

Exact Versus Conceptual Replication: Internet-based Research Investigating the Replicability of Cognitive Effects

Maria Rosa Miccoli* and Ulf-Dietrich Reips†‡

Research Methods, Assessment, and iScience, Department of Psychology, University of Konstanz, Germany
E-mail: maria-rosa.miccoli@uni-konstanz.de

Abstract. In recent years, the need for replication efforts has grown. Replication science faces key challenges, including achieving generalizability across heterogeneous samples and environments while streamlining the theory-experiment cycle to facilitate research efforts. Systematic replication projects using Internet-based methodologies address these challenges by facilitating access to diverse samples, employing rigorous testing, reducing costs, and ensuring materials are readily available. Standards for Internet-based experimenting provides transparency and reproducibility. We present three remote experiments, including one exact replication (N: 410) and two conceptual replications (N: 270; N: 365), which test the mental accounting effect based on Kahneman and Tversky's classic paradigm. The remote version of the exact replication maintained the same experimental design, instructions, and procedure as the original paradigm. In the two conceptual replications, we adapted the original price to the current value of money: Ticket price and the monetary loss were changed from 10\$ to 40€. In the first conceptual replication, we varied the original experimental design: the mental account variable was manipulated within-subjects. In the second conceptual replication, we varied the price stimulus while retaining the mental account manipulation in a between-subjects design. The exact replication replicated the original findings with an effect of small size, while the two conceptual replications replicated the results with an effect of increased size that is more comparable to the original findings. The results highlight the importance of adapting experimental paradigms to the current times, and the advantages of conducting remote replication projects step-by-step.

Keywords: replication, mental accounting, cognitive effects, remote research

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

[DOI: 10.2352/J.Percept.Imaging.2025.7.000408]

1. INTRODUCTION

Since the replication crisis and several scandals that uncovered cases of misinterpretation of statistical testing and results [45, 59, 60], great efforts have been made to increase the number of published replications and improve the methodologies that are used for replicating studies [19]. Still, some researchers consider that replications are only useful for the purpose of confirming the replicability of research

findings under specific circumstances [11], rather than to benefit the advancement of knowledge on a broader scale (generalizability to different contexts, contemporaneity of the research, and moderator effects) to enrich science [62].

Making use of the opportunities associated with the Internet revolution that were emphasized by the early pioneers of remote research [3, 32, 37, 38], we have witnessed the development of several “Big Team” efforts [7] and uses of crowdsourcing (the Hagen Cumulative Science Project [15] and the Psychological Science Accelerator [31]) to conduct replication studies in different laboratories around the world. These initiatives are beneficial to science, as they test and increase the generalizability of the results to different populations and contexts. Replication is crucial for distinguishing between true effects and chance findings in scientific research, particularly in fields with low base rates of real effects [57]. This is where remote replications offer a distinct advantage: they permit larger, more diverse participant pools and greater geographical and cultural variation, which help in assessing the generalizability of results. Remote experiments, for instance, permit easier replication and testing across different contexts, contributing to the evaluation of the robustness of findings. In this way, remote studies and Internet-based research play a significant role in improving the reliability of scientific knowledge.

Switching to Internet-based experiments that were originally conducted in the laboratory can be taken a step further. Evidence from various fields, for example, research on music perception and cognition [16], has shown that concerns about switching from laboratory to Internet-based experiments are often unwarranted, as results are often comparable. However, it remains fundamental to adopt rigorous methodologies in Internet-based research to ensure the reliability and comparability of results. Few concerns have mostly been focused on the reliability of data collection, and the lack of complete control of the experimental conditions [17]. For instance, issues such as repeated use of the same MTurk samples, internal validity concerns due to conditional dropout, and the potential for increased error variance have been highlighted [5, 34, 37, 38]. These challenges can lead to concerns about the quality of data obtained online, but it is important to note that Internet-based research has a long-standing tradition of rigorous methodologies [2, 40, 41, 43], predating the massive use of MTurk samples. In Internet-based research, several

*<https://orcid.org/0000-0002-4455-2654>

†<https://orcid.org/0000-0002-1566-4745>

‡[ResearcherID:A-5177-2012](https://orcid.org/0000-0002-1566-4745)

Received Feb. 5, 2024; accepted for publication Dec. 27, 2024; published online Feb. 24, 2025. Associate Editor: Kimberly A. Jameson.

2575-8144/2025/7/000408/15/\$00.00

techniques have been implemented to enhance internal validity by addressing dropout issues. For instance, the multiple-site entry technique [39] was used to monitor sample effects and differing dropout patterns. Implementing seriousness [1] and quality control [12] checks improve data quality. By employing these methods, researchers can better ensure data integrity and enhance the reliability of findings in remote studies. Moreover, psychological research has made significant strides in improving techniques for conducting remote studies, particularly in the aftermath of COVID-19 pandemic, which accelerated the development of best practices for ensuring data reliability in online environments [46].

The increase in the number of Internet-based studies has prompted the development of rigorous research methodologies [40, 41] that provide Internet-based replication projects a great opportunity. The advantages of replicating experiments in Internet-based environments are several. From a practical perspective, subject pool diversity, lower costs, and lower hurdles to run studies make the theory-experiment cycle very smooth.

From an educational perspective, introduction to Open Science practices training and replication studies have become a part of basic research training, which would benefit from complete and accessible materials [24, 25, 61]. Moreover, meta-analyses are essential for advancing research and serving educational purposes. They help to determine whether an effect exists and reduce uncertainty, but their execution is often hindered by difficulties in accessing data from multiple studies. While the growing interest in Open Science is gradually easing this process, digitalized data from remote studies also play a crucial role in facilitating broader access to previous studies [22].

From a strictly methodological perspective, criticisms of replication studies have centered on concerns regarding internal validity, as noted by Feest [11], who pointed to overlooked confounding variables and the potential misidentification of relevant causal features of stimuli. Remote experiments offer the opportunity to easily make experimental materials available and thus facilitate simplified reproducibility and mitigate threats to internal validity [1, 12, 39]. It is important to note that repeated investigations of the same phenomenon inherently result in heterogeneity of effect sizes, which can be attributed to random variation across studies [23]. However, this variability should be valued, as it contributes to the generalizability of findings [37, 38].

Machery [29] emphasized the importance of sample heterogeneity in replication studies and ensuring that sample characteristics are clearly reported. While the criticism of replication studies' internal validity [11] can be debated, Internet-based research offers advantages that address these concerns. The transparency afforded by making materials fully available online allows researchers to detect confounding factors that may have been missed in the original studies and provide additional information about the potential sources of random variation across studies. Furthermore, Internet-based studies permit recruitment of more

heterogeneous samples, addressing Machery's concerns [29] of sample diversity. Online platforms also facilitate techniques, such as Reips' *multiple site entry technique* [39], which supports diverse data collection and sample characteristic reports.

To summarize, Internet-based (remote) replication studies can be easily implemented with access to large, heterogeneous samples and promote transparency through materials that are readily available for replication efforts. Additionally, they offer valuable learning opportunities for students and researchers, fostering broader engagement in replication efforts.

Replication projects about cognitive effects are fundamental. They help in developing the understanding of theories, assessing the methodological characteristics of previous studies, and identifying the conditions that moderate or limit the generalizability of specific findings. The cognitive effect "mental accounting" is defined as the human tendency to sort resources into mental categories, each characterized by different subjective values denoted as x units that impact decision-making processes [56]. When applied to money, mental accounting refers to the tendency to categorize financial resources differently based on the *mental accounts* they are assigned to, and the subjective value associated with each account.

