**Revisiting mental accounting classic paradigms:**

**Replication of the problems reviewed in Thaler (1999)**

Mengfei Li

ORCID: 0000-0001-6460-555X

The University of Hong Kong, Hong Kong SAR

[lmf0312@hku.hk](mailto:lmf0312@connect.hku.hk) / [lmf0312@connect.hku.hk](mailto: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

\*Contributed equally, joint first author

^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:

The author(s) received no financial support for the research and/or authorship of this article.

## Authorship declaration:

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

Gilad Feldman guided and supervised each step in the project, (later: 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](mailto:gfeldman@hku.hk) ; 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))

​

## Contributor Roles Taxonomy

In the table below, employ CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to the url (<https://www.casrai.org/credit.html> ) on details and definitions of each of the roles listed below.

|  |  |  |
| --- | --- | --- |
| **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

**[IMPORTANT:   
Method and results sections were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. This is written in past tense yet no pre-registration or data collection have been conducted.]**

Mental accounting, the internal categorization system individuals adopt to manage their financial activities, makes people prone to irrational decision-making. In a pre-registered study with an American online Amazon Mechanical Turk sample (*N* = 1000), we replicated 17 mental accounting problems reviewed in Thaler (1999). Out of the 17 problems, we found empirical support for X with effect sizes ranging from X.XX [X.XX, X.XX] to X.XX [X.XX, X.XX], and no empirical support for Y with effect sizes ranging from X.XX [X.XX, X.XX] to X.XX [X.XX, X.XX]. Extending the replication, we provided an initial test of four predictions not previously empirically tested that were described in Thaler’s (1999) paper as predictions. The replications and extensions examined different parts of the mental accounting theory and the results were interpreted separately. Materials, dataset, and analysis code were made available on the OSF: <https://osf.io/pwa68/>.

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

# PCIRR-Study Design Table

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

*Note*. Requested as part of the PCI-RR submissions

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

## Background

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

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

We report a very close replication of the work reviewed in Thaler (1999), with the following goals (closeness evaluation based on the LeBel et al., 2018 criteria). Our first goal was to conduct close replications of the classic mental accounting problems reviewed by Thaler (1999) by an external independent lab (KNAW, 2018; Peels, 2019). Our second goal was to examine several predictions made by Thaler regarding mental accounting behaviors that the review did not cover empirical tests for.

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

## Mental accounting

Mental accounting has long been a heated topic in the field of behavioral economics, psychology, and judgment and decision making. The earliest empirical evidence on mental accounting behaviors dates back to Tversky and Kahneman’s (1981) famous theater-ticket experiment (one of our replication problems). Tversky and Kahneman proposed that mental accounting is a form of decision framing by which people formulate (psychological) accounts to evaluate events and options (as cited in Henderson & Peterson, 1992). Their findings suggested that in people’s minds, losses tend to be labeled into different categories, violating the long-standing economic notion of fungibility (Thaler, 1999). Ever since, the concept of mental accounting has been used to understand a wide range of decision-making behaviors, such as gambling, risk taking, and investment (Thaler & Johnson, 1990). Although these subsequent mental accounting studies differ in specific objectives and orientations, nearly all research has touched upon gains and losses, and indicated the violation of fundamental economic norms (Henderson & Peterson, 1992). In Thaler (1985) and our target article-Thaler (1999), the mental accounting phenomenon was further elaborated and summarized into a broad theory on consumer choices (Muehlbacher & Kirchler, 2019). In 2011, Soman and Ahn reviewed substantial mental accounting research focusing on the framing effects. More recently, Zhang and Sussman (2018)’s review paper again outlined the categorization process of mental accounting, and they summarized it as a way for people to “group expenses into categories, assign funds to these categories, determine budgets, and perform elements of cost–benefit analyses.” (p.65). The two recent review papers cited very similar research to Thaler (1999), such as Heath and Soll (1996), Tversky and Kahneman (1981), Thaler (1980), and Thaler and Johnson (1990). This further exemplifies the necessity in revisiting these classic findings and testing the reproducibility, robustness, and generalizability of these influential and pioneering works, to substantiate and strengthen the empirical foundations of the theoretical framework of mental accounting. We aimed to examine the evidence reviewed in Thaler (1999) targeting different subsets of the mental accounting framework.

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

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

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

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

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

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

## Original hypotheses and findings in target article

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

Table 1

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

| **Problem** | **Google scholar citation** | **Description and Explanation** | **Hypothesis** |
| --- | --- | --- | --- |
| 1 | Based on **Tversky and Kahneman, 1986**  (608) | **(Risk Taking)**  Manipulation with two conditions testing diminishing sensitivity towards gain and loss.  People perceive outcomes based on the value function in the prospect theory. | **H1:** People are risk-averse for gains and risk-seeking for losses. |
| 2 | 23905) | **(Time investment versus price reference point)**  Manipulation with two conditions testing reference points: Spending 20 minutes to save $5 out of $15 versus $5 out of $125 | **H2:** Reference points shift evaluations of value.  People are more likely to spend 20 minutes to save $5 out of $15 than to save $5 out of $125 |
| 3 | (23905) | **(Theater Play Ticket)**  Manipulation with two conditions testing the impact of an existing account on decisions.  The sunk cost effect arises when the decision is referred to an existing account set up by a related act. | **H3:** Not explicitly reported |
| 4 | (8237) | **(Events and Happiness)**  Four pairs of scenarios testing the hedonic framing.  People tend to frame outcomes in ways that make them the happiest. | **H4:** People follow four principles (a) segregate gains, (b) integrate loss, (c) cancel losses against larger gains, and (d) segregate “silver linings”. |
| 5 | (3194) | **(Same Day or Two Weeks Apart)**  Three pairs of events testing the temporal spacing of hedonic editing.  People tend to simplify and encode multiple outcomes in a hedonically optimal manner. | **H5:** The hedonic editing hypothesis argues that subjects choose to have the events occur “apart” when segregation is preferred, and “together” when integration is hedonically optimal. |
| 6 | (3194) | **(Emotional Impact of Losing $9)**  Manipulation with two conditions testing the effect of a prior loss.  People do not integrate subsequent losses with the prior loss when faced with a two-stage gamble. | **H6:** Not explicitly reported |
| 7 | **Thaler, 1985**  (8237) | **(Location and Price)**  Manipulation with two conditions testing transaction utility.  Consumption decisions are influenced by people’s perceived value of the “deal”. | **H7:** Transaction utility influences willingness to pay. |
| 8 | (8237) | **(Selling Ticket)**  Manipulation with three conditions testing the determinants of the reference point.  Fairness is the dominant factor in determining reference price. | **H8:** Not explicitly reported |
|  | (5) | **(Wine Bottle)**  Manipulation with two conditions testing the value of wine.  People hold mixed perceptions of the value of items when the consumption and purchase is temporally separated. | **H9:** Not explicitly reported |
| 10 | **Shafir and Thaler, 1998**  (5) | **(Feelings about Purchase)**  Three statements examining “investment” purchases.  People can avoid the feeling of spending when the purchase is perceived as investment. | **H10:** Not explicitly reported |
| 11 | **Heath and Soll, 1996**  (1035) | **(Previous Events and New Payment)**  Manipulation with two conditions testing the underconsumption of a typical target.  People set budgets for different accounts and recomputed the remaining budgets periodically. They will resist further expenses if a particular budget is depleted. | **H11a:** The budget-setting process promotes greater underconsumption in the $50 condition than the $20 condition.  **H11b:**The expense-tracking process promotes greater underconsumption for more typical purchases. |
| 12 | (706) | **(The Performance)**  Manipulation with two conditions testing whether the value of time is influenced by price-related characteristics of the decision situation.  The value of time is influenced by contextual effects. | **H12:** Subjects are willing to pay more money to avoid waiting the same amount of time for a higher-priced food or service than for a lower-priced product. |
| 13-15 | (4637) | **(Choices)**  A gain (Q13) vs. loss (Q14-15) scenario examines prior outcomes and risky choices.  When gambles are bracketed together, the outcome of the prior gamble can influence subsequent choices. | **H13:** “House Money” - Prior gain stimulates risk seeking  **H14-15**: Weaker to no effects for prior loss. |
| 16 | **Samuelson, 1963**  (983) | **(Coin Flip Bet)**  A scenario testing how bracketing the gambles affects the attractiveness of individual bets. | **H16**: People shift between single gambles and long-term repeating gambles. |
| 17 | (4637) | Two scenarios examining the myopic loss aversion effect.  Narrow framing inhibits risk-taking. | **H17:** Not explicitly reported |

