |  | 143 | 57% | / |  |
| 13 | **Imagine that you have just won $30.** | 504 |  |  |  | *Not successful* |
| A 50% chance to gain $9 and a 50% chance to lose $9.  | 143 | 28% | 70% |
| No further gain or loss.  | 361 | 72% | 30% |
| 14 | **Imagine that you have just lost $30.** | 504 |  |  |  | Successful |
| A 50% chance to gain $9 and a 50% chance to lose $9. | 85 | 17% | 40% |
| No further gain or loss.  | 419 | 83% | 60% |
| 15 | **Imagine that you have just lost $30.** | 504 |  |  |  | *Not successful* |
| A 33% chance to gain $30 and a 67% chance to gain nothing. | 119 | 24% | 60% |
| A sure $10.  | 385 | 76% | 40% |
| 16d | **A single coin flip, heads you win $200, tail you lose $100.**  | 506 |  |  |  | Successful |
| 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.**  |  |  |  |
| Take the bet. | 247 | 49% | / |
| Not taking the bet | 259 | 51% | / |
| 17e | **A project: 50% chance to gain $2 million, 50% chance to lose $1 million**. | 504 |  |  |  |  |
| Undertake the project | 152 | 30% | 12% |  |
| Not undertaking the project | 352 | 70% | 88% | Successful |
|  | **A portfolio of 25 of investments: Each has a 50% chance of gaining $2 million and 50% chance of losing $1 million.** |  |  |  |  |  |
|  | Undertake the investments |  | 236 | 47% | / |  |
|  | Not undertaking the investments |  | 268 | 53% | / |  |
| 18 | **Paid $40 for tickets:** | 502 |  |  | N/A | Extension: supported |
| Go to the game | 160 | 32% |
| Stay home | 342 | 68% |
| **Tickets given by friends:** |  |  |
| Go to the game | 68 | 14% |
| Stay home | 434 | 86% |
| 19 | **Imagine that you joined a tennis club and paid a $300 yearly membership fee.**  |  |  |  | N/A | Extension: *not supported* |
| Stop playing  | 502 | 380 | 76% |
| Continue to play  |  | 122 | 24% |

*Note.* *n* represents sample size per condition .

a In Problem 6, Condition A-1, 2, 3 had two samples, so results are both presented, separated by a slash. The statistical results reported in Condition A-5 added up to 110% rather than 100%, suggesting a possible reporting mistake in the original article. For the calculation of the effect size we will assume “no differences” is equal to 11% and not 21%.

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 under-consume 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |

d

e In Problem 17, 3 of the 25 executives accepted the single investment, and the CEO accepted the portfolio of 25 of these investments.

###### Table 10*Descriptive statistics for problems with scale/text entry and a between-subjects design*

| **Problem**  | **Condition** | **Replication** | **Original findings** | **Replication Interpretation** |
| --- | --- | --- | --- | --- |
|  |  | *N* | *Mean* | *SD* |
| 7 | Hotel (fancy) purchase |  | 254254 | 7.09 | 19.85 | *Median*=$5*Median*=$3 | Successful |
| Grocery (run-down) purchase |  |  | 4.17 | 3.98 |
| 8  | Free ticket | Market value $5 |  | 166 | 2.37 | 3.49 | 68% answer 0, 26% answer 5, 3% answer 10, and 3% answer Other  | Successful |  |  |
| Stranger | 6.04 | 5.24 | 6% answer 0, 77% answer 5, 10% answer 10, and 6% answer Other  |  |  |
| Market value $10 |  | 4.26 | 5.43 | 65% answer 0, 26% answer 5, 6% answer 10, and 3% answer Other  |  |  |
| Stranger | 10.32 | 6.79 | 6% answer 0, 16% answer 5, 58% answer 10, and 19% answer Other  |  |  |
| Paid 5 | Market value $5 |  | 169 | 3.72 | 2.47 | 14% answer 0, 79% answer 5, 0% answer 10, and 7% answer Other  |  |  |
| Stranger | 6.68 | 3.49 | 0% answer 0, 79% answer 5, 7% answer 10, and 14% answer Other  |  |  |
| Market value $10 |  | 6.12 | 4.20 | 7% answer 0, 79% answer 5, 4% answer 10, and 9% answer Other |  |  |
| Stranger | 11.51 | 5.59 | 0% answer 0, 14% answer 5, 57% answer 10, and 29% answer Other  |  |  |
| Paid 10 | Market value $5 |  | 162 | 5.01 | 2.68 | 0% answer 0, 69% answer 5, 23% answer 10, and 8% answer Other  |  |  |
| Stranger | 8.19 | 3.58 | 0% answer 0, 42% answer 5, 46% answer 10, and 12% answer Other  |  |  |
| Market value $10 |  | 7.55 | 3.82 | 0% answer 0, 15% answer 5, 69% answer 10, and 15% answer Other  |  |  |
| Stranger | 11.43 | 4.01 | 0% answer 0, 0% answer 5, 73% answer 10, and 27% answer Other |  |  |
| 12 | $15 condition |  | 235 | 8.14 | 11.32 | People are willing to pay twice as much to avoid waiting for the $40 ticket than for the $15 ticket (=$7.20 vs. =$3.60, *t*=1.92(39), *p*=.06).  | Successful |
| $40 condition |  | 222 | 10.34 | 7.66 |
| 21 | Day expression |  | 167 | 44.53 | 32.32 | N/A | Extension: supported. |
| Year expression |  | 166 | 26.04 | 28.05 |
| Both expressions | Day | 170 | 45.16 | 31.71 |
| Year |  | 170 | 35.72 | 29.05 |

*SD* represents standard deviation.