In this article, we integrate two goals: (1) to provide evidence emphasizing the importance of systematic replication studies in advancing science, and (2) to illustrate how Internet-based methodologies support replication studies through methodological rigor and the facilitation of diverse participant recruitment, thereby broadening the generalizability of research findings.

We conducted three remote replication experiments on the mental accounting effect [21], adopting them as empirical examples to evaluate the strengths and weaknesses of *exact* versus *conceptual* replication approaches online to increase the likelihood of replicating cognitive effects.

1.1 Exact Versus Conceptual Replications

The two major types of replication are *exact* and *conceptual* replications.

In *exact replication* attempts, researchers execute experiments rigorously similar as possible to the original experiment, i.e. adopt the same experimental designs and use the original stimuli. The extent to which the study must adhere to the original conditions (laboratories, physical settings, instructions, demography of experimenters) is a subject of methodological debate [18].

Given the inherent impossibility of precisely replicating the exact same conditions and historical time to assess the generalizability of effects [37, 38, 52], exact replication studies should extend beyond identical settings. Therefore, we advocate the importance of replicating experiments not only in different physical environments but also in Internet-based settings.

The term "exact replication" is used in a broad sense [18] here, encompassing replication studies that maintain the

same core elements such as experimental designs, main statistical analyses, and key stimuli as in the original experiments; for instance, the same prices in experiments involving monetary values. Following this definition, exact replications can involve few adjustments, such as shifting from a laboratory to an Internet-based setting, but the aforementioned core elements are preserved. While exact replications aim to maintain the same core elements, practical adjustments like modifying stimulus timing, incentives, or instructions—especially when necessary due to changes in the testing environment—do not constitute major conceptual changes and are classified as exact replication. These minor adjustments ensure that the replication remains feasible without altering the theoretical conceptualization and, in principle, operationalization of the original study.

In cognitive research, the main goal and strength of exact replication studies is to assess the internal validity of cognitive effects [18]. However, exact replications have been criticized for the risk of overlooking confounding variables due to maintaining precisely the same specific setting conditions [11]. Ultimately, full exact replications are impossible because of changes in (historical) times.

A practical example in cognition is the perception of money's value, which plays a major role in the mental accounting of money. Indeed, inflation rates should be considered when conducting studies involving money. Inflation decreases the purchasing power of currency over time, i.e., the same amount of money will have a different value at a later time. For example, 1 dollar in 1984 (original article publication by Kahneman & Tversky) was worth ~ 2.62 dollars in July 2021 (Experiment 1 data collection) according to the CPI Inflation Calculator [8]. The cognitive processing of mismatch [33] between real-world prices and prices used in the experiment may have impacted participants' decisions. Based on this effect of inflation, performing a replication project assessing the importance of price adjustments to contemporary times is fundamental.

Conceptual replication attempts maintain the original experiment's independent and dependent variables while varying the operationalization of the variables to different degrees [28]. Conceptual replications encompass both variations in experimental designs and modifications to key components of the experiments, such as altering stimuli or defining the vignettes' characteristics (key elements of variable operationalization) differently. In the first type of conceptual replication, variables remain constant, and alternative operationalizations are explored. The second type introduces changes to the variables themselves, such as altering the characteristics of the stimuli or vignettes (key elements of variable operationalization).

Conceptual replications assess the external validity of the cognitive effects: Adopting different main components and experimental designs enhance the understanding of cause-effect relationships across different measures, populations, and stimuli [54]. They aim to support the theoretical hypotheses of original experiments by operationalizing the experimental design and the stimuli in various ways [55].

However, conceptual replications pose a significant challenge: Maintaining and replicating the cognitive effects when varying the experimental circumstances requires a deep understanding of the concept and underlying psychological mechanisms [47]. To increase the chances of a well-executed conceptual replication, focusing on the theoretical constructs behind the theory requires considering a few variables to be changed.

Both exact and conceptual replications play critical roles in advancing cognitive sciences. Exact replications assess the internal validity of cognitive effects by testing whether they can be observed under similar conditions, ensuring the reliability of the original findings. Conceptual replications extend the generalizability and external validity of these effects by testing whether the underlying theory holds when critical aspects of the experimental design (e.g., key stimuli) are altered. In this study, we propose a process-oriented replication approach, combining exact and conceptual replications to ensure the robustness and generalizability of the cognitive effects under investigation. This approach not only strengthens the validity of the findings but also sheds light on how context-sensitive cognitive processes may change under different conditions, such as variations in monetary value.

By specifically recognizing the profound impact of sensitivity to context changes [58] and prices on individual willingness to pay [14], we make a comparison among one exact and two conceptual replication studies. We propose a process-oriented approach that integrates exact and conceptual replications for three main reasons: (1) addressing internal and external validity of the cognitive effects [18]; (2) defining a step-by-step replication project to progressively evaluate the key components of the experiment that replicate the effects; (3) conducting exact and conceptual replications in a process helps to understand the current status of the theory [19]. In line with previous studies adopting step-by-step procedures for testing new cognitive effects [19], we posit that old theories must also be tested with a similar approach. This approach evaluates the original methods and whether changes are needed for updating a specific theory, with reference to the general context [58] – e.g., inflation.

1.2 *The Original Study by Kahneman and Tversky*

To provide context for the present study, we first summarize the original study and its purpose. Kahneman and Tversky [21] adopted the theater ticket experiment to test the mental accounting of money effect: Participants changed their hypothetical behavior in response to a hypothetical loss of the same objective value (i.e., 10\$). Crucially, this change in behavior was contingent on the mental account to which the loss was associated (ticket versus bill). Kahneman and Tversky [21] tested mental accounting of money with their theater ticket paradigm, in a laboratory setting. For this purpose, participants were instructed to envision a scenario in which they had recently incurred a loss of 10\$, either as a theater ticket (same mental account) or as a dollar bill

(different mental account). Losing a ticket was classified as “same” mental account based on the hypothesis that the cost of the lost theater ticket is already mentally associated with the expense category of going to the theater. In contrast, losing a \$10 bill is not connected to the theater category, and thus, represents a “different” mental account. A factorial design varying the mental account variable (same – ticket loss – versus different – bill loss) between-subjects¹ was implemented. The dependent variable was operationalized as a hypothetical behavior referring to the willingness to pay for the theater ticket even after losing 10\$ (as a ticket or bill). The hypothetical behavior was assessed using a binary response format (yes versus no). The results showed a difference among participants willing to pay for a ticket in the ticket condition (yes – 46% versus no – 54%) and in the bill condition (yes – 88% versus no – 12%). The authors explained their results with the concept of “topical organization” of mental accounts: People categorize, organize, and value their financial transactions differently based on the topic or context with which such transactions are associated. Attending a theater performance is perceived as a transaction where the ticket cost is traded for the experience. Buying a second ticket would raise the overall cost. Conversely, the cash loss is not directly linked to the theater experience.