*Note.* The Google scholar citations were of April 2022.

For Shafir and Thaler, 1998, it was later published titled “Invest Now, Drink Later, Spend Never: The Mental Accounting of Delayed Consumption” and there were 229 Google Scholar citations.

We provided a summary of sample characteristics and the findings of the original studies in Table 2.

Table 2

*Summary of sample characteristics and findings in the original studies*

| **Factors** | **Sample characteristics** | **Statistical results reported** |
| --- | --- | --- |
| Problem 1 | Gain condition: *N*=126  Loss condition: *N*=128 | Gain condition:  [72%]-Sure gain  [28%]-Uncertain gain  Loss condition:  [36%]-Sure loss  [64%]-Uncertain loss |
| Problem 2 | $15 Calculator Condition (jacket for $125, calculator for $15): *N*=93  $15 Calculator Condition (jacket for $15, calculator for $125): *N*=88 | $15 Calculator Condition:  [68%]-Yes  $125 Calculator Condition:  [29%]-Yes |
| Problem 3 | Lost a $10 bill condition: *N*=183  Lost the ticket condition: *N*=200 | Lost a $10 bill condition:  [88%]-Yes [12%]-No  Lost the ticket condition  [46%]-Yes [54%]-No |
| Problem 4 | 87 students in an undergraduate statistical class at Cornell University | World Series lottery:  [56%]-A [16%]-B [15%]-No difference  Letter from IRS:  [66%]-A [14%]-B [7%]-No difference  New York State lottery:  [22%]-A [61%]-B [4%]-No difference  Car damage  [19%]-A [63%]-B [5%]-No difference |
| Problem 5 | For all three events: *N*=65 | Office lottery:  [25%]-A [63%]-B [12%]-No difference  Arithmetical mistake:  [57%]-A [34%]-B [9%]-No difference  Parking ticket:  [75%]-A [17%]-B [7%]-No difference |
| Problem | Two groups of Cornell MBA students  First group: *N*=87  Second group: *N*= 81 | First group of Problems  1. [70%]-A [9%]-B [21%]-No difference  2. [13%]-A [55%]-B [31%]-No difference  3. [39%]-A [38%]-B [23%]-No difference  4. [50%]-A [33%]-B [17%]-No difference  5. [51%]-A [38%]-B [21%]-No difference  Second group of Problems  6. [7%]-A [64%]-B [28%]-No difference  7. [11%]-A [65%]-B [23%]-No difference  8. [12%]-A [62%]-B [26%]-No difference  9. [14%]-A [65%]-B [21%]-No difference  10. [7%]-A [68%]-B [25%]-No difference |
| Problem 7 | Participants are in an executive development program, and are regular beer drinkers | Fancy resort hotel condition:  *Median*=$2.65  Small run-down grocery store condition:  *Median*=$1.50 |
| Problem 8 | First-year MBA students  Free condition: *N*=31  Paid $5 condition: *N*=28  Paid $10 condition: *N*=26 | Free condition:  *When market value is $5:*  68% answer 0, 26% answer 5, 3% answer 10, and 3% answer Other if selling to a friend.  6% answer 0, 77% answer 5, 10% answer 10, and 6% answer Other if selling to a stranger.  *When market value is $10:*  65% answer 0, 26% answer 5, 6% answer 10, and 3% answer Other if selling to a friend.  6% answer 0, 16% answer 5, 58% answer 10, and 19% answer Other if selling to a stranger.  Paid $5 condition:  *When market value is $5:*  14% answer 0, 79% answer 5, 0% answer 10, and 7% answer Other if selling to a friend.  0% answer 0, 79% answer 5, 7% answer 10, and 14% answer Other if selling to a stranger.  *When market value is $10:*  7% answer 0, 79% answer 5, 4% answer 10, and 9% answer Other if selling to a friend.  0% answer 0, 14% answer 5, 57% answer 10, and 29% answer Other if selling to a stranger.  Paid $10 condition:  *When market value is $5:*  0% answer 0, 69% answer 5, 23% answer 10, and 8% answer Other if selling to a friend.  0% answer 0, 42% answer 5, 46% answer 10, and 12% answer Other if selling to a stranger.  *When market value is $10:*  0% answer 0, 15% answer 5, 69% answer 10, and 15% answer Other if selling to a friend.  0% answer 0, 0% answer 5, 73% answer 10, and 27% answer Other if selling to a stranger. |
| Problem 9 | Participants are subscribers to a wine newsletter, *Liquid Assets*. They are highly knowledgeable wine consumers with substantial home cellars.  Giving away condition: *N*=97  Drinking condition: *N*=76 | Giving away condition:  [30%]-(a) [17%]-(b) [9%]-(c)  [30%]-(d) [14%]-(e)  Drinking condition:  [30%]-(a) [18%]-(b) [7%]-(c)  [20%]-(d) [25%]-(e) |
| Problem 10 | Participants are subscribers to a wine newsletter, *Liquid Assets*. | Statement (a): *Mean*=3.31  Statement (b): *Mean*= 1.94  Statement (c): *Mean*=2.88 |
| Problem | 66 MBA students, split evenly between conditions. Participants receive pizza and beer in exchange. | Heath and Soll (1996) revealed that a larger proportion of people are more likely to underconsume in the high-cost than in the low-cost condition (*t*(26)=2.17, *p*<.05 by paired *t*-test). The proportion of subjects who underconsume the target is highly correlated with typicality for both high-cost (*r*(25)=.80, *p*<.01) and low-cost conditions (*r*(25)=.67, *p*<.01). |
| Problem 12 | MBA students  N=67, 37 male and 30 female | 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). |
| Problem 13 | MBA students who played for real money | [70%]-A 50% chance to gain $9 and a 50% chance to lose $9  [30%]-No further gain or loss |
| Problem 14 | MBA students who played for real money | [40%]-A 50% chance to gain $9 and a 50% chance to lose $9  [60%]-No further gain or loss |
| Problem 15 | MBA students who played for real money | [60%]-A 33% chance to gain $30 and a 67% chance to gain nothing  [40%]-A sure $10 |
| Problem 16 | An economist colleague | No for A single coin flip  Yes for playing the bet 100 times |
| Problem 17 | The CEO and a group of 25 executives from one firm, each of whom was responsible for managing a separate division | 3 of the 25 executives accepted the single investment, and the CEO accepted the portfolio of 25 of these investments. |

*Note*. a The statistical results reported in Problem 6-Condition A-5 added up to 110% rather than 100%, suggesting a possible reporting mistake in the original article.  
b Also, we are unsure about the paired t-test reported in Problem 11 as the experiment seems to adopt a between-subject design.