###### Table 11*Comparison and descriptive statistics for problems with scale and a within-subjects design*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Problem**  | **Options** | **Replication** | **Original findings** | **Replication Interpretation** |
|  |  | *N* | *Mean* | *SD* |
| 10a | I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway. |  | 502 | 2.98 | 1.45 | *Mean*=3.31  | Successful |
| I feel like I made a $400 investment which I will gradually consume after a period of years. | 3.56 | 1.30 | *Mean*= 1.94 |  |  |  |
| I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered.  |  |  | 3.08 | 1.36 | *Mean*=2.88 |  |
| I cannot understand this question. |  | 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 | N/A | Extension: *Not supported* |
| 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. |  |  | 3.45 | 1.33 |

*Note.* *N* represents sample size. *SD* represents standard deviation.

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 12*Problems 1-6, 9, and 19: Summary of* ***χ²*** *tests*

| **Problem**  | **Conditions and Sub Questions** | ***χ²*** | ***df*** | ***p*** |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | One-sample, single condition | 85.03 | 1 | < .001 |  |  |  |  |
| 2 | One-sample, single condition | 3.39 | 1 |  |  |  |  |
| 3 | One-sample, single condition | 16.18 | 1 | < .001 |  |  |  |  |
|  | 4 | A. Two wins: $50, and $25 versus One win: $75.  | 123.24 | 2 | < .001 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | B. Two mistakes: $100, and $50 versus One mistake: $150 | 88.22 | 2 | < .001 |
|  |
|  | C. Two events: Win $100, and loss $80 versus One event: Win $20 | 422.21 | 2 | < .001 |
|  |
|  | D. Two events: Loss $200, and win $25 versus One event: Loss $175 | 206.10 | 2 | < .001 |
|  |
| 5 | A. Two events: (1) win $25 (2) win $50 | 19.65 | 2 | < .001 |
|  | B. Two events: (1) $100 must be paid (2) $50 must be paid.  | 173.13 | 2 | < .001 |
|  | C. Two events: (1) a $20 parking ticket (2) a $25 bill  | 116.59 | 2 | < .001 |
| 6 |  |  |  |  |
|  | 1. (A) Lose $9. (B) Lose $9 after having gained $30. |  | 2 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2. (A) Lose $9. (B) Lose $9 after having lost $30. | 317.84 | 2 | < .001 |
| 3. (A) Lose $9. (B) Lose $9 after having lost $250. | 278.81 | 2 | < .001 |
| 4. (A) Lose $9. (B) Lose $9 after suffering a loss of $1000. | 244.96 | 2 | < .001 |
| 5. (A) Lose $9 after suffering a loss of $30. (B) Lose $9 after suffering a loss of $1000. | 183.14 | 2 | < .001 |
| Second group of questions: |  |  |  |
| 1. (A) Lose $9. (B) Lose $9 after suffering a loss of $9. |  | 2 | < .001 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2. (A) Lose $9. (B) Lose $9 after suffering a loss of $18. | 385.93 | 2 | < .001 |
| 3. (A) Lose $9. (B) Lose $9 after suffering a loss of $36. | 370.31 | 2 | < .001 |
| 4. (A) Lose $9. (B) Lose $9 after suffering a loss of $45.  | 380.67 | 2 | < .001 |
| 5. (A) Lose $9 after suffering a loss of $9. (B) Lose $9 after suffering a loss of $36. | 350.10 | 2 | < .001 |
|
|  | Giving away | 333 | 4 |  |
| 9 | Drinking | 298 | 4 |  |
|  | Giving away versus drinking |  | 4 |  |
| 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 13*Problems 5, 13-18: Summary of paired-samples McNemar tests*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Problem**  | **Comparisons** | ***χ²*** | ***df*** | ***p*** | **Cohen’s g** |  |
| 5a | Comparing A to BA: (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 | / |  |
| Comparing A to CA: (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 | / |  |
| 13-14 | Problems 13 vs. 14*Won* $30 loss potential vs. *Lost* $30 loss potential | 22.73 | 1 | < .001 | 0.20 [0.12, 0.26] |  |
| 14-15 | Problems 14 vs. 15 (exploratory)Lost $30 *loss* potential vs. Lost $30 *no loss* potential | 7.9 | 1 | .005 | 0.12 [0.04, 0.20] |
| 13-15 | Problems 13 vs. 15*Won* $30 *loss* potential vs. *Lost* $30 *no loss* potential | 3.27 | 1 | .070 | 0.07 [-0.01, 0.14] |  |
| 16 | 1 bet vs. 100 bets | 84.90 | 1 | < .001 | 0.36 [0.30, 0.40] |  |
| 17 | 1 investment vs. 25 investments | 45.82 | 1 | < .001 | 0.27 [0.20, 0.33] |  |
| 18 | Paid $40 vs. Given by friends | 86.37 | 1 | < .001 | 0.47 [0.41, 0.49] |  |

*Note. df* indicates degree of freedom.

a Problem 5 compared the same day to two weeks apart, higher same day for negative than for positive.

###### Table 14*Problems 7, 12, 20, and 21: Summary of all t-tests results*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Problem and Test Type** | **Statistic** | ***df*** | ***p*** | **Mean difference** | **SE difference** | **Effect size** | **95% CI** |  |
| 7 Independent samples t-test | Student’s *t*=2.30 | 506 | .011 | 2.92 | 1.27 | 0.20 | [0.03, 0.38] |  |
| 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] |  |
| 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] |  |
| Statement 2: Student’s *t*=7.53 | 506.00 | < .001 | 0.45 | / | 0.33 | [0.24, 0.42] |
| 21 Independent samples t-test Paired sample t-test | Student’s *t*=5.57 | 331 | < .001 | 18.48 | 3.32 | 0.61 | [0.39, 0.83] |  |
| Welch’s *t*=5.57 | 325.07 | < .001 | 18.48 | 3.32 | 0.61 | / |
| 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.
Effect size for independent samples t-test is Cohen’s *d*, effect size for paired sample t-test is Cohen’s *dz*.