1.3 The Present Study

In our study, we employed the theater ticket experiment as a test case to highlight that due to historical changes, reproducing the original study’s results with similar effect sizes is unlikely when attempting to conduct full exact replications. In the case of perception value of money, factors such as shifts in the perceived value of money, inflation, and other historical developments (e.g., changes in payment methods) must be considered when conducting replication studies. Thus, we assessed the importance of adapting the monetary values to contemporary times, and included an additional variable, along with changes in payment methods: the purchase medium independent variable. Over the past few decades payment methods have evolved following the advent of the Internet, allowing consumers to purchase items remotely. Researchers have explored the role of purchase media (traditional – in-person - versus modern – online - purchases) on purchase intentions [10], suggesting that the purchase media may play a role in consumers decisions.

We replicate the classic theater ticket experiment with the inclusion of an additional variable “purchase medium”. The purpose of this experiment is to investigate two parameters: the replicability of the effect and the potential influence of adapting the experimental context to contemporary settings on the subjective perception of money.

We conducted three remote experiments, including one exact replication and two conceptual replications testing the mental accounting effect (Experiments, data sets, and analysis scripts are available [here](https://osf.io/s3a8b/) (https://osf.io/s3a8b/) through the Open Science Framework.). The remote version of the exact replication maintains the same experimental design, instructions,² ticket price, and between-subjects manipulation for the mental account variable as in the original paradigm to adhere to the broad definition of “exact replication” from Hudson [18]. The two conceptual replications maintain the same instructions, but they are characterized by an adaptation of the monetary value to the contemporary times, altering the characteristics of the ticket price in terms of monetary value: The ticket price and the monetary loss were changed from 10\$ to 40€, considering inflation and conversion rates. Additionally, in the first conceptual replication, we introduced a variation in the experimental design by manipulating the mental account variable within-subjects. The decision to switch from a between-subjects to a within-subjects design in the first conceptual replication was motivated by one of its methodological advantages [6]. A within-subjects design allows participants to serve as their own control, reducing variability from individual differences and increasing the statistical power of the study. This design is particularly useful for examining cognitive effects, such as mental accounting. In the second conceptual replication, as in the original experiment and our exact replication, the mental account variable was manipulated between-subjects, and we exclusively altered a key characteristic: the ticket price. This decision was based on two objectives: first, to evaluate whether the design itself plays a role in replicating the cognitive effect; and second, to balance the strengths and limitations of within- and between-subjects approaches. While a within-subjects design increases sensitivity and controls for individual differences, a between-subjects design reduces potential carryover effects and ensures independent treatment conditions [6]. Combining both designs allowed us to rigorously assess the robustness of the mental accounting effect across different methodological frameworks.

We modified the monetary value of the ticket price in two conceptual replications to assess whether the original cognitive effects hold in a different currency context and when adapting the monetary value to contemporary times.

Across all three replications, we explored the impact of the purchase medium on individual behavior by varying this variable between-subjects. All three replications were presented in English to the participants, and participants were informed of this before starting the study.

Hypothesis 1 pertains to the cognitive phenomenon being tested here, mental accounting of money. Building on the foundational work of Kahneman and Tversky [21], we formulated our first hypothesis, as follows.

¹ Kahneman and Tversky [21] mention that both conditions were presented to a sub-sample of the participants in an exploratory manner, but they did not provide a detailed report of the results from this exploratory analysis.

² Participants were asked to imagine being on their way for a theater play and finding out that they had lost 10\$ (as a ticket or as a bill). They were asked whether they would still pay for the ticket, with a binary response (yes versus no).

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

Yes

No

Figure 1. Example of a scenario web page from Experiment 1. The example shows one of the experimental scenarios (same mental account and traditional purchase medium) and the options provided to participants to report their willingness to pay.

Hypothesis 1. For a hypothetical monetary loss, individuals will display a greater propensity to allocate additional financial resources towards a theater ticket when the monetary loss is defined as a different mental account (bill), compared to when defined as the same mental account (ticket). A higher number of people will report a hypothetical intention to buy the theater ticket after losing cash than after losing the ticket.

Drawing from prior research on the effect of the purchase medium on consumer intentions [10], our second hypothesis seeks to explore how this factor may influence the willingness to buy a ticket in the context of our study.

Hypothesis 2. When the purchase medium is modern (online booking), it is more likely that people will display a higher willingness to buy a ticket than when the purchase medium is traditional (box office).

Informed by discussions on the importance of replication in scientific research [19, 50], our third hypothesis addresses the relevance of conceptual replication studies within the context of our investigation.

Hypothesis 3. We hypothesize that the mental accounting effect will be replicated with higher strength (effect size) in the conceptual replication studies than in the exact replication study.

2. EXPERIMENT 1

Experiment 1 was conducted to evaluate whether the mental accounting effect can be replicated by using the original ticket and loss price (10\$). To pursue the goal of adapting the experiment to contemporary times, we explored the role of purchase medium impact on individual decisions.

2.1 Method

2.1.1 Participants

Participants were recruited via social media such as Reddit (144), Facebook (45), WhatsApp (15) and Telegram (11) platforms; SurveyCircle (55), a mailing list for Internet Researchers – the Association of Internet Researchers (AoIR) – (53), an independent platform aggregating psychological studies – Psychological Research on the Net – (48), and the University of Konstanz SONA system (39), a platform

where students gain credits for taking part in research studies. We adopted the seriousness check as an inclusion criterion [1, 40]: Only data from participants who confirmed their commitment to serious participation prior to starting the experiment were included in the analyses. Of 482 datasets from participants who indicated their intention to seriously participate in the experiment, a final sample of 410 datasets was used for analyses. Detecting IP addresses has been recommended to exclude possible multiple submissions from the same participants [42]: Seventeen cases were not included in the analysis for this reason, only the first submission belonging to such IP addresses was included in data analyses. Additionally, the data of 55 participants were not included because of missing items that referred to the main questions (two items, one investigating mental accounting and one investigating the preference for theater).

The reported age of the participants ($N = 410$; 236 females) showed a median of 25 years ($SD = 11.37$, range 10–70).

2.1.2 Design and Procedure

In WEXTOR [44], we implemented a 2×2 between-subjects factorial design varying the mental account (same – ticket loss versus different – bill loss) and the purchase medium (traditional – box office versus modern – online booking).

Participants were each shown one scenario, depending on the condition they were randomly assigned to. As in the original experiment by Kahneman and Tversky [21], participants were instructed to envision a scenario, in which they had recently experienced a loss of 10\$, either as a theater ticket (same mental account) or as a bill (different mental account). Moreover, depending on the purchase medium condition, the hypothetical scenario described a scene where the theater ticket was acquired either directly at the box office (traditional) or through online booking (modern).

The dependent variable was operationalized as a hypothetical behavior referring to the willingness to pay the theater ticket after losing 10\$ (as ticket or bill), as shown in Figure 1. As in the original experiment, the hypothetical behavior was measured through a choice between mutually exclusive options (Yes versus No), asking participants whether they would be willing to pay for a 10\$ ticket, after the experienced loss.

The remote experiment took approximately 2 minutes and included information about informed consent, a seriousness check [1, 40], socio-demographic questions, one question referring to the research question, and one question about the preference for theater. The question about preference for theater (“How much do you like going to the theater?”) was included as an exploratory measure (Appendix A, Figure A2) to assess whether individual differences in preference for going to the theater could be linked to variations in responses to the main questions. Although this question was not part of the original study and is not central to testing our hypotheses, we report the results in Appendix B.

Experiment 1 was conducted from May to July 2021.