## Extensions - Prediction extensions

We extended the replication of the problems reviewed by also adding a test of four predictions that Thaler (1999) reflected on but did not review empirical evidence that directly tested these predictions. We summarized our extensions in Table 3.



## Overview of replication and extension

Thaler’s (1999) review paper covered a long list of classic mental accounting problems, and we focused on 17 of those. For each of the replication problems, we followed the original experimental design with minor adjustments needed to update those to current times make those suitable for our target sample. We summarized the minor changes we made in Table 8. We then added four additional experiments to examine predictions that Thaler made that were not reviewed with supporting empirical evidence.

## Pre-registration and open-science

*[Note: Written in past tense as a simulation for the final manuscript, but no pre-registration has taken part yet, and will only be completed after RR Stage 1 IPA.]*

We first pre-registered the study on the Open Science Framework (OSF) and data collection was launched later. Pre-registrations, power analyses, and all materials used in these problems are available in the supplementary materials. We pre-registered and provided all materials, data, code for all studies on OSF: <https://osf.io/v7fbj/> .   
 We also provided further open-science details and disclosures in the supplementary materials section “Open-science Disclosures”. All measures, manipulations, exclusions conducted for this investigation are reported, all studies were pre-registered with power analyses reported in the supplementary materials section “Power analysis of original study effect to assess required sample for replication”, and data collection was completed before analyses.

Table 3

*Extension: Summary of predictions made by Thaler (1999) that had no reviewed supporting evidence*

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

*Note*. The papers listed are the sources of the predictions and none of the predictions have been tested directly to the best of our knowledge.

a For Problem 20, we aimed to examine how much participants identify with Thaler’s prediction.  
b The penny-a-day effect in Problem 21 has been investigated in follow-up research in marketing (e.g. Gourville, 1998).

# Method

**[IMPORTANT:   
Method and results sections were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. This is written in past tense yet no pre-registration or data collection have been conducted.]**

## Power analysis

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

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

## Participants

For the Stage 1 PCI Registered Report, we stimulated a dataset of 200 participants using Qualtrics (*Mage* = 50.07, *SD* = 29.85; 52 females, 48 males, 40 others, and 60 rather not disclose). This allowed us to clearly demonstrate what we intended to include in the Results section after data analysis. We summarized sample differences between the current replication and the original studies in Table 4 .

In the actual data collection, we will recruit native English speakers who were born, raised, and located in the US on Amazon Mechanical Turk using the CloudResearch/Turkprime platform (Litman et al., 2017). Based on our extensive experience of running similar judgment and decision-making replications on MTurk, to ensure high-quality data collection, we will employ 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 will also employ the [Qualtrics fraud and spam prevention measures](https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/): reCAPTCHA, prevent multiple submission, prevent ballotstuffing, bot detection, security scan monitor, relevantID, etc.

Assignment pay is based on the federal wage of 7.25USD/hour, per minute, so for example - 5-8 minutes survey would be paid 1USD per participant. We first pretested survey duration with 30 participants to make sure our time run estimate was accurate and adjusted pay as needed, the data of the 30 participants was not analyzed other than to assess survey completion duration and needed pay adjustments. [For those pretest participants, if survey duration was longer than expected, they would be paid a bonus as pay adjustment. All of these will be reported after data collection.]

Table 4

*Summary of samples in the original studies and our replication*

|  | **Sample size** | **Characteristics** | **Medium (location) and Compensation** |
| --- | --- | --- | --- |
| The current replication | 200 (52 females, 48 males, 40 others, and 60 rather not disclose) | US American (Median age=53.00 years, Average age=50.07 years, Standard deviation age=29.85 years, age range=0-100 years) | Computer (online) |
| Problem 1 | 254 | 126 for Gain condition  and 128 for Loss condition | Unreported |
| Problem 2 | 181 | 93 for $15 Calculator Condition and 88 for $125 Calculator Condition | Unreported |
| Problem 3 | 383 | 183 for Lost a $10 bill condition and 200 for Lost the ticket condition | Unreported |
| Problem 4 | 87 | Undergraduate students in a statistical class at Cornell University | (In person) |
| Problem 5 | 195 | 65 for each condition | Unreported |
| Problem 6 | 168 | Cornell MBA students,  87 for Condition A and 81 for Condition B | Unreported |
| Problem 7 | Unreported | Regular beer drinkers in an executive development program | (In person) |
| Problem 8 | 85 | First-year MBA students, 31 for Free condition, 28 for Paid $5 condition, 26 for Paid $10 condition | Unreported |
| Problem 9 | 173 | Subscribers to a wine newsletter 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 | Unreported |
| Problem 11 | 66 | MBA students, split evenly across conditions | (In person),  Pizza and beer |
| Problem 12 | 67 (37 male and 30 female) | MBA students | Unreported |
| Problem 13 | Unreported | MBA students | (In person) |
| Problem 14 | Unreported | MBA students | (In person) |
| Problem 15 | Unreported | MBA students | (In person) |
| Problem 16 | 1 | An economist | (In person) |
| Problem 17 | 26 | A CEO and 25 executives | (In person) |

## Design and procedure

We summarized the experimental designs in Table 5. We mapped the designs used in the problems, which included one-sample, between-subject, and within-subject experimental designs. We set up all the Problems using Qualtrics. Adopting a formatting method that can best reduce participants' cognitive load, we had a total of 18 Qualtrics blocks. Participants were randomly assigned to complete 9 of the 18 blocks, in order to address reviewer’s feedback to decrease the length of the survey and the burden on participants. The display of Problems and conditions was counterbalanced using the randomizer “evenly present” function in Qualtrics. Problems were presented in random order and participants were randomly and evenly assigned into different conditions. We previously tested this method in many other replications and extensions conducted by our team, for example, a similar replication of an influential review paper by Kahneman and Tversky (1972) (Wan & Feldman, 2021). Our findings from projects using a similar design suggest that combining several experiments in a single data collection in random order did not impact likelihood of replication success, and allowed for important additional insights.