###### Table 15*Problems 8, 10, 11: Summary of all ANOVA results*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Problem and Test Type** | **Source of variation** | ***SS*** | ***df*** | ***MS*** | ***F*** | ***p*** | ***η²*** |
| 8 Mixed ANOVA | Friend vs. Stranger | 8718.31 | 1 | 8718.31 | 594.75 | < .001 | 0.16 |
|  | Market worth $5 vs. $10 | 5072.36 | 1 | 5072.36 | 870.07 | < .001 | 0.09 |
| Free ticket vs. $5 vs. $10 | 1736.17 | 2 | 868.08 | 15.69 | < .001 | 0.03 |
| Friend vs. Stranger x Free ticket vs. $5 vs. $10 | 146.59 | 2 | 73.29 | 5.00 | .007 | 0.00 |
| Friend vs. Stranger x Market worth $5 vs. $10 | 418.89 | 1 | 418.89 | 205.04 | < .001 | 0.01 |
| Market worth $5 vs. $10 x Free ticket vs. $5 vs. $10 | 47.91 | 2 | 23.95 | 4.11 | .017 | 0.00 |
| Friend vs. Stranger x Market worth $5 vs. $10 x Free ticket vs. $5 vs. $10 | 79.70 | 2 | 39.85 | 19.50 | < .001 | 0.00 |
| 10 Repeated measures ANOVA | Accuracy of feelings  | 98.08 | 2 | 49.04 | 25.26 | < .001 | 0.03 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |
| --- |
|  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 11 Mixed ANOVA | Given vs. Spent | 83.36 | 1 | 83.36 | 382.59 | < .001 | 0.17 |
|  | Dinner vs. Ticket | 4.09 | 1 | 4.09 | 55.52 | < .001 | 0.01 |
| High cost vs. Low cost | 3.15 | 1 | 3.15 | 7.37 | .007 | 0.01 |
| Given vs. Spent x High cost vs. Low cost | 1.98 | 1 | 1.98 | 9.11 | .003 | 0.00 |
| Given vs. Spent x Dinner vs. Ticket | 0.83 | 1 | 0.83 | 16.49 | <.001 | 0.00 |
| Dinner vs. Ticket x High cost vs. Low cost | 0.01 | 1 | 0.01 | 0.08 | .783 | 0.00 |
| Given vs. Spent x Dinner vs. Ticket x High cost vs. Low cost | 0.01 | 1 | 0.01 | 0.26 | .610 | 0.00 |

*Note. SS* represents Type 3 sums of squares, *df* represents degree of freedom, and *MS* represents mean square

##

### Problem 1 (Tversky and Kahneman, 1986) [Framing effect]

In our replication of Problem 1, we found support for Tversky and Kahneman (1986)’s findings that people were more risk-averse for gains (84%) than for losses (56%; *χ²*(1) = 85.03, *p* < .001; Cramer’s V = 0.37 [0.22, 0.54]).

### Problem 2 (Tversky and Kahneman, 1981)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |
| --- |
|  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

##

In our replication of Problem 2, we failed to find support for Tversky and Kahneman (1981)’s findings that people were more inclined to spend 20 minutes to save $5 out of $15 (save $5 out of $15: 11%; save $5 out of $125: 7%; *χ²*(1) = 3.39, *p* = .066; Cramer’s V = 0.06 [0.00, 0.17]).

### Problem 3 (Tversky and Kahneman, 1981) [Sunk-cost effect]

In our replication of Problem 3, we found support for Tversky and Kahneman (1981)’s findings that sunk costs impact decisions less when sunk costs are not from the same mental account (90% bought ticket after losing $10), compared to when sunk costs are from the same mental account but less so for an unrelated act (77% bought ticket after losing a ticket; *χ²*(1) = 16.18, *p* < .001; Cramer’s V = 0.17 [0.09, 0.26]).

### Problem 4: Hedonic editing (Thaler, 1985)

In our replication of Problem 4, we found support for Thaler (1985)’s findings that people prefer to segregate gains (Cramer’s V = 0.26 [0.20, 0.33]), integrate losses (Cramer’s V = 0.23 [0.17, 0.29]), integrate smaller losses with larger gains (Cramer’s V = 0.43 [0.37, 0.49]), and segregate small gains from large losses (Cramer’s V = 0.31 [0.25, 0.37]). Effects were consistent with yet weaker than the original studies, and a larger proportion of the participants in the replication were indifferent.

### Problem 5 (Thaler and Johnson, 1990)

In our replication of Problem 5, we found support for Thaler and Johnson (1990)’s findings that integrated losses were less upsetting than separate losses (5b: Cramer’s V = 0.28 [0.21, 0.34]; 5c: Cramer’s V = 0.22 [0.17, 0.29]), but opposite findings to the original’s that separate gains were happier than integrated gains (5a: Cramer’s V = 0.09 [0.03, 0.16]).

### Problem 6 (Thaler and Johnson, 1990)

In our replication of Problem 6, we found some support for Thaler and Johnson (1990)’s findings: seven effects replicated well with larger effects, whereas three effects were in the opposite direction. Participants perceived the loss of $9 as less upsetting when it occurs after a prior gain (Cramer’s V = 0.50 [0.41, 0.59]) but as more upsetting if the $9 loss followed prior losses, yet - inconsistent with the target article, the magnitude of the prior losses did not seem to have much impact (Cramer’s V = 0.47 - 0.60). When prior losses were compared directly in the comparison between losing $9 after a loss of either $30 or $1000 - participants perceived the loss following $1000 to be more painful, opposite from the target’s findings.

Regardless of the comparison between which hurts more, the core argument is that people are not indifferent to prior losses and our findings support this idea - participants indicated that they think they would not ignore prior losses and integrate this in their evaluations of loss.