2.1.3 Data Analyses

In line with the original study by Kahneman and Tversky [21], we calculated participants’ willingness to pay for different conditions and used a contingency table to replicate the basic structure of their analysis. This provided an overview of participants’ responses across conditions. The contingency table approach was our main analysis method, aligning with the data analysis approach from the original study [21]. Additionally, we reported the response time in different conditions to evaluate whether experimental conditions impacted the response time. All analyses were performed using R Statistical Software [36].

To supplement this, we employed binary logistic regression to explore the probability of participants’ willingness to pay for a theater ticket based on the main variables under consideration—mental account and purchase medium—while accounting for theater preference as a covariate. This exploratory analysis enhances our understanding of the data by assessing the individual contributions of each predictor [53]. Given that we included two main predictors, the binary logistic regression also served as a statistical control for the results of the main analyses. The regression analysis was conducted using the `glm()` function from the core package `stats` (Version 4.3.3), corresponding to the R version used during the analyses. It is reiterated that this analysis was exploratory, aiming to further clarify the role of our additional variable and covariate and provide deeper insights into the effects of the predictors.

2.2 Results

In line with the original theater ticket experiment [21], we determined participants’ willingness to pay under different conditions. Our aim was to examine the replicability of the mental accounting of money effect (Hypothesis 1) and to assess if the online booking medium (i.e., modern purchase medium) would increase participants’ willingness to pay (Hypothesis 2).

2.2.1 Main Analyses

First, we focus on the mental account condition, without considering the purchase medium, to compare our results to the original study’s results. In the same mental account

Table 1. Number of participants willing and unwilling to pay (another) theater ticket.

		Ticket loss		Bill loss	
		Yes	No	Yes	No
Traditional	Count	61	34	80	25
	% of condition	64	36	76	24
Modern	Count	71	36	86	17
	% of condition	66	34	83	17

Note. $N = 410$. Frequencies and proportions of participants answering *yes* or *no* to paying for a theater ticket, depending on the mental account condition (between-subjects, ticket versus bill), further stratified by the purchase medium condition (between-subjects, traditional versus modern).

condition, the proportion of individuals willing to pay (65%) for a theater ticket was notably higher than those expressing an unwillingness to pay (35%), after incurring a 10\$ ticket loss. Similarly, in the different mental account condition, the percentage of individuals willing to pay for a theater ticket (80%) outnumbered the percentage of those who reported being unwilling to pay for a theater ticket (20%) after losing a bill worth 10\$.

Second, we present a contingency table that includes the purchase medium condition. Table 1 shows the number and proportion of participants reporting their willingness to pay (yes versus no) depending on the mental account condition (columns) and the purchase medium condition (rows).

Participants, on average, took 1 minute (i.e., 66'') to complete the experiment (Min: 18''; Max: 346'').³ No noteworthy difference in average session length was observed among different between-subjects conditions (traditional \times ticket: 61''; traditional \times bill: 66''; modern \times ticket: 66''; modern \times bill: 70''). Interestingly, the average session length of participants answering *yes* was slightly faster (traditional: 64''; modern: 69'') than that of participants answering *no* (traditional: 72''; modern: 77'') in the bill condition (also see Appendix B).

2.2.2 Binary Logistic Regression (Exploratory Analysis)

In addition to conducting the analysis conducted in the original experiment, we conducted a binary logistic regression (Appendix B) to examine the impact of the independent variables (mental account, purchase medium) and the preference for theater on the binary dependent variable (willingness to pay). This additional analysis helped to investigate whether the preference for theater played a role in participants’ willingness to pay for a ticket, after experiencing a 10\$ loss.

We fitted a logistic model (estimated using maximum likelihood) to predict the willingness to pay (WTP) with mental account, purchase medium and theater preference

³ Based on the laboratory pre-test and on the general average, participants who took more than 10 minutes were not included in this calculation. Therefore, data from sixteen participants are not included in this specific calculation.

(formula: $WTP \sim \text{'Mental account'} + \text{'Purchase medium'} + \text{'Theater preference'}$). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. Confidence intervals (95% CIs) and p -values were computed using a Wald z -distribution approximation. The model's explanatory power was weak (Tjur's $R^2 = 0.07$). The model's intercept, corresponding to mental account = bill, purchase medium = traditional and theater preference = 0, was 0.20 with 95% CI $[-0.41, 0.82]$, $p = 0.5$. The effect of mental account (ticket) was statistically significant and negative: beta = -0.84 , 95% CI $(-1.31, -0.38)$, $p < 0.001$. So, the probability that participants would show a willingness to pay was significantly lower in the ticket condition than in the bill condition. The purchase medium (modern) showed a positive and non-statistically significant effect, beta = 0.28, 95% CI $[-0.17, 0.73]$, $p = 0.23$. The effect of theater preference was statistically significant and positive, beta = 0.01, 95% CI $[0.005, 0.01]$, $p < 0.001$: As the preference for theater increased, the likelihood that participants would report a willingness to pay for a theater ticket increased as well.⁴

2.3 Discussion

Experiment 1 does not exactly replicate the results from Kahneman and Tversky's experiment [21]. They observed a large difference between willingness to pay for a ticket in the ticket condition (46%) and in the bill condition (88%). In this study, the results of the replication in Experiment 1 indicate a smaller difference in willingness to pay for a ticket in the ticket condition compared to the bill condition, with a 15-percentage-point difference (versus a 42-percentage-point difference in the original study). When focusing exclusively on the mental account condition, we observed a narrower difference between willingness to pay for a ticket in the ticket condition (65%) and in the bill condition (80%).⁵ This holds true for traditional (ticket – 64% versus bill – 76%) and modern (ticket – 66% versus bill – 83%) purchase medium conditions.

The binary logistic regression reiterates that the mental account condition significantly influences the likelihood of individuals purchasing a ticket, after experiencing a 10\$ loss. However, given the exploratory nature of this analysis and that it was not conducted in the original study, we treat this result with caution.

There are two potential explanations for the observed results in the main analyses, which show a replication with a smaller effect than the original findings. First, in line with our hypothesis emphasizing the importance of adjusting monetary values to contemporary times, our findings indicate that inflation weakens the mental accounting of money effect.

⁴ Appendix B depicts the detailed results from the binary logistic regression.

⁵ We report the magnitude of the percentage-point difference in this study. In the [R scripts](#) on the Open Science Framework, we present the Cramér's V values for our experiments as an additional measure of effect size, along with the estimated Cramér's V from the original experiment [21].

As suggested by prior research [33], cognitive processing of discrepancies between real-world prices and experimental prices significantly influences individual decision-making.

Second, we included the purchase medium variable to explore its role in individual decisions. This choice is motivated by prior studies that have examined the role of contextual sensitivity in scientific studies [58] and emphasized that people may vary in their degree of sensitivity to the context. However, our results indicate that solely adjusting the purchase medium to align an experiment with contemporary times is insufficient to replicate cognitive effects at their original effect size. When focusing only on the mental account condition, our main analyses revealed a difference of 15-percentage-point in willingness to pay for a ticket between the ticket and the bill conditions, whereas Kahneman and Tversky [21] reported a difference of 42-percentage-point. Additionally, when examining the purchase medium conditions across the ticket conditions (traditional and modern), we observed that a greater number of participants were willing to pay for another ticket compared to those who were not. This finding contrasts with the original study, which reported the opposite trend in the ticket condition.