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

[*For review: The Qualtrics survey .QSF file and an exported DOCX file are provided on the OSF folder. A preview link of the Qualtrics survey is provided on -* [*https://hku.au1.qualtrics.com/jfe/preview/SV\_3POMRveqS86noY6?Q\_CHL=preview&Q\_SurveyVersionID=current*](https://hku.au1.qualtrics.com/jfe/preview/SV_3POMRveqS86noY6?Q_CHL=preview&Q_SurveyVersionID=current)]

Table 5

*Replication and extension experimental design*

|  |  |  |  |
| --- | --- | --- | --- |
| **Problem 1:**  Between | **IV: Gain condition**  Choices between sure/uncertain *gain* | | **IV: Loss condition**  Choices between sure/uncertain *loss* |
| DV: Risk taking preference (choice) | | |
| **Problem 2:**  Between | **IV: $15 Calculator Condition**  Jacket is $125; Calculator is $15 | | **IV: $125 Calculator Condition**  Jacket is $15; Calculator is $125 |
| DV:Willingness to travel to another store (choice) | | |
| **Problem 3:**  Between | **IV: “Lost a bill” Condition**  Lost a $10 bill as you enter the theater | | **IV: “Lost the ticket” Condition**  Lost the $10 ticket as you enter the theater |
| DV: Willingness to buy (another) ticket (choice) | | |
| **Problem 4:**  Multiple experiments, one sample proportions | IV: Hedonic framing | | |
| DV: Whether perceived as emotionally equivalent (choice)  Specific DV items: After reading the scenario, participants choose who was happier/more upset. | | |
| **Problem 5:**  Within | IV: Temporal spacing | | |
| DV: Whether perceived as emotionally equivalent (choice)  Specific DV items: After reading the scenario, participants are to choose who was happier/more unhappy. | | |
| **Problem 6:**  Multiple experiments, one sample proportions | IV: Incremental impact of loss  Manipulation: Different prior outcomes | | |
| DV: Emotional Impact of Losing $9 (choice)  Specific DV items: After reading the scenario, participants are to choose which event hurts more. | | |
| **Problem 7:**  Between | **IV: Hotel condition**  The soda is sold at a fancy resort hotel. | | **IV: Grocery store condition**  The soda is sold at a small, run-down grocery store. |
| DV: Price willing to pay (continuous) | | |
| **Problem 8:**  Mixed:  Between-subject design (Free vs. $5 vs. $10) Within: Friend vs. Stranger. Within: Market price $5 vs. $10 | **IV: Free ticket condition**  The tickets were given for free by a friend. | **IV: Paid $5 condition**  The tickets were bought at $5. | **IV: Paid $10 condition**  The tickets were bought at $10. |
| DV: Price willing to sell (continuous)  Specific DV items: Participants are to indicate their selling price when the customer is a friend/stranger when the going price is $5/$10. | | |
| **Problem 9:**  Between | **IV: Drinking condition**  Participants are to imagine drinking a bottle of the wine with dinner | | **IV: Giving away condition**  Participants are to imagine giving one bottle of the wine to a friend as a gift |
| DV: Feeling of the cost  Specific DV items: Participants are to choose which statement best captures their feeling of the cost. | | |
| **Problem 10:**  Within | IV: Purchase of Bordeaux futures at $400 | | |
| DV: Feeling about purchase (ordinal)  Specific DV items: Participants are to choose which statement best captures their feeling at the purchase. | | |
| **Problem :**  Mixed:  Between: $50 vs. $20  Within: Dinner vs. Ticket vs. Flu  Within: Spent/given | **IV: $50 condition**  Spent $50 on the previous event | | **IV: $20 condition**  Spent $20 on the previous event |
| DV: Willingness to buy a $25 theater ticket later in the week (choice) | | |
| **Problem :**  Between | **IV: $15 condition**  The tickets will cost $15 each | | **IV: $40 condition**  The tickets will cost $40 each |
| DV: Price willing to pay to avoid waiting (continuous) | | |
| **Problem 13:**  Within | IV: Won $30 scenario | | |
| DV: Risk taking preference  Specific DV items: Imagine winning $30, participants are to choose between uncertain gain/loss or no further gain/loss. | | |
| **Problem 14:**  Within | IV: Lost $30 scenario A | | |
| DV: Risk taking preference  Specific DV items: Imagine losing $30, participants are to choose between uncertain gain/loss or no further gain/loss. | | |
| **Problem 15:**  Within | IV: Lost $30 scenario B | | |
| DV: Risk taking preference  Specific DV items: Imagine losing $30, participants are to choose between uncertain gain/loss or a sure gain. | | |
| **Problem 16:**  Within | IV: Coin Flip Bet | | |
| DV: Willingness to take the bet  Specific DV items: Decision under a single coin flip/100 coin flips | | |
| **Problem 17:**  Within | IV: Division Investment | | |
| DV: Willingness to undertake the investment  Specific DV items: Decision under a single project/ a portfolio of 25 projects | | |
| **Problem 18:**  Within (extension) | IV: The cost of the ticket | | |
| DV: Willingness to go to the game  Specific DV items: Decision between go to the game and stay home when the ticket is bought/given | | |
| **Problem 19:**  Within  (extension) | IV: Membership at tennis club | | |
| DV: Willingness to play  Specific DV items: Decision between stop playing and continue playing | | |
| **Problem 20:**  Within  (extension) | IV: Shoe purchase scenario | | |
| DV: Sunk cost effects  Specific DV items: Participants are to indicate how accurately the statements apply to them. | | |
| **Problem 21:**  Between  (extension) | **IV: Day expression condition** | **IV: Year expression condition** | **IV: Both expressions condition** |
| DV: Attractiveness of the membership plan | | |

*Note.* Please see Table 11 for the options of each Problem.

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

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

## Manipulations

We provided full details of the manipulations between the conditions and the experimental designs in the supplementary materials section “Materials and scales used in the replication + extension problems”. Problems 1, 2, 3, 6, 7, 8, 9, 11, 12, and 21, involved between-subjects manipulations, and participants were randomly assigned to conditions separately in each of those*.*The order of the problems was also randomized. Please see Table 5 for a summary of all problems and manipulations.

## Measures

### Replications: 17 problems testing mental accounting paradigms

We summarized the measures and data analysis strategery for all replication problems in Table 6.

Table 6

*Measures and data analysis strategery for replication problems*

|  |  | **Data analysis strategery** | |
| --- | --- | --- | --- |
| **Problem** | **Measure** | **In the original** | **Deduced additional analysis** |
| 1 | Choose between two choices (displayed in random order) | Calculated the cumulative percent frequency for each choice | Chi-square |
| 2 | Answer the Yes/No question (options displayed in random order) | Calculated the cumulative percent frequency for the Yes option | Chi-square |
| 3 | Answer the Yes/No question (options displayed in random order) | Cumulative percent frequency for each choice | Chi-square |
| 4 | Four pairs of scenarios are presented in random order. Choose among three choices. | Calculated the cumulative percent frequency for all three choices. | Proportion tests |
| 5 | Three pairs of scenarios are presented in random order. Choose among three choices. | Calculated the cumulative percent frequency for all three choices. | Proportion tests  McNemar paired-samples tests: A-B A-C |
| 6 | Five pairs of events displayed in random order. Choose among three choices. | Calculated the cumulative percent frequency for all three choices. | Proportion tests |
| 7 | Report what price they will tell the friend. | Calculated the median for the two conditions | Independent samples t-test |
|  | Report what price they will ask under different condition | Calculated percent of subjects giving common answers (0, 5, 10, Other). | Mixed ANOVA:  3 between: free vs. paid 5 vs. paid 10  2 within: friend vs. stranger  2 within:  market worth 5 vs. 10 |
| 9 | Choose among five statements (displayed in random order). | Calculated the cumulative percent frequency for all five statements. | Chi-square |
| 10 | Indicate which statement more accurately captures their feelings on a 5-point Likert scale (four statements displayed in random order). | Calculated the mean score of all four 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 vs. 50  2 within: given vs. spent  2 within:  dinner vs. ticket |
| 12 | Report how much they would be willing to pay to avoid waiting. | Calculated the mean score for each condition and conducted independent sample t-tests. | Subtract the price of the ticket 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*: For Problem 8, there was no explanation provided regarding the classification of “common answers”, so we can only assume that any value other than 0, 5, and 10 were counted as “Other”.