### Problem 7 (Thaler, 1985)

In our replication of Problem 7, we found support for Thaler (1985)’s findings that people are willing to adjust their spending based on the purchasing context, people were willing to pay higher prices for the same soda in a fancy resort hotel than in a grocery store (*t*(374.55) = 6.18, *p* < .001; Cohen’s *d* = 0.20 [0.03, 0.38]).

### Problem 8 (Thaler, 1985)

In our replication of Problem 8, we found support for Thaler (1985)’s findings that people request a price equal to cost when selling to a friend and a price equal to market price when selling to a stranger (unless their cost exceeds market price) (Friend vs. Stranger: *F* = 594.75, *p* < .001, *η²p =* 0.55 [0.49, 0.59]; interaction of the three factors - buyer, cost, and market price: *F* = 19.51, *p* < .001, *η²p =* 0.07 [0.03, 0.12]).

The interaction pattern generally captured the spirit of having different mental accounts used for determining asking price when selling to friends versus to strangers, yet deviated slightly from the reported findings, as can be seen in Figure 1. Amount paid and market price impacted the asking price both for a friend and for a seller, yet the asking price for a stranger was indeed closer to the market price whereas the asking price for a friend was closer to cost.

###### Figure 1*Problem 8: Asking price - interaction between buyer closeness, amount paid, and market price*



Plot using JAMOVI “jmv” R package (Selker et al., 2022)

### Problem 9 (Shafir and Thaler, 1998)

In our replication of Problem 9, we found support for Shafir and Thaler’s (1998) findings that people’s estimation of the cost of a bottle of wine - which gains value over time - differs from a rational economic assessment in which the cost reflects current market price (giving away: *χ*²(4) = 333, *p* < .001, Cramer’s V = 0.80 [0.72, 0.89]; drinking: *χ*²(4) = 298, *p* < .001, Cramer’s V = 0.77 [0.68, 0.85]; comparing giving away versus drinking: *χ*²(4) = 3.64, *p* = .457, Cramer’s V = 0.00 [0.00, 0.14]).

### Problem 10: (Shafir and Thaler, 1998)

In our replication of Problem 10, we found support for Shafir and Thaler (1998)’s findings that people tend to perceive the wine purchase as an investment (Spent $400: *M* = 2.98, Invest $400: *M* = 3.56, *SD* = 1.30; Save $100: *M* = 3.08, *SD* = 1.36; *F*(2, 1002) *=* 25.26, *η²p* = 0.05).

### Problem 11 (Heath and Soll, 1996; Study 2)

Heath and Soll (1996) examined whether people would spend $25 on a theater ticket, manipulating two factors about events prior to the decision: 1) whether the prior event was an expenditure or a gift, and 2) whether the event was related (sports ticket) or unrelated (dinner / flu vaccination) to the theater ticket. They showed that people were less willing to spend money on a theater ticket the more they previously spent, and especially if it was spent on something that seems related - such as a sports event ticket, compared to something that was unrelated.

In our replication of Problem 11, we found mixed support for their hypothesis. We found the intended pattern for dinner versus. sports ticket, participants were more willing to buy a theater ticket when they spent their money on the unrelated dinner compared to the related sports event ticket (Cohen’s *d* = 0.26 [0.13, 0.39]) , yet the findings of the unrelated flu vaccination mirrored that of the related sports event ticket (Cohen’s *d* = 0.05 [-0.08, 0.17). We plotted the results in Figure 2.

###### Figure 2*Problem 11: Willingness to buy theater ticket after related versus unrelated and given versus spent activities*



###

### Problem 12 (Leclerc et al., 1995)

In our replication of Problem 12, we found support for Leclerc et al. (1995)’s findings that the price of a ticket influences people’s willingness to pay for the time to wait to obtain that ticket ($15 ticket: *M* = 8.14, *SD* = 11.32, vs. $40 ticket: *M* = 10.34, *SD* = 7.66; Cohen’s *d* = 0.23 [0.04, 0.41]).

### Problems 13-15 (Thaler, 1999)

In Problems 13-15, Thaler (1999) manipulated gain and loss scenarios and showed that the outcomes of prior gambles could influence subsequent decisions. In our replication of Problems 13-15, we found mixed evidence for the hypotheses.

We did not find support for prior gains as stimulating risk-seeking behavior (Problem 13: 28%), yet we found support for a greater inclination towards risk-taking when there is a prior loss (Problem 14: 17% vs Problem 15: 24%). Going beyond the pre-registered singular problem statistical analyses, in Stage 2, we also contrasted each of the three problems and found that prior loss stimulated stronger risk aversion than prior gain (Problem 14 > Problem 13; Cohen’s *g* = 0.20 [0.12, 0.26]), and that in cases of similar prior loss, there was higher risk aversion when there was an additional loss potential (Problem 14 > Problem 13: Cohen’s *g* = 0.12 [0.04, 0.20]; no support for differences between Problems 13 and 15: Cohen’s *g* = 0.07 [-0.01, 0.14]).

### Problem 16 (Samuelson, 1963)

In our replication of Problem 16, we found support for Samuelson (1963)’s findings that bracketing gambles together increases the attractiveness of individual bets (single coin flip: 25%; 100 coin flips: 49%; Cohen’s *g* = 0.36 [0.30, 0.40]).

### Problem 17 (Thaler, 1999)

In our replication of Problem 17, we found support for Thaler (1999)’s finding that narrow framing suppresses risk-seeking behaviors (single project: 30%; 25 investments: 47%; Cohen’s *g* = 0.27 [0.20, 0.33]).

## Extensions testing review’s untested predictions

### Problem 18 (Thaler, 1980)

Thaler (1980) proposed that families would be more inclined to go to a basketball game in bad weather when they paid for the tickets, compared to when they received the tickets as a gift. In our Problem 18 extension, we found support for this prediction (paid $40: 32%; given by friends: 14%; Cohen’s *g* = 0.47 [0.41, 0.49]).