Third, even though the experiment's setting shifted from a laboratory to an Internet-based environment, previous studies have pointed out that this change should not pose a problem per se, as data collected in laboratory and Internet-based settings have been shown to yield comparable results in testing cognitive effects [9, 49, 51]. Given this information, the rigorous methodologies used in Internet-based studies, and the cost-effectiveness of remote research, we advocate remote replications, even for cognitive effects originally tested in laboratory settings. Nevertheless, it is possible that differences in participant motivation [20] or social desirability [48] between the two settings may have played a role in not exactly replicating the cognitive effect.

As Experiment 1 did not replicate the size of the effect that was found in the original study, in the following experiments, we conducted conceptual replications. The purpose of the following replications was to find out whether indeed the assumed inflation effect on the motivational value of money was responsible for the reduced size of the effect. We thus adapted the price in the experimental scenario to the contemporary context. Moreover, we assessed whether varying the mental account variable within- versus between-subjects could alter the effect size of the cognitive effect. Additionally, we explored the role of the frequency with which participants go to the theater and the maximum price they are willing to pay for a theater ticket. The aim was to examine the potential relationship between participants' answers with (1) their theater habits and (2) the monetary value they associate with a theater ticket.

3. EXPERIMENT 2

Experiment 2 was conducted to evaluate whether the mental accounting effect can be replicated with an effect size similar to the original study, specifically adapting

Imagine that you have decided to see a play and paid the admission price of €40 per ticket via online booking. As you enter the theater, you discover that you lost your printed ticket. You cannot find the confirmation e-mail and the ticket cannot be recovered. Would you pay €40 for another ticket?

Yes

No

Figure 2. Example of a scenario web page from Experiment 2. The example shows one of the experimental scenarios (same mental account and modern purchase medium) and the options that participants could select to report their willingness to pay for another ticket.

the monetary values of a theater ticket and associated losses to contemporary standards (40€).⁶ Additionally, we varied the experiment by manipulating the mental account variable within-subjects, to understand whether the reduced variability associated with individual differences could play a role in the effect size of the mental accounting effect.

3.1 Method

3.1.1 Participants

Participants were recruited through social media such as Reddit (141) and Facebook (2), and AoIR mailing list (29) as well as the University of Konstanz SONA platform (98).

Akin to Experiment 1, in Experiment 2 we utilized the seriousness check as an inclusion criterion [1, 40]. Of 334 participants who indicated their intention to seriously participate in the experiment [1, 40], a final sample of 270 datasets was used for analyses. Nineteen cases were not included in the analysis due to multiple submissions with the same IP; only the first submission from an IP address was included in data analyses [42]. Additionally, 45 datasets were not included because of missing items that referred to the main questions (two research questions investigating the mental accounting principle, and one question about the maximum price that participants were willing to pay in real life for a theater ticket).

The reported age of the participants ($N = 270$; 176 females) showed a median of 20 years ($SD = 9.39$, range 15–70).⁷ Most participants reported being from Germany (129), from other European countries (59), or from the United States (44), and being students (149).

3.1.2 Design and Procedure

The main change from Experiment 1 was the inflation adjusted adaptation of the ticket price to current value.

⁶ Before starting the official data collection for Experiment 2, we assessed the appropriateness of the adjusted ticket price by conducting a pre-test via the University of Konstanz – SONA platform (49). Participants were asked to report the maximum price they were willing to pay for a theater ticket, with the majority reporting a willingness to pay 31–40€. These data were not included in the analyses reported here.

⁷ The standard deviation for reported age in Experiments 1 and 2 must be interpreted with caution due to the use of broad categories for the youngest (“below 10”) and oldest (“over 69”) participants.

Otherwise similar, in WEXTOR [44], we implemented a 2×2 mixed factorial design varying the purchase medium condition between-subjects (traditional – box office versus modern – online booking), but varied the mental account condition (same – ticket loss versus different – bill loss) within-subjects.

The participants were randomly assigned to one of the two purchase medium conditions (traditional versus modern). Participants were each shown two scenarios, in which they were asked to imagine that they had just lost 40€ either as a theater ticket (same mental account) or as bills (different mental account). The measure was a yes-no choice, as shown in Figure 2.

The remote experiment took approximately 2 minutes. Everything was done as in Experiment 1, except for the value of the ticket price (upgraded to 40€) and two additional exploratory questions: One question about the frequency with which participants go to the theater, and one about the maximum price that participants would be willing to pay in real-world for a theater ticket (Appendix A).

Akin to Experiment 1, the preference for going to the theater was included as an exploratory measure. Additionally, in Experiment 2, we included an exploratory measure about the frequency of going to the theater.

The experimental conditions were counterbalanced to control for order effects. Experiment 2 was conducted from January to June 2022.

3.1.3 Data Analyses

Akin to Experiment 1, and similar to the original study [21], we calculated the number of participants who were willing to pay, and we used a contingency table to replicate the basic structure of analysis and an overview of participants’ responses across conditions as well as the response time in different conditions.

Similar to Experiment 1, in Experiment 2 we employed binary logistic regression to explore the probability of participants’ willingness to pay for a theater ticket based on the main variables under consideration—mental account and purchase medium—while accounting for the maximum price they were willing to pay for a theater ticket. The binary logistic regression analysis was conducted as in Experiment 1. Due to the small effects observed in Experiment 1, it was necessary to investigate the role of participants’ individual

Table II. Number of participants willing and unwilling to pay for (another) theater ticket.

		Ticket loss		Bill loss	
		Yes	No	Yes	No
Traditional	Count	42	84	79	47
	% of condition	33	67	63	37
Modern	Count	56	88	107	37
	% of condition	39	61	74	26

Note. $N = 270$. Frequencies and proportions of participants answering *yes* or *no* to paying for a ticket, depending on the combination of mental account (within-subjects, ticket versus bill) and purchase medium (between-subjects, traditional versus modern) conditions.

maximum ticket price, as well as the impact of the independent variables (mental account and purchase medium) on the binary dependent variable (willingness to pay). This exploratory analysis aimed to examine whether the maximum price that participants were willing to pay for a theater ticket influenced their willingness to pay (Appendix B).

3.2 Results

3.2.1 Main Analyses

Akin to Experiment 1, the number of participants willing to pay in different conditions was calculated through a contingency table analysis to test Hypotheses 1 and 2.

First, we focused on the mental account condition. In the ticket condition, the proportion of participants willing to pay (36%) for a theater ticket was lower than that of participants unwilling to pay (64%). Conversely, in the bill condition, the proportion of participants willing to pay for a theater ticket (69%) was higher than the proportion of participants unwilling to pay for a theater ticket (31%).

Second, Table II depicts the count and proportion of participants reporting their willingness to pay for a theater ticket depending on the mental account conditions (columns) and the purchase medium conditions (rows).

Participants took 2 minutes on average (i.e., 118'') to complete the experiment (Min: 41''; Max: 10').⁸ No remarkable difference in average session length was observed among different between-subjects conditions (traditional: 114''; modern: 120''). However, participants were, on average, slightly slower in answering *yes* (traditional: 122''; modern: 124'') than in answering *no* (traditional: 107''; modern: 115''), as shown in detail in Appendix B.