### Extensions: Testing predictions by Thaler with no reviewed supporting evidence

We added four problems that were not originally tested in the review article, and we summarized the measures and data analysis strategy for these extensions in Table 7.

Table 7

*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. | Paired samples t-test |
| 21 | Rate the attractiveness of the membership plan on a 0-100 scale. | Independent samples t-test  Paired sample t-test |

## Deviations from the original studies

Our replication deviations from the original’s studies include participants’ characteristics, delivery mode, and the extensions. We summarized sample deviations in Table 4 and technical deviations in Table 8.

Table 8

*Deviations for specific problems*

| **Problem number** | **Deviation details** | **Reason for change** |
| --- | --- | --- |
| Replication Problem 1 | We adopted the wording Thaler used in his work. For the Gain condition, the original second option was framed as “*A 50 % chance to gain $200 and a 50% chance to lose $0*”. We changed it to “*A 50 % chance to gain $200 and a 50% chance to gain $0*” in our current replication. | We tried to be as close as possible to Thaler. While comparing the option with the loss condition, we suspected it as a typo. |
| Replication Problem 4 | The description of the problem is slightly simplified and we randomized the sequence of the scenarios. | Minor wording changes for clarity.  Randomization to eliminate order bias. |
| Replication Problem 5, 6 | The description of the problem and the options are revised and simplified. We also randomized the sequence of the scenarios. | Minor wording changes for clarity.  Randomization to eliminate order bias. |
| Replication Problem 7 | The original study used beer in the two conditions and we changed it into soda. | Some of our targeted participants may not drink. |
| Replication Problem 9 | Added “Imagine that you enjoy drinking wine” at the beginning of the scenario. We also randomized the sequence of the statements. | Our targeted population would mostly come from the working/lower-middle class and might not enjoy drinking wine.  Randomization to eliminate order bias. |
| Replication Problem 10 | 1. Added “Imagine that you enjoy drinking wine” at the beginning of the scenario.  2. Added another option “*I cannot understand this question*”  3. Randomized the sequence of the statements. | 1. Our targeted population would mostly come from the working/lower-middle class and might not enjoy drinking wine.  2. Our pretest showed that this scenario might be too vague and difficult to comprehend for our targeted participants so we added another statement to check for understanding.By adding this option, we ensure that participants will not just choose a random option when they cannot understand the question.  3. Randomization to eliminate order bias. |
| Replication Problem 11 | The instruction of the problem is simplified. | Minor wording changes for clarity and understandability. |
| Replication Problem 12 | The original study used student tickets at the student window in the scenarios. We changed it into discounted tickets and discount windows.  The question is revised. | Our targeted population would have a wide age range from 18 to 80 so many of them might not be students.  Our pretest showed that the original framing of the question caused misunderstandings. We revised for greater clarity. |
| Replication Problem 13, 14, 15 | Added “Imagine that…”at the beginning of the scenarios. | To facilitate perspective-taking |
| Extension Problem 21 | Thaler (1999) used “local public radio station” in his prediction while we changed it into “music online streaming service”. | The original scenario does not apply to 2022 so we change it to update to current time. |

## Evaluation criteria for replication findings

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

[*Note: Evaluation is not currently possible in Stage 1 since the current sample used for data analysis was randomly generated on Qualtrics. This part will be further revised after the actual data collection.*]

## Replication closeness evaluation

We provided details on the classification of the replications using the criteria by LeBel et al. (2018) in Table 9 below (also see Section *“*Replication Evaluation” in the supplementary). We summarized the replication as a "very close" replication.

Table 9

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

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

## Data analyses plans: Outliers and exclusions

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

# Results

**[IMPORTANT:   
Method and results sections were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. This is written in past tense yet no pre-registration or data collection have been conducted.]**

## Replications and extensions

Descriptives for all the problems are provided in Tables 10 and 11. Statistical tests for all problems are summarized in Tables 12-16.

[Note: *Please refer to the JAMOVI PDF output file Thaler1999repext-JAMOVIoutput.pdf on the OSF for plots of the problems, these will be updated in the manuscript after actual data collection.*]