### Problem 19 (Thaler, 1980)

Thaler (1980) suggested that due to the sunk cost effect, people are more inclined to continue playing despite pain if they have already paid a membership fee. However, in our Problem 19 extension, we failed to find support for the prediction, where 76% of the participants chose to stop playing (*χ*² = 132.60, *p* < .001, opposite direction to prediction; Cramer’s V = 0.26 [0.20, 0.33]).

### Problem 20 (Thaler, 1999)

Thaler (1999) proposed that when people buy a pair of uncomfortable shoes, the more expensive the shoes are, the more times they will be tried on and the longer they will be kept. In our Problem 20 extension, we only found partial support for his prediction. Although our participants were inclined to keep the shoes for a longer time (*t*(506)= 7.53, *p* < .001*,* Cohen’s *d* = 0.33 [0.24, 0.42], but they were not as inclined to keep wearing them (*t*(506)= 1.64, *p* = .51, Cohen’s *d* = 0.07 [-0.01, 0.16]; Stage 2 added comparison: *t*(506) = 4.31, *p* < .001, Cohen’s *dz* = 0.19 [0.10, 0.28])

### Problem 21 (Thaler, 1999)

Thaler (1999) predicted that in subscription services people will be more likely to purchase smaller more often repeating costs over a larger less often costs. In our Problem 21 extension, we indeed found that when presented individually, the framing “merely 27 cents a day” (*M* = 44.53, *SD* = 32.32) was rated as more attractive than “100 US$ a year” (*M* = 26.04, *SD* = 28.05; *t*(331) = 5.57, *p* < .001; Cohen’s *d* = 0.61 [0.39, 0.83]). We found similar results when the two options were presented together in a within-subject design (“merely 27 cents a day”: *M* = 45.16, *SD* = 31.71; “100 US$ a year”: *M* = 35.72, *SD* = 29.05; *t*(169) = 3.82; Cohen’s *dz* = 0.29, [0.14, 0.45]).

##

We summarized the findings, the comparison to the original findings, and our interpretation of our findings in comparison to the original findings in Table 16.

We used the replication evaluation criteria by LeBel et al. (2019). In Stage 1 and our initial submission of Stage 2 we were sometimes unable to deduce the effects, yet in our revision of Stage 2, we were able to use a guide by Jané et al. (2024) to provide effect size for most effects in the original studies, and for all of the replication effects. In the original problems where we did not have enough information, we simply used “signal” versus “no signal” and the direction of the effect in our replication interpretation.