3.2.2 Binary Logistic Regression (Exploratory Analysis)

We fitted a binary logistic model (estimated using maximum likelihood) to predict the willingness to pay for a ticket with mental account, purchase medium and the maximum price willing to pay for a theater ticket. The model's explanatory power was substantial (Tjur's $R^2 = 0.30$). The model's

⁸ Based on the pre-test, participants who took more than 10 minutes were not included in this calculation. Therefore, data from 10 participants were not included in this specific calculation.

intercept, corresponding to mental account = bill, purchase medium = traditional and maximum price willing to pay = between 1 and 10€, was -1.16 with 95% CI $[-2.43, -0.13]$, $p = 0.04$. The effect of mental account [ticket] was statistically significant and negative, $\beta = -1.76$, 95% CI $[-2.20, -1.34]$, $p < 0.001$: The probability that participants would show a willingness to pay was significantly lower in the ticket condition than in the bill condition. Purchase medium's [Modern] impact was statistically non-significant and positive: $\beta = 0.25$, 95% CI $[-0.17, 0.66]$, $p = 0.24$. The maximum price participants were willing to pay had a significant impact on participants' likelihood of willingness to pay for a ticket. Specifically, individuals indicating a willingness to pay more than 21€ for a theater ticket in the real world were significantly more likely to express a willingness to pay for a ticket compared to those who were willing to pay between 1 and 10€ for a theater ticket (Appendix B).

3.3 Discussion

The findings from Experiment 2 indicate that adjusting the monetary value of the theater ticket and the corresponding financial loss to reflect contemporary conditions successfully replicated the original mental accounting effect with a similar magnitude. The results obtained by Kahneman and Tversky [21] were replicated with a similar effect size in this experiment (Hypothesis 3). Hypothesis 1 was supported by the results both in the frequency analysis and binary logistic regression. Respondents were keener on buying a ticket after losing cash (in the form of bills) than after losing a ticket, even though the ticket and the cash were worth the same amount of money. When focusing solely on the mental account condition, we observed a substantial difference in the willingness to pay for a ticket between the ticket loss condition (36%) and the bill loss condition (69%). This 33-percentage-point difference aligns more closely with the original findings by Kahneman and Tversky [21], who reported a 42-percentage-point difference, than the 15-percentage-point difference observed in Experiment 1. Interestingly, when considering both the mental account and purchase medium conditions, a mirroring effect becomes apparent, particularly in the traditional purchase medium condition. In the ticket condition, the majority of the participants (67%) of participants indicated they would not buy a new ticket, while in the bill condition, a similar percentage (63%) would choose to buy a theater ticket. These results echo those of Kahneman and Tversky [21] and further support the concept of mental accounts' topical organization greatly impacting the subjective value of monetary losses. Specifically, theater tickets are intertwined with the theater experience, while the lack of association between the bills and the theater experience weakens the impact of the monetary loss.

Referring to the purchase medium (Hypothesis 2), the increased willingness to pay for a ticket in the modern condition was mild (a 5-percentage-point increase in the ticket condition and a 12-percentage-point increase in the bill

condition), and the binary logistic regression analysis indicated that the purchase medium was not significantly related to a difference in terms of willingness to pay for a ticket. Maity and Dass [30] showed that the cognitive costs and the extent of richness contained in the purchase mediums were the actual indicators of previously demonstrated consumer preference associated with modern purchase media rather than the modern purchase medium itself (i.e., online booking). The lack of a difference in information richness between the traditional and modern purchase scenarios may explain the non-significant impact of the purchase medium on the willingness to pay for a theater ticket in our study.

Including the maximum price participants were willing to pay as a variable in this study was essential to examine the role of price in their willingness to pay for a ticket, addressing the absence of replicated results for effect size in Experiment 1. Indeed, the results suggest that individuals who expressed a willingness to pay a minimum of 21€ for a theater ticket in the real-world were keener to purchase a ticket after experiencing a ticket or bill loss compared to those who would spend less than 21€. These results suggest that the personal value attributed to an experience in terms of monetary price has a significant role in purchase intentions. They align with previous cognitive studies demonstrating that the subjective prices attributed to experiences are significant predictors of purchase intentions [27].

Experiment 2 emphasizes the importance of adapting prices in experiments to contemporary context for replicating psychological effects involving money. A within-subjects manipulation was chosen to minimize the variability resulting from inter-individual variance for the mental accounting effect. Still, for methodological reasons it is necessary to investigate further whether the results can be replicated in a between-subjects experiment.

4. EXPERIMENT 3

Experiment 3 was conducted to check if the results of Experiment 2 were not artifacts of a within-subjects design (e.g., participants contrasting the conditions artificially – demand effect) but rather reflected the importance of adapting the experiment to contemporary times.

4.1 Method

4.1.1 Participants

The recruitment occurred via social media platforms such as Reddit (82) and Facebook (11); AoIR mailing list (99), SurveyCircle (26) and the University of Konstanz SONA platform (147).

Akin to Experiments 1 and 2, in Experiment 3 we employed the seriousness check [1, 40] as an inclusion criterion of datasets of participants for data analysis. Of 437 datasets from participants who indicated their intention to seriously participate in the experiment, a final sample of 365 datasets was used for analyses. Twenty cases were not included in the analysis due to multiple submissions with the same IP; only the first submission belonging to such IP addresses was included in the data analyses [42].

Additionally, 52 datasets were not included because of missing items that referred to the main questions (one question investigating the mental accounting principle, and one question about the maximum price that participants were willing to pay in the real world for a theater ticket).

The reported age of the participants ($N = 365$; 256 females) showed a median of 25 years ($M = 30$, $SD = 12.6$, range 14–75). Most participants reported being from Germany (181), other European countries (84), or from the United Kingdom (39), and were students (184).

4.1.2 Design and Procedure

The main change from Experiment 1 was the inflation-adjusted adaptation of the ticket price to current value. In WEXTOR [44], we implemented a 2×2 between-subjects factorial design varying the mental account condition (same – ticket loss versus different – bill loss) and the purchase medium condition (traditional – box office versus modern – online booking), as in Experiment 1.

Participants were randomly assigned to one of the four conditions. They were each shown one scenario where they were asked to imagine that they had just lost 40€ either as a theater ticket (same mental account) or as a bill (different mental account) and where the purchase was either made at the box office (traditional purchase medium) or via online booking (modern purchase medium). As in previous experiments, the measure was a yes-no choice. It is shown in Figure 3.

The remote experiment took approximately 2 minutes. Referring to the experiment procedure, everything was done as in Experiment 2.

Experiment 3 was conducted from May 2022 to February 2023.

4.1.3 Data Analyses

Akin to Experiments 1 and 2, and in line with the original study [21], in Experiment 3 we determined the participants' willingness to pay, and used a contingency table to replicate the basic structure of their analysis, which provided an overview of participants' responses across conditions. Similar to Experiments 1 and 2, we report the response time in different conditions.

Binary logistic regression was conducted as in Experiments 1 and 2. As in Experiment 2, the binary logistic regression helped to explore whether the participants' maximum price willing to pay for a ticket played a role in their willingness to pay (Appendix B).

4.2 Results

4.2.1 Main Analyses

Akin to Experiments 1 and 2, we calculated participants' willingness to pay under different conditions to test Hypotheses 1 and 2.