Table 10

*Descriptive statistics for problems with options*

| **Number** | **Condition**  **(if applicable)** | **N** | **Options** | **Count** |
| --- | --- | --- | --- | --- |
| Problem 1 | Gain condition | 100 | A sure gain of $100 | 56 |
| A 50 % chance to gain $200 and a 50% chance to gain $0 | 44 |
| Loss condition | 100 | A sure loss of $100 | 43 |
| A 50% chance to lose $200 and a 50% chance to lose $0 | 57 |
| Problem 2 | $15 Calculator condition | 100 | Make the trip to the other store | 41 |
| Not making the trip to the other store | 59 |
| $125 Calculator condition | 100 | Make the trip to the other store | 62 |
| Not making the trip to the other store | 38 |
| Problem 3 | Lost a $10 bil condition | 100 | Buy another ticket | 41 |
| Not buying another ticket | 59 |
| Lost the ticket bill condition | 100 | Buy another ticket | 46 |
| Not buying another ticket | 54 |
| Problem 4 | Two wins: $50, and $25  versus  One win: $75.  Who was happier? | 200 | Two wins is happier | 63 |
| One win is happier | 75 |
| No difference | 62 |
| Two mistakes: $100, and $50  versus  One mistake: $150  Who was more upset? | 200 | Two mistakes is more upset | 63 |
| One mistake is more upset | 64 |
| No difference | 73 |
| Two events: Win $100, and loss $80  versus  One event: Win $20  Who was happier? | 200 | Two events is happier | 67 |
| One event is happier | 64 |
| No difference | 69 |
| Two events: Loss $200, and win $25  versus  One event: Loss $175  Who was more upset? | 200 | Two events is more upset | 82 |
| One event is more upset | 63 |
| No difference | 55 |
| Problem 5 | Two events:  (1) win $25  (2) win $50  Who is happier? | 200 | Happier when occur on the same day | 79 |
| Happier when two events are two weeks apart | 65 |
| No difference | 56 |
| Two events:  (1) $100 must be paid  (2) $50 must be paid.  Who is more unhappy? | 200 | More unhappy when occur on the same day | 73 |
| More unhappy when two events are two weeks apart | 56 |
| No difference | 71 |
| Two events:  (1) a $20 parking ticket  (2) a $25 bill  Who is more unhappy? | 200 | More unhappy occur on the same day | 67 |
| More unhappy when two events are two weeks apart | 69 |
| No difference | 64 |
| Problem 6 | (A) You lose $9. (B) You lose $9 after having gained $30. | 100 | A hurts more than B | 40 |
| B hurts more than A | 30 |
| No difference | 30 |
| (A) You lose $9. (B) You lose $9 after having lost $30. | 100 | A hurts more than B | 29 |
| B hurts more than A | 39 |
| No difference | 32 |
| (A) You lose $9. (B) You lose $9 after having lost $250. | 100 | A hurts more than B | 33 |
| B hurts more than A | 34 |
| No difference | 33 |
| (A) You lose $9. (B) You lose $9 after suffering a loss of $1000. | 100 | A hurts more than B | 28 |
| B hurts more than A | 39 |
| No difference | 33 |
| (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000. | 100 | A hurts more than B | 31 |
| B hurts more than A | 36 |
| No difference | 33 |
| (A) You lose $9. (B) You lose $9 after suffering a loss of $9. | 100 | A hurts more than B | 37 |
| B hurts more than A | 36 |
| No difference | 27 |
| (A) You lose $9. (B) You lose $9 after suffering a loss of $18. | 100 | A hurts more than B | 31 |
| B hurts more than A | 28 |
| No difference | 41 |
| (A) You lose $9. (B) You lose $9 after suffering a loss of $36. | 100 | A hurts more than B | 38 |
| B hurts more than A | 33 |
| No difference | 29 |
| (A) You lose $9. (B) You lose $9 after suffering a loss of $45. | 100 | A hurts more than B | 35 |
| B hurts more than A | 33 |
| No difference | 32 |
| (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36. | 100 | A hurts more than B | 30 |
| B hurts more than A | 40 |
| No difference | 30 |
| Problem 9 | Giving away condition | 101 | $0 | 20 |
| $20 | 19 |
| $20 plus interest | 27 |
| $75 | 15 |
| A $55 saving | 20 |
| Drinking condition | 99 | $0 | 20 |
| $20 | 22 |
| $20 plus interest | 17 |
| $75 | 16 |
| A $55 saving | 24 |
| Problem 11 | Spend $50 on dinner.  Would you buy a $25 theater ticket later in the week? | 100 | Buy the ticket | 48 |
| Not buying the ticket | 52 |
| Given a $50 dinner.  Would you buy a $25 theater ticket-later in the week? | 100 | Buy the ticket | 48 |
| Not buying the ticket | 52 |
| Spend $50 on a sports ticket. Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 47 |
| Not buying the ticket | 53 |
| Given a $50 sports ticket.  Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 51 |
| Not buying the ticket | 49 |
| Spend $50 for an inoculation. Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 52 |
| Not buying the ticket | 48 |
| Spend $20 on dinner.  Would you buy a $25 theater ticket later in the week? | 100 | Buy the ticket | 53 |
| Not buying the ticket | 47 |
| Given a $20 dinner.  Would you buy a $25 theater ticket-later in the week? | 100 | Buy the ticket | 48 |
| Not buying the ticket | 52 |
| Spend $20 on a sports ticket. Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 47 |
| Not buying the ticket | 53 |
| Given a $20 sports ticket.  Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 47 |
| Not buying the ticket | 53 |
| Spend $20 for an inoculation. Would you purchase a $25 theater ticket later in the week? | 100 | Buy the ticket | 49 |
| Not buying the ticket | 51 |
| Problem 13 | Imagine that you have just won $30. | 200 | A 50% chance to gain $9 and a 50% chance to lose $9. | 99 |
| No further gain or loss. | 101 |
| Problem 14 | Imagine that you have just lost $30. | 200 | A 50% chance to gain $9 and a 50% chance to lose $9. | 99 |
| No further gain or loss. | 101 |
| Problem 15 | Imagine that you have just lost $30. | 200 | A 33% chance to gain $30 and a 67% chance to gain nothing. | 105 |
| A sure $10. | 95 |
| Problem 16 | A single coin flip, heads you win $200, tail you lose $100.  Would you take the bet? | 200 | Take the bet. | 99 |
| Not taking the bet | 101 |
| A package bet of 100 coin flips, each coin flip you either win $200 or lose $100.  Would you take the bet? | 200 | Take the bet. | 90 |
| Not taking the bet | 110 |
| Problem 17 | A project:  50% chance to gain $2 million, 50% chance to lose $1 million. | 200 | Undertake the project | 103 |
| Not undertaking the project | 97 |
| A portfolio of 25 of investments:  each has a 50% chance of gaining $2 million and 50% chance of losing $1 million. | 200 | Undertake the investments | 93 |
| Not undertaking the investments | 107 |
| Problem 18 | Paid $40 for tickets |  | Go to the game | 96 |
|  | Stay home | 104 |
| Tickets given by friends |  | Go to the game | 102 |
|  | Stay home | 98 |
| Problem 19 | Imagine that you joined a tennis club and paid a $300 yearly membership fee. | 200 | Stop playing | 102 |
| Continue to play | 98 |

*Note.* N represents sample size.

Table 11

*Descriptive statistics for problems with scale and text entry*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number** | **Condition**  **(if applicable)** | **N** | **Mean** | **Standard deviation** |
| Problem 7 | Hotel condition | 100 | 49.67 | 29.99 |
| Grocery condition | 100 | 47.79 | 26.74 |
| Problem 8 | Free condition-Market value $5-Friend | 66 | 89.98 | 54.32 |
| Free condition-Market value $5-Stranger | 66 | 97.06 | 58.72 |
| Free condition-Market value $10-Friend | 66 | 104.88 | 54.73 |
| Free condition-Market value $10-Stranger | 66 | 107.73 | 56.28 |
| Paid 5 condition-Market value $5-Friend | 66 | 101.36 | 55.07 |
| Paid 5 condition-Market value $5-Stranger | 66 | 104.14 | 61.49 |
| Paid 5 condition-Market value $10-Friend | 66 | 101.71 | 62.03 |
| Paid 5 condition-Market value $10-Stranger | 66 | 93.58 | 56.28 |
| Paid 10 condition-Market value $5-Friend | 68 | 100.50 | 51.75 |
| Paid 10 condition-Market value $5-Stranger | 68 | 96.62 | 54.41 |
| Paid 10 condition-Market value $10-Friend | 68 | 89.04 | 55.85 |
| Paid 10 condition-Market value $10-Stranger | 68 | 96.21 | 53.11 |
| Problem 10 | I feel like I just spent $400, much as I would feel if I spent $400 on a weekend getaway. | 200 | 2.88 | 1.35 |
| I feel like I made a $400 investment which I will gradually consume after a period of years. | 200 | 3.04 | 1.35 |
| I feel like I just saved $100, the difference between what the futures cost and what the wine will sell for when delivered. | 200 | 2.94 | 1.45 |
| I cannot understand this question. | 200 | 3.00 | 1.44 |
| Problem 12 | $15 condition | 101 | 32.46 | 28.71 |
| $40 condition | 99 | 5.83 | 27.78 |
| Problem 20 | The more you paid for the shoes, the more times you will try to wear them. | 200 | 2.76 | 1.45 |
| 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. | 200 | 2.87 | 1.38 |
| Problem 21 | Day expression | 67 | 47.12 | 27.12 |
| Year expression | 66 | 56.27 | 30.32 |
| Both expressions-Day | 67 | 49.60 | 31.04 |
| Both expressions-Year | 67 | 53.46 | 30.04 |

Table 12.