###### Table 16*Comparison of replication and original findings*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

#

| **Problem** | **Condition/Sub Questions** | **ES** | **Original** Effect size [95% CI] | **Replication** Effect size [95% CI] | **Interpretation**  |
| --- | --- | --- | --- | --- | --- |
| 1 | Gain condition | *h* | 0.46 [0.28; 0.63] | 0.75 [0.62; 0.87] |  |
| Loss condition  | *h* | 0.28 [0.11; 0.46]  | 0.12 [-0.00, 0.24] |  |
|  | **Overall: Two conditions compared**  | **Cramer’s V** | **0.43 [0.32, 0.57]** | **0.37 [0.22, 0.54]** | **Successful: signal-consistent** |
| 2  | $15 calculator condition | *h* | 0.37 [0.17; 0.57] | -0.89 [-1.02; -0.77] |  |
| $125 calculator condition | *h* | 0.43 [0.22; 0.64] | 1.04 [0.91; 1.16] |  |
|  | **Overall: Two conditions compared**  | **Cramer’s V** | **0.38 [0.24, 0.54]** | **0.06 [0.00, 0.17]** | ***Unsuccessful:******no signal, inconsistent*** |
| 3 | Lost a $10 bill condition  | *h* | 0.86 [0.72; 1.01] | 0.93 [0.80; 1.05] |  |
| Lost the ticket condition  | *h* | 0.08 [-0.06; 0.22] | -0.57 [-0.69; -0.45] |  |
|  | **Overall: Two conditions compared**  | **Cramer’s V** | **0.44 [0.34, 0.54]** | **0.17 [0.09, 0.26]** | **Successful: signal, inconsistent, smaller** |
| 4 | 4A Segregate gains | Cramer’s V | 0.29 [0.14, 0.45] | 0.26 [0.20, 0.33] | signal-consistent |
| 4B Integrate loss  | Cramer’s V | 0.42 [0.28, 0.58] | 0.23 [0.17, 0.29] | signal-inconsistent, smaller |
| 4C Cancel losses against larger gains | Cramer’s V | 0.41 [0.27, 0.57] | 0.43 [0.37, 0.49] | signal-consistent |
| 4D Segregate small gains from larger losses | Cramer’s V | 0.41 [0.27, 0.57] | 0.31 [0.25, 0.37] | signal-inconsistent, smaller |
|  | **Overall** |  |  |  | **Successful. 4 supported.** |
| 5 | 5A Prefer segregation - Happier two weeks apart  | Cramer’s V | 0.30 [0.13, 0.48] | 0.09 [0.03, 0.16] | *signal-inconsistent, opposite* |
| 5B Prefer integration - More unhappy two weeks apart  | Cramer’s V | 0.29 [0.13, 0.48] | 0.28 [0.21, 0.34] | signal-consistent |
| 5C Prefer integration - More unhappy two weeks apart  | Cramer’s V | 0.42 [0.26, 0.60] | 0.22 [0.17, 0.29] | signal-inconsistent, smaller |
|  | **Overall** |  |  |  | **Mixed. 2 supported, 1 unsupported-opposite.** |
| 6 | First group of questions: (undergrad & MBA1 samples) |  |  |  |  |
|  | 1. (A) Lose $9. (B) Lose $9 after having gained $30. - A hurts more | Cramer’s V | 0.37 [0.22, 0.53]0.51 [0.39, 0.63] | 0.50 [0.41, 0.59] | signal-inconsistent, largersignal-consistent |
| 2. (A) Lose $9. (B) Lose $9 after having lost $30 - B hurts more | Cramer’s V | 0.25 [0.10, 0.41]0.46 [0.34, 0.58] | 0.54 [0.45, 0.62] | signal-inconsistent, largersignal-consistent |
| 3. (A) Lose $9. (B) Lose $9 after having lost $250.- A hurts more | Cramer’s V | \*-0.04 [-0.25,-0.00]\*-0.30 [-0.42, -0.18] | 0.51 [0.42, 0.59] | *signal-inconsistent, opposite* |
| Second group of questions: (MBA1 sample only) |  |  |  |  |
| 4. (A) Lose $9. (B) Lose $9 after a loss of $1000.- A hurts more | Cramer’s V | \*-0.18 [-0.35, -0.01] | 0.47 [0.39, 0.56] | *signal-inconsistent, opposite* |
| 5. (A) Lose $9 after a loss of $30. (B) Lose $9 after a loss of $1000.- A hurts more | Cramer’s V | \*-0.14 [-0.31, -0.00] | 0.41 [0.33, 0.50] | *signal-inconsistent, opposite* |
| Third group of questions: (MBA2 sample) |  |  |  |  |
| 1. (A) Lose $9. (B) Lose $9 after a loss of $9.- B hurts more | Cramer’s V | 0.35 [0.20, 0.52] | 0.51 [0.43, 0.60] | signal-inconsistent, larger |
| 2. (A) Lose $9. (B) Lose $9 after a loss of $18.- B hurts more | Cramer’s V | 0.32 [0.18, 0.49] | 0.60 [0.51, 0.69] | signal-inconsistent, larger |
| 3. (A) Lose $9. (B) Lose $9 after a loss of $36.- B hurts more | Cramer’s V | 0.29 [0.14, 0.46] | 0.59 [0.50, 0.67] | signal-inconsistent, larger |
| 4. (A) Lose $9. (B) Lose after a loss of $45. - B hurts more | Cramer’s V | 0.30 [0.16, 0.47] | 0.59 [0.51, 0.68] | signal-inconsistent, larger |
| 5. (A) Lose $9 after a loss of $9. (B) Lose $9 after a loss of $36.- B hurts more | Cramer’s V | 0.37 [0.22, 0.54] | 0.56 [0.48, 0.66] | signal-inconsistent, larger |
|  | **Overall** |  |  |  | **Mixed. 7 supported, 3 unsupported-opposite.**  |
| 7 | Higher willingness to spend in hotel than in grocery | Cohen’s *d* | N/A |  0.20 [0.03, 0.38] | **Successful:****signal, supported** |
| 8 | Friend < stranger | *η²p* | N/A | 0.55 [0.49, 0.59] | signal, supported |
|  | Interaction:Friend - closer to cost, stranger - closer to market price | *η²p* | N/A | 0.07 [0.03, 0.12] | signal, supported |
|  | **Overall** |  |  |  | **Successful, both supported** |
| 9 | Giving away | Cramer’s V | 0.72 [0.58, 0.87] | 0.80 [0.72, 0.89] | signal-consistent |
|  | Drinking | Cramer’s V | 0.80 [0.64, 0.96] | 0.77 [0.68, 0.85] | signal-consistent |
|  | Giving away versus Drinking | Cramer’s V | 0.05 [0.00, 0.27] | 0.00 [0.00, 0.14] | no signal-consistent |
|  | **Overall** |  |  |  | **Successful, 3 as expected** |
| 10 | Investment $400 > Cost $400Investment $400 > Saved $100 | Cohen’s *d* (paired) | ~0.50 [~0.29, ~0.71] | 0.43 [0.33, 0.52]0.36 [0.27, 0.45] | **Successful:signal, supported** |
| 11 | Prior activity cost - 20 versus 50 (not core) | *η²p* | N/A | 0.01 | signal, consistent |
|  | Dinner (unrelated) versus Sports ticket (related) | Cohen’s *d* (paired) | N/A | 0.26 [0.13, 0.39] | signal, consistent |
|  | Flu vaccination (unrelated) versus Sports ticket (related) | Cohen’s *d* (paired) | N/A | 0.05 [-0.08, 0.17] | *no signal, inconsistent* |
|  | **Overall** |  |  |  | **Mixed. One supported, one unsupported** |
| 12 | $15 ticket versus $40 ticket | Cohen’s *d* | 0.48 [-0.02, 0.97] | 0.23 [0.04, 0.41] | **Successful**:signal, inconsistent, weaker |
| 13 | Won $30: Risk-seeking vs. risk-averse (loss potential)  | Cohen’s *h* | 0.41 | \*-0.46 [-0.54; -0.37] | ***Unsuccessful:*** *signal-inconsistent, opposite*not supported |
| 14  | Lost $30: Risk-seeking vs. risk-averse (loss potential) | Cohen’s *h* | 0.20 | 0.72 [0.63; 0.81] | **Successful**:signal-inconsistent, larger |
| 15 | Lost $30: Risk-seeking vs. risk-averse (no loss potential)  | Cohen’s *h* | 0.20 | \*-0.55 [-0.63; -0.46] | *signal-inconsistent, opposite* |
|  | Problems 13 and 14 compared (exploratory) | Cohen’s *g* | N/A | 0.20 [0.12, 0.26] | signal; more risk averse following loss |
|  | Problems 13 and 15 compared  | Cohen’s *g* | N/A | 0.07 [-0.01, 0.14] | no signal |
|  | Problems 14 and 15 compared (exploratory) | Cohen’s *g* | N/A | 0.12 [0.04, 0.20] | signal, expected direction; (gain only)1. supported |
| 13-15 | **Overall** |  |  |  | **Mixed:** P13/P15 unsuccessful, P14 successful |
| 16 | 1 bet vs. 100 bets | Cohen’s *g* | N/A | 0.36 [0.30, 0.40] | **Successful**: signal, supported |
| 17 | 1 investment vs. 25 investments | Cohen’s *g* | N/A | 0.27 [0.20, 0.33] | **Successful**:signal, supported |
| Extensions |
| 18 | Paid $40 vs. Given by friends | Cohen’s *g* | N/A | 0.47 [0.41, 0.49] | signal, supported |
| 19 | Paid $300, pain, stop preferred over continue playing | Cramer’s V | N/A | 0.26 [0.20, 0.33] | *signal, inconsistent, opposite* |
| 20 | Paid more -> will wear more  | Cohen’s *d* | N/A | 0.07 [-0.01, 0.16 | signal, supported |
|  | Paid more -> will keep longer | Cohen’s *d* | N/A | 0.33 [0.25, 0.42] | *no signal, unsupported* |
|  | Compare: Keeping > wearing | Cohen’s *dz* | N/A | 0.19 [0.10, 0.28] | signal, supported |
| 21 | “merely 27 cents a day” more attractive than “100 US$ a year”Independent-samples | Cohen’s *d* | N/A | 0.61 [0.39, 0.83] | signal, supported |
|  | Dependent-samples | Cohen’s *dz* | N/A | 0.29 [0.14, 0.45] | signal, supported |