First, we focused on the mental account condition. We observed that in the ticket (same mental account) condition, a lower proportion of individuals (33%) expressed willingness to pay for a theater ticket than those who

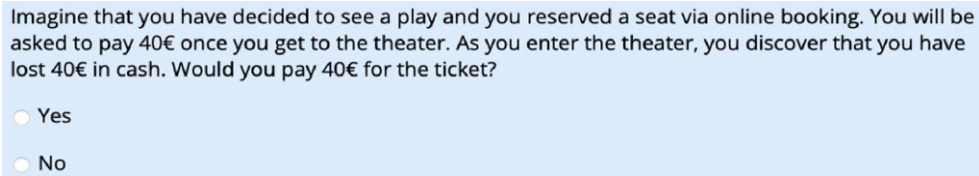


Figure 3. Example of a scenario web page from Experiment 3. The example shows one of the experimental scenarios (different mental account and modern purchase medium) and the options that participants could select to report their willingness to pay.

Table III. Number of participants willing and unwilling to pay for (another) theater ticket.

		Ticket loss		Bill loss	
		Yes	No	Yes	No
Traditional	Count	35	59	42	56
	% of condition	37	63	43	57
Modern	Count	26	63	63	21
	% of condition	29	71	75	25

Note. $N = 365$. Frequencies and proportions of participants answering *yes* or *no* to paying for a theater ticket, depending on the mental account (between-subjects, ticket versus bill), and the purchase medium (traditional versus modern) conditions they were randomly assigned to.

were unwilling to pay (67%). Conversely, in the bill (different mental account) condition, a higher proportion of individuals (58%) indicated their willingness to pay for a theater ticket in contrast to those who reported to be unwilling to pay (42%).

Second, Table III depicts the count and proportion of participants who reported their willingness to pay (*yes* versus *no*) based on the mental account conditions (columns) and the purchase medium conditions (rows).

Participants took 2 minutes on average (i.e., 113'') to complete the experiment (Min: 28''; Max: 575'').⁹ No noteworthy difference in average session length was observed among different between-subjects conditions (traditional x ticket: 112''; traditional x bill: 99''; modern x ticket: 128''; modern x bill: 115''). Interestingly, the average session length of participants answering *yes* in the modern x ticket condition was relatively slower (156'') than that of participants answering *yes* in the modern x bill condition (116''), as shown in detail in Appendix B.

4.2.2 Binary Logistic Regression (Exploratory Analysis)

We fitted a logistic model (estimated using maximum likelihood) to predict the willingness to pay for a ticket. The model included the variables mental account, purchase medium and the maximum price willing to pay. The model's explanatory power was moderate (Tjur's $R^2 = 0.17$). The model's intercept, corresponding to mental account = bill,

purchase medium = traditional, and maximum price willing to pay = 1-10€, was -0.95 with 95% CI $[-2.03, -0.01]$, $p = 0.06$. Within this model, the effect of mental account [ticket] was statistically significant and negative, $\beta = -1.12$, 95% CI $[-1.60, -0.66]$, $p < 0.001$. Participants were significantly less likely to exhibit a willingness to pay in the ticket condition compared to the bill condition. Furthermore, the impact of the purchase medium [modern] was statistically significant and positive, $\beta = 0.56$, 95% CI $[0.11, 1.03]$, $p = 0.02$. It was observed that the maximum price participants were willing to pay had a significant and positive impact on participants' likelihood of willingness to pay for a ticket when participants reported they would be willing to pay more than 31€ for a theater ticket. Participants who reported a willingness to spend more than 31€ for a theater ticket in real-world were likely keen on paying for a ticket even after experiencing a loss of 40€ (either as a ticket or bill) compared to those who expressed they would spend less than 31€ for a theater ticket in the real world scenario (Appendix B).

4.3 Discussion

The results of Experiment 3 shed light on the impact of mental account conditions on the willingness to buy a ticket in specific conditions.

When focusing solely on the mental account condition, we observed a notable difference in the willingness to pay for a ticket between the ticket loss condition (33%) and the bill loss condition (58%). This 25-percentage-point difference, while notable, aligns less closely with the original findings of Kahneman and Tversky [21], who reported a 42-percentage-point difference, than does the difference observed in Experiment 2. However, when the mental account and purchase medium conditions are considered together, it is evident that the replication of the original results is more pronounced in the modern purchase medium condition, with a large difference between the ticket loss condition, with a large difference between the ticket loss condition (29%) and the bill loss condition (75%). By contrast, in the traditional purchase medium condition, the difference is much smaller, with 37% willing to pay for a ticket in the ticket loss condition and 43% in the bill loss condition.¹⁰ We observe that individuals tend to express a greater willingness to pay for a ticket following a 40€ loss

⁹ Based on the pre-test, participants who took more than 10 minutes were not included in this calculation. Therefore, data from 17 participants were not included in this specific calculation.

¹⁰ In Appendix B, we show that the small effect may be attributed to age differences and the preference for theater average score. However, further research is needed to clarify these results.

represented as bills, while this pattern reverses in the context of the ticket condition. This is corroborated by the results of the binary logistic regression, which suggest that participants are significantly less willing to pay in the ticket condition compared to the bill condition. The results of Experiment 3 confirm Hypothesis 1.

In reference to the purchase medium condition, the binary logistic regression indicated that the modern purchase medium contributed significantly to the willingness to pay for a ticket, in line with Hypothesis 2. Given the lack of similar results in the previous experiments, this observation needs to be interpreted with caution. The number of participants willing to pay for a ticket in the modern conditions suggest an interaction of bill \times modern purchase medium, but conducting interaction analyses based on a binary logistic regression is not recommended. Therefore, further replications should include the possibility of an interaction between mental account and purchase medium. To pursue this goal, future conceptual replications should employ continuous (versus yes-no) measurements (e.g., a visual analogue scale - VAS) to assess the mental accounting of money effect with increased precision.

As previously explained for Experiment 2, including the maximum price participants were willing to pay as a variable in this study was necessary to gain deeper insights into the lack of a replicated mental accounting effect for effect size in Experiment 1. The results of Experiment 3 suggest that participants who report a maximum price they were willing to pay higher than 31€ for a theater ticket were significantly more likely to buy a ticket, after experiencing a 40€ loss. The results may be attributed to individual preferences for specific activities, which are intrinsically linked to the assessment of benefits associated with a product or service - a concept known as value consciousness [14]. This value consciousness and the subjective prices associated with specific experiences are predictors of purchase intentions [27].

Experiment 3 revealed that mental accounting could be replicated when using a between-subjects design. Additionally, it shows the importance of adapting the experimental scenarios to the contemporary context refers not only to adjusting the monetary values, but also to other aspects of change in history (e.g., modern purchase media).

5. GENERAL DISCUSSION

The current study tested whether adapting the experimental context to the current historical context (i.e., monetary value, modern purchase medium) increased the likelihood of replicating a cognitive effect with a similar effect size to the original study. We described the magnitude of the effect observed in the main analyses, which we assessed by comparing the differences in key outcomes between conditions in our study and those reported in the original study. To provide empirical evidence, we conducted three remote experiments testing the mental accounting effect based on Kahneman and Tversky's theater ticket paradigm [21].