*Summary of all chi square results*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Problem** | ***χ²*** | ***χ²* continuity correction** | **Difference in 2 proportions** | **Odds ratio** | **Phi-coefficient** | **Cramer’s V** |
| 1 | 3.38 (df=1, p=0.066) | 2.88 (df=1, p=0.090) | -0.13  95% CI  [-0.27; 0.01] | 0.59  95%CI  [0.34; 1.04] | 0.13 | 0.13 |
| 2 | 8.83 (df=1, p=0.003) | 8.01 (df=1, p=0.005) | 0.21  95%CI  [0.07; 0.35] | 2.35  95%CI  [1.33; 4.14] | 0.21 | 0.21 |
| 3 | 0.51 (df=1, p=0.476) | 0.33 (df=1, p=0.568) | 0.05  95% CI  [-0.09; 0.19] | 1.23  95%CI  [0.70; 2.15] | 0.05 | 0.05 |
| 9 | 2.87 (df=4, p=0.580) | 2.87 (df=4, p=0.580) | / | / | / | 0.12 |

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

Table 13.

*Summary of all proportion tests results*

| **Problem** | **Conditions** | **Options** | **Observed proportion** | **Expected proportion** | ***χ²*** | **df** | **p** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 4 | Two wins: $50, and $25  versus  One win: $75.  Who was happier? | Two wins is happier | 0.32 | 0.33 | 1.57 | 2 | 0.456 |
|  | One win is happier | 0.38 |
|  | No difference | 0.31 |
|  | Two mistakes: $100, and $50  versus  One mistake: $150  Who was more upset? | Two mistakes is more upset | 0.32 | 0.33 | 0.91 | 2 | 0.634 |
|  | One mistake is more upset | 0.32 |  |  |  |  |
|  | No difference | 0.36 |  |  |  |  |
|  | Two events: Win $100, and loss $80  versus  One event: Win $20  Who was happier? | Two events is happier | 0.34 | 0.33 | 0.19 | 2 | 0.909 |
|  | One event is happier | 0.32 |  |  |  |  |
|  | No difference | 0.34 |  |  |  |  |
|  | Two events: Loss $200, and win $26  versus  One event: Loss $175  Who was more upset? | Two events is more upset | 0.41 | 0.33 | 5.77 | 2 | 0.056 |
|  | One event is more upset | 0.32 |  |  |  |  |
|  | No difference | 0.28 |  |  |  |  |
| 5 | Two events:  (1) win $25  (2) win $50  Who is happier? | Happier when occur on the same day | 0.40 | 0.33 | 4.03 | 2 | 0.133 |
|  | Happier when two events are two weeks apart | 0.33 |  |  |  |  |
|  | No difference | 0.28 |  |  |  |  |
|  | Two events:  (1) $100 must be paid  (2) $50 must be paid.  Who is more unhappy? | More unhappy when occur on the same day | 0.36 | 0.33 | 2.59 | 2 | 0.274 |
|  | More unhappy when two events are two weeks apart | 0.28 |  |  |  |  |
|  | No difference | 0.35 |  |  |  |  |
|  | Two events:  (1) a $20 parking ticket  (2) a $25 bill  Who is more unhappy? | More unhappy when occur on the same day | 0.34 | 0.33 | 0.19 | 2 | 0.909 |
|  | More unhappy when two events are two weeks apart | 0.34 |  |  |  |  |
|  | No difference | 0.32 |  |  |  |  |
| 6 | (A) You lose $9. (B) You lose $9 after having gained $30. | A hurts more | 0.40 | 0.33 | 2.00 | 2 | 0.368 |
|  | B hurts more | 0.30 |  |  |  |  |
|  | No difference | 0.30 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after having lost $30. | A hurts more | 0.29 | 0.33 | 1.58 | 2 | 0.454 |
|  | B hurts more | 0.39 |  |  |  |  |
|  | No difference | 0.32 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after having lost $250. | A hurts more | 0.33 | 0.33 | 0.02 | 2 | 0.990 |
|  | B hurts more | 0.34 |  |  |  |  |
|  | No difference | 0.33 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after suffering a loss of $1000. | A hurts more | 0.28 | 0.33 | 1.82 | 2 | 0.403 |
|  | B hurts more | 0.39 |  |  |  |  |
|  | No difference | 0.33 |  |  |  |  |
|  | (A) You lose $9 after suffering a loss of $30. (B) You lose $9 after suffering a loss of $1000. | A hurts more | 0.31 | 0.33 | 0.38 | 2 | 0.827 |
|  | B hurts more | 0.36 |  |  |  |  |
|  | No difference | 0.33 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after suffering a loss of $9. | A hurts more | 0.37 | 0.33 | 1.82 | 2 | 0.403 |
|  | B hurts more | 0.36 |  |  |  |  |
|  | No difference | 0.27 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after suffering a loss of $18. | A hurts more | 0.31 | 0.33 | 2.78 | 2 | 0.249 |
|  | B hurts more | 0.28 |  |  |  |  |
|  | No difference | 0.41 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after suffering a loss of $36. | A hurts more | 0.38 | 0.33 | 1.22 | 2 | 0.543 |
|  | B hurts more | 0.33 |  |  |  |  |
|  | No difference | 0.29 |  |  |  |  |
|  | (A) You lose $9. (B) You lose $9 after suffering a loss of $45. | A hurts more | 0.35 | 0.33 | 0.14 | 2 | 0.932 |
|  | B hurts more | 0.33 |  |  |  |  |
|  | No difference | 0.32 |  |  |  |  |
|  | (A) You lose $9 after suffering a loss of $9. (B) You lose $9 after suffering a loss of $36. | A hurts more | 0.30 | 0.33 | 2.00 | 2 | 0.368 |
|  | B hurts more | 0.40 |  |  |  |  |
|  | No difference | 0.30 |  |  |  |  |
| 19 | Imagine that you joined a tennis club and paid a $300 yearly membership fee. | Stop playing | 0.51 | 0.50 | 0.08 | 1 | 0.777 |
|  | Continue to play | 0.49 |  |  |  |  |

*Note.* df indicates degree of freedom

Table 14

*Summary of all McNemar paired-samples tests*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Problem** | **Comparisons** | ***χ²*** | **df** | **p** | ***χ²* continuity correction** | **Log odds ratio exact** |
| 5 | Comparing A to B  A-Two events:  (1) win $25  (2) win $50  Who is happier?  B-Two events:  (1) $100 must be paid  (2) $50 must be paid.  Who is more unhappy? | 3.52 | 3 | 0.318 | 3.52  (df=3, p=0.318) | / |
|  | Comparing A to C  A-Two events:  (1) win $25  (2) win $50  Who is happier?  C-Two events:  (1) a $20 parking ticket  (2) a $25 bill  Who is more unhappy? | 1.78 | 3 | 0.619 | 1.78  (df=3, p=0.619) | / |
| 13 | Baseline for Problem 14 and 15 | / | / | / | / | / |
| 14 | Problem 13 vs. 14 | 0.00 | 1 | 1.000 | / | 0.00 (p=1.000) |
| 15 | Problem 13 vs. 15 | 0.37 | 1 | 0.544 | / | -0.12 (p=0.614) |
| 16 | A single bet vs. 100 bets | 0.85 | 1 | 0.356 | / | 0.19 (p=0.412) |
| 17 | A project vs 25 investments | 1.11 | 1 | 0.292 | / | 0.22 (p=0.343) |
| 18 | Paid $40 vs. Given | 0.35 | 1 | 0.552 |  | -0.12 (p=0.621) |