\* Original study predicted these would be opposite to the other items. Cramer’s V only includes positive values, yet we converted it in the negative effect when the direction (noted with \*) is the opposite.
1 We note that Problems 13-15 from Thaler (1999) p. 198 were not easy to deduce hypotheses and analyses from. We pre-registered testing each problem and then contrasting 14 and 15 against 13, yet given that Problem 13 was not supported, we realized the need to reframe these contrasts and also run an exploratory contrast between Problems 14 and 15, which in hindsight we should have included a hypothesis for in Stage 1. Overall, we categorized the main demonstration in Problem 13 as failed yet the contrasts to 14 and 15 as aligned with the arguments made, and so as successful. We conclude with our impressions that the demonstrations provided are not empirically aligned with the arguments made, specifically the claim that this is about “chance to break even”, and are more likely the result of the contrast between loss potential and no-loss potential. We reframed this post-hoc in Stage 2 throughout to align with that insight.

#

We carried out a well-powered Registered Report to replicate and extend classic decision-making and behavioral economics problems that were reviewed by Thaler (1999) on or related to the topic of mental accounting. More than twenty six years after the publication of Thaler’s review paper, in our replication we were able to find support for 11 out of the 17 problems reviewed. Specifically, we found consistent results for Problems 1, 3, 4, 7, 8, 9, 10, 12, 14, 16, and 17. Yet the results for Problems 5, 6, and 11 were mixed, and the results of Problems 2, 13, and 15 were inconsistent with the original findings. In the following section we discuss our replication’s findings in comparison to the original studies’, and review the results of the extensions.

##

We categorized our findings regarding Problems 5, 6, and 11 as mixed.

Problems 5 and 6 offered new findings that were not entirely consistent with the original’s findings. In Problem 5, we found no indication for a preference to spread out gains, though if we take a broader view on the comparisons between gains and losses, they seemed to be in the same direction as in the original.

Problem 6 also yielded mixed 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.

In Problem 11, the findings followed the expected pattern of higher likelihood to spend on a theater ticket after spending on an unrelated dinner compared to a related sport event ticket, yet we did not find an indication for such differences when comparing between the related sport event ticket and the unrelated flu vaccination. This suggests the need for a more comprehensive replication of the studies by Heath and Soll (1996) to examine all of their events (e.g., boat tour, party snacks, jeans, watch, etc.) and examine whether the willingness to consume and mentally budget is indeed about relatedness, typicality, or perhaps some other factor.

To summarize, this hedonic editing effect 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. Replications, therefore, serve as an important method to identify and set limits on certain effects.

### Problems with inconsistent results

Results from Problems 2, 13, and 15 were not in alignment 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. In hindsight, this makes sense, and raises an important dilemma in conducting replication studies - whether or not to update and adjust the prices to current days. As one of the participants in the feedback section pointed out, it may cost more than $5 to drive 20 minutes to the other store with the changing costs of driving and the devaluing of money since the problem was first presented in the 1980s. What would have been plausible to some participants in the 1980s, may no longer be plausible to most of current time participants. Therefore, the inconsistent findings may not be due to lack of an effect, but rather because of the changing circumstances. The differences in the findings could be due to the passage of time, and could also be due to our sample’s demographics compared to the original’s, or because of the unified design. The exact reason for the different findings is down to speculation. Future replications could further examine adjustments to the scenarios, or to assess different moderating factors.

We felt that it is important to first conduct a direct replication using unadjusted prices because otherwise the differences may have been attributed to our changes. We at least now have the insight that the phenomenon cannot be observed using the same questions in the current context, and future studies can now have a better estimate of the likelihood of replicating the study without adjustments.

Problems 13 and 15 were questions on risk attitudes, with findings that differed from the original claim, with effects in the 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.

## Extensions

Going beyond the replication of original studies that reported empirical evidence, we also ran four extensions to examine the predictions Thaler made or referred to, that were not accompanied by empirical evidence. Among them, we found empirical support for Problems 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 weather the snowstorm and make an effort to go to the sports game when they paid for the ticket as compared to when they received the ticket as a gift. However, in Problem 19, the large majority of the participants said they will not continue playing tennis after suffering an elbow injury, 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, yet with no indication for the prediction that they would also try to wear them more. Taking into account the replication success of Problem 3, these together revealed that the sunk cost effect might be context-based, such that sunk cost effects may not manifest when there is anticipated physical pain.