*Note.* df indicates degree of freedom

Table 15

*Summary of all t-tests results*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Problem (test type)** | **Statistic** | **df** | **p** | **Mean difference** | **SE difference** | **Effect size** | **95% CI** |
| 7  (Independent samples t-test) | Student’s t=-0.47 | 198 | 0.640 | -1.88 | 4.02 | Cohen’s d =-0.07 | [-0.34;0,21] |
|  | Welch’s t=-0.47 | 195.46 | 0.640 | -1.88 | 4.02 | Cohen’s d =-0.07 | / |
| 12  (Independent samples t-test) | Student’s t=6.66 | 198 | <.001 | 26.63 | 4.00 | Cohen’s d =0.94 | [0.63; 1.25] |
| Welch’s t=6.67 | 197.97 | <.001 | 26.63 | 3.99 | Cohen’s d =0.94 |  |
| 20  (Paired samples t-test) | Student’s t=-0.73 | 199 | 0.467 | -0.10 | 0.14 | Cohen’s d =-0.05 | [-0.19; 0.09] |
| 21  (Independent samples t-test) | Student’s t=1.84 | 131 | 0.069 | 9.15 | 4.99 | Cohen’s d =0.32 | [-0.03; 0.66] |
| Welch’s t=1.83 | 128.95 | 0.069 | 9.15 | 4.99 | Cohen’s d =0.32 | / |
| 21  (Paired sample t-test) | Student’s t=-0.77 | 66 | 0.445 | -3.87 | 5.03 | Cohen’s d=-0.09 | [-0.33; 0.15] |

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

Table 16

*Summary of all ANOVA results*

|  |
| --- |
| **Problem 8-Mixed ANOVA** |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Within Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Friend vs. Stranger |  | 341.41 |  | 1 |  | 341.41 |  | 0.11 |  | 0.736 |  | 0.00 |  | | Friend vs. Stranger ✻ Q8-Condition |  | 1939.38 |  | 2 |  | 969.69 |  | 0.32 |  | 0.724 |  | 0.00 |  | | Residual |  | 590361.40 |  | 197 |  | 2996.76 |  |  |  |  |  |  |  | | Market worth $5 vs. $10 |  | 67.30 |  | 1 |  | 67.30 |  | 0.03 |  | 0.873 |  | 0.00 |  | | Market worth $5 vs. $10 ✻ Q8-Condition |  | 14842.18 |  | 2 |  | 7421.09 |  | 2.82 |  | 0.062 |  | 0.01 |  | | Residual |  | 518350.77 |  | 197 |  | 2631.22 |  |  |  |  |  |  |  | | Friend vs. Stranger ✻ Market worth $5 vs. $10 |  | 93.02 |  | 1 |  | 93.02 |  | 0.03 |  | 0.874 |  | 0.00 |  | | Friend vs. Stranger ✻ Market worth $5 vs. $10 ✻ Q8-Condition |  | 4255.14 |  | 2 |  | 2127.57 |  | 0.58 |  | 0.561 |  | 0.00 |  | | Residual |  | 723035.98 |  | 197 |  | 3670.23 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Between Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Q8-Condition |  | 3586.42 |  | 2 |  | 1793.21 |  | 0.54 |  | 0.586 |  | 0.00 |  | | Residual |  | 659388.46 |  | 197 |  | 3347.15 |  |  |  |  |  |  |  | |
| **Problem 10-Repeated measures ANOVA** |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Within Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Feeling |  | 2.47 |  | 2 |  | 1.23 |  | 0.63 |  | 0.532 |  | 0.00 |  | | Residual |  | 777.53 |  | 398 |  | 1.95 |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Between Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Residual |  | 370.50 |  | 199 |  | 1.86 |  |  |  |  |  |  |  | | | | | | | | | | | | | | | |
| **Problem 11-Mixed ANOVA** |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Within Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Given/Spent |  | 0.10 |  | 1 |  | 0.10 |  | 0.41 |  | 0.524 |  | 0.00 |  | | Given/Spent ✻ Q11-Condition |  | 0.00 |  | 1 |  | 0.00 |  | 0.01 |  | 0.943 |  | 0.00 |  | | Residual |  | 49.15 |  | 198 |  | 0.25 |  |  |  |  |  |  |  | | Dinner/Ticket |  | 0.03 |  | 1 |  | 0.03 |  | 0.14 |  | 0.708 |  | 0.00 |  | | Dinner/Ticket ✻ Q11-Condition |  | 0.10 |  | 1 |  | 0.10 |  | 0.45 |  | 0.501 |  | 0.00 |  | | Residual |  | 44.12 |  | 198 |  | 0.22 |  |  |  |  |  |  |  | | Given/Spent ✻ Dinner/Ticket |  | 0.00 |  | 1 |  | 0.00 |  | 0.00 |  | 0.949 |  | 0.00 |  | | Given/Spent ✻ Dinner/Ticket ✻ Q11-Condition |  | 0.10 |  | 1 |  | 0.10 |  | 0.33 |  | 0.568 |  | 0.00 |  | | Residual |  | 61.15 |  | 198 |  | 0.31 |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Between Subjects Effects | | | | | | | | | | | | | | |  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | | Q11-Condition |  | 0.00 |  | 1 |  | 0.00 |  | 0.01 |  | 0.941 |  | 0.00 |  | | Residual |  | 45.10 |  | 198 |  | 0.23 |  |  |  |  |  |  |  | | | | | | | | | | | | | | | |

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

## Exploratory analysis

For Problem 9 and 10, if we fail to replicate the original findings, we will re-run the analysis after excluding participants who report that they do not drink alcohol and dislike wine. For Problem 21, if we fail to replicate the original findings, we will try log-transforming the prices and removing all answers that are 3 standard deviations above the mean (with the criteria of p < .01 to adjust for multiple analyses). Meanwhile, in the actual data collection, we aim to examine the intercorrelations among mental accounting problems that support the original findings.

We planned to conduct a pooled analysis for all Problems with a clear indication of mental accounting effect. If the answer is influenced by mental accounting, we will count as 1; otherwise count as 0. We will report a correlation table among the coding of the problems.

## Comparing replication to original findings

We compared the results of the replication to the original findings when applicable based on LeBel et al. (2019) outcome interpretation criteria - 1) signal / no signal, 2) consistency / inconsistency, 3) larger / smaller / opposite effect, by comparing replication effect confidence intervals to the original effect size.

*[Note: The comparisons are not possible now since the current sample is randomly generated on Qualtrics. This part will be further revised after the actual data collection.]*

# Discussion

*[Note: The discussion is only to be completed in Stage 2 following data collection]*

## Limitations of present work and directions for future research

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

# References

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

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

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

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

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

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

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

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

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

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

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

Peels, R. (2019). Replicability and replication in the humanities. *Research Integrity and Peer Review*, *4*(1), 1–12. <https://doi.org/10.1186/s41073-018-0060-4>

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

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

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

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

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

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, *36*(6), 643–660. <https://doi.org/10.1287/mnsc.36.6.643>

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

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

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

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

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

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