We found support for the pennies-a-day effect in Problem 21 (Gourville, 1998). Price frames seemed 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 a review article

This replication project differs from a typical replication registered reports, 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. In addition to attempting a replication of original studies we added extensions testing predictions which did not have any reports of findings. Anecdotal evidence and untested assumptions and predictions are useful, as they provide ideas for future research to build on existing empirical studies. Replications and extensions of a review article can help tackle both aspects, by systematically mapping the studies reported as well as untested claims that can be empirically tested. We hope to see more systematic replications and extensions of impactful review papers, taking a similar approach to ours.

## 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. 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 replications’, 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.

Our participants were exclusively from the US and recruited using an online platform, which is a limitation to generalizability (Simons et al., 2017). Follow-up studies may aim to rerun the same problems using non-US samples to explore the cross-cultural reliability of the mental accounting phenomenon. For instance, a follow-up mass collaboration project conducted by Priolo et al. (2023) was a promising attempt in examining the robustness of mental accounting across cultural contexts. 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. For example, Yue et al. (2020) argued that households altered their risk preference and became more risk-averse due to the pandemic. Thus, the temporal specificity sets another constraint on generalizability.

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 try and discuss implications for mental accounting theory as a whole. Therefore, the results of our replications for each of the problems should be interpreted separately and cautiously. We also did not address the mental accounting tendency at the individual level. Following the suggestion from one of the reviewers, we encourage future research to delve deeper in this regard to undertake broader theoretical integrations and examine individual tendencies. We see much promise in further studies of the links between the different aspects of the mental accounting framework.

#

# Conclusion

We examined the replicability of the mental accounting studies summarized by Thaler (1999). We successfully replicated 11 problems, found mixed support for 3 problems, and failed to find support for 3 problems, suggesting that mental accounting effects generally tend to replicate well but that some effects may be more complex and contextual than originally documented. We believe our reconstruction and reanalysis of classic experiments as well as our exploratory analyses provide an impetus and practical guide 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/bes2.1801c>

Au, N., & Feldman, G. (2020). Revisiting “Goals as Reference Points”: Replication and extensions of Heath et al. (1999). <https://doi.org/10.17605/OSF.IO/WMQTB>

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0, [https://CRAN.R-project.org/package=pwr](https://cran.r-project.org/package%3Dpwr)

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

Feldman, G. (2023). Registered Report Stage 1 manuscript template. <https://doi.org/10.17605/OSF.IO/YQXTP>

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., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive psychology*, *38*(1), 79-109. <https://doi.org/10.1006/cogp.1998.0708>

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](https://doi.org/10.1016/0749-5978%2892%2990006-S)

Hong, C., & Feldman, G. (2025). Revisiting the “Belief in the law of small numbers”: Conceptual replication and extensions Registered Report of problems reviewed in Tversky and Kahneman (1971). Stage 1 In-principle acceptance from Peer Community in Registered Report. <https://doi.org/10.17605/OSF.IO/MNS7J>

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>

Jané, M., Xiao, Q., Yeung, S., Ben-Shachar, M. S., Caldwell, A., Cousineau, D., Dunleavy, D. J., Elsherif, M., Johnson, B., Moreau, D., Riesthuis, P., Röseler, L., Steele, J., Vieira, F., Zloteanu, M., & Feldman, G. (2024). Guide to effect sizes and confidence intervals. <http://dx.doi.org/10.17605/OSF.IO/D8C4G>

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>

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454. [https://doi.org/10.1016/0010-0285(72)90016-3](https://doi.org/10.1016/0010-0285%2872%2990016-3)

Kelley, K. (2023). MBESS: The MBESS R Package. R package version 4.9.3, [https://CRAN.R-project.org/package=MBESS](https://cran.r-project.org/package%3DMBESS) .

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>

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>

Mayiwar, L., Wan, K., Løhre, E., & Feldman, G. (2024). Revisiting representativeness classic paradigms: Replication and extensions of the problems reviewed in Kahneman and Tversky (1972). *Quarterly Journal of Experimental Psychology.* <https://doi.org/10.17605/OSF.IO/NHQC4>

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>

Muehlbacher, S., & Kirchler, E. (2019). Individual differences in mental accounting. *Frontiers in psychology*, *10*, 2866. <https://doi.org/10.3389/fpsyg.2019.02866>

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>

Priolo, G., Stablum, F., Vacondio, M., D'Ambrogio, S., Caserotti, M., Conte, B., … Rubaltelli, E. (2023, July 18). The robustness of mental accounting: A global perspective. <https://doi.org/10.31219/osf.io/apc26>

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [https://www.R-project.org/](https://www.r-project.org/)

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice Bracketing. *Journal of Risk and Uncertainty*, 19(1), 171–197. [https://doi.org/10.1023/A:1007879411489](https://doi.org/10.1023/A%3A1007879411489)

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

Selker, R., Love, J., Dropmann, D., Moreno, V., & Agosti, M. (2022). Package ‘jmv’. Retrieved from CRAN <https://cran.r-project.org/web/packages/jmv/jmv.pdf>

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](https://doi.org/10.1016/0167-2681%2880%2990051-7)

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](https://doi.org/10.1002/%28SICI%291099-0771%28199909%2912%3A3)<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. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105-110. <http://dx.doi.org/10.1037/h0031322>

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

Wong, C., & Feldman, G. (2025). Choice Bracketing revisited: Replication and extensions Registered Report of seven experiments reviewed in Read et al. (1999). *Royal Society Open Science*. <https://doi.org/10.17605/OSF.IO/VDQEK>

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>

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. <https://doi.org/10.1017/S0140525X17001972>