Replicating the experiments using Internet-based methodologies was a step forward compared with the

original experiment conducted in a laboratory setting. We observed that conducting the experiments remotely was advantageous because it allowed us to reach a large and heterogeneous sample [39], and create easily reproducible materials [38], while adapting the replication paradigms to contemporary contexts and methodologies. These adaptations help ensure that replication studies remain relevant and robust, addressing the evolving demands of scientific inquiry [35]. Due to the ease of sharing links to remote experiments, using various recruitment platforms or services that allow pre-screening parameters (e.g., Prolific) can significantly enhance the heterogeneity of samples compared to traditional laboratory settings. The samples recruited in the three experiments had similar median ages—25, 20, and 25 for Experiments 1, 2, and 3, respectively, but were characterized by relatively broad age ranges, which contributes to the generalizability of the findings. In this study, we intentionally utilized multiple recruitment platforms to increase sample diversity and be able to investigate robustness across samples via this multiple site entry technique [39]. Notably, in Experiments 2 and 3, where we collected additional demographic information such as occupation status and country of residence, we demonstrated the feasibility of reaching participants from varied backgrounds. Unlike laboratory-based experiments, which often rely on participants affiliated with the institution or require time-consuming coordination with external labs, conducting these experiments remotely allowed us to recruit participants from four platforms in Experiment 2, five platforms in Experiment 3, and eight platforms in Experiment 1, which is considered beneficial to increase the sample's diversity.

In response to Feest's concerns about internal validity [11]—i.e., the risk of overlooked confounding variables and misidentification of relevant causal features—we found that combining a process-oriented approach with remote replications offered several advantages. The process-oriented approach allowed us to identify the monetary value of the ticket price and monetary loss as an influential factor in the tested effect. Additionally, conducting remote replications facilitated the online sharing of easily reproducible experimental materials, offering future researchers the opportunity to further investigate the effects or explore potential additional factors more easily. Thus, remote replications support greater transparency and broader scrutiny of experimental materials.

Literature has shown that direct comparisons between data collected in laboratory and Internet-based settings yield similar results for replicating and testing cognitive effects [51], even for time-sensitive tasks, if best practices are being followed [13]. Additionally, researchers observed that results and variances from remote experiments are comparable to those obtained from laboratory-based studies [49]. The abovementioned advantages of remote experiments, coupled with the high data quality provided by rigorous online methodologies, support the argument that exact replications do not need to be confined to laboratory settings when replicating original studies conducted in such environments. Indeed, exact replication should maintain the main

components (same experimental designs, main statistical analyses, and key stimuli) of the original experiments [18], rather than striving for a (practically unattainable) full replication of the original experiment's conditions. Our study contributes to the body of knowledge by showing that the mental accounting effect is replicable using Internet-based research methodologies.

Our step-by-step approach in developing replication studies of relatively older theories led us to focus on minor adjustments when conducting the conceptual replications. This prevented premature conclusions about the inability to replicate the cognitive effect with a similar effect size to the original results in Experiment 1. Experiment 2, utilizing a within-subjects design and adjusting the ticket prices to contemporary standards, successfully reproduced the original results with a similar effect size to the original study. Similarly, in Experiment 3, adapting the price of the theater ticket to contemporary times but adopting a between-subjects design, resulted in a general replication of the original findings. Results from Experiment 3 suggest that adapting the purchase medium to the current times may play a role; however, these findings should be interpreted with caution, as they were observed only in this experiment.

The results of the three experiments align with prior research, indicating that the cognitive processing of mismatch among real-life prices and prices in the experiment might alter participants' decisions [33]. Consistent with Hypothesis 3, the lack of replicated results with a similar effect size to the original study in Experiment 1, and the outcomes of Experiments 2 and 3 suggest that for replication studies of relatively older experiments involving money, it is recommended to consider factors such as inflation and the role of cognitive processing mismatch between the real world prices and prices used in the experiments [33].

The mental accounting of money effect was replicated (Hypothesis 1), in line with the original study from Kahneman and Tversky [21]. However, further studies are needed to clarify whether and how specific prices affect this cognitive effect. The extent of sensitivity to the context impacting replication studies [58], or the mismatch among real-world prices and the prices considered in the experiment [27] may impact the effect size of the replicated cognitive effects. We observed that participants who associate higher perceived prices with the theater (i.e., they attribute an increased monetary value to the experience) are more likely to buy a ticket, independently from the experimental conditions. This aspect was not part of the experimental manipulation, and it was not part of the original experiment, but it might be inspiring for future research. The price attributed to the theater ticket (and the monetary loss) is highly relevant, and studying in depth the relationship between monetary prices and cognitive effects is valuable for research in cognitive psychology. This includes examining the cognitive effect's influence on price perceptions [4] and the potential impact of price on cognitive

effects [26]. Considering participants' attitudes toward ticket prices and current theater pricing trends, we adjusted the ticket price to 40€ in the conceptual replications. However, this adjustment resulted in participants losing multiple bills in the bill conditions of the two conceptual replications, as no 40€ bill exists. Although our results do not indicate a significant impact on participants' decisions, future research should investigate this confound. Referring to the importance of tailoring paradigms to specific contexts, future studies could explore adapting the currency used in experiments to participants' geographic locations to evaluate whether any differences emerge. In this study, we prioritized increased generalizability and sample heterogeneity by using the same currency across all experiments, regardless of participants' locations. This decision ensured comparability across conditions, particularly in relation to the cognitive effects we were testing, which relied on the relative comparison of values (e.g., 40€) rather than on specific purchasing power. Moreover, we controlled for participants' maximum willingness to pay for a ticket in our binary logistic regression analysis to mitigate concerns about individual differences in price perception. Future process-oriented replications that focus on cultural differences and variations in purchasing power could further investigate these aspects.

We observed that the purchase medium did not have a significant effect on participants' willingness to pay for a ticket, and thus, Hypothesis 2 was not confirmed. We propose that, despite the recognized influence of modern purchase mediums on purchase intentions [10], it is the richness of the purchase medium's description [30] that significantly impacts purchase intentions. Thus, only changing the purchase medium was not sufficient to observe significant differences in the purchase intentions.

We conclude that the mental accounting effect is replicable when appropriately adapted to contemporary contexts and circumstances. Moreover, we emphasize the importance of process-oriented replication projects to better understand the fundamental components of experimental paradigms used to test cognitive theories, such as mental accounting. Notably, Internet-based methodologies offer significant advantages for replication studies, including the ability to easily share experimental materials and access large samples with ease. The primary objective of replication studies should be to evaluate the robustness of the cognitive effect under investigation, which can be achieved through rigorous methodologies and the inclusion and comparison of several different participant samples.

SUPPLEMENTARY MATERIALS

Supplementary materials, including Appendices A and B, are available online at: <https://osf.io/s3a8b/>.

ACKNOWLEDGMENT

The authors express their gratitude to Andrey Andreev and Isabel Helm for their support in conducting the experiments as student assistants, and to Annika-Tave Overlander and Patrick Slayer for proofreading this work.

REFERENCES

- 1 F. Aust, B. Diedenhofen, S. Ullrich, and J. Musch, "Seriousness checks are useful to improve data validity in online research," *Behav. Res. Methods*, **45**, 527–535 (2013).
- 2 M. H. Birnbaum, "Introduction to psychological experiments on the Internet," in *Psychological Experiments on the Internet*, edited by M. H. Birnbaum (Academic Press, San Diego, 2000), pp. XV–XX.
- 3 M. H. Birnbaum, "Human research and data collection via the Internet," *Annu. Rev. Psychol.* **55**, 803–832 (2004).
- 4 M. C. Campbell, "'Says Who?'" How the source of price information and affect influence perceived price (un)fairness," *J. Mark. Res.* **44**, 261–271 (2007).
- 5 J. J. Chandler and G. Paolacci, "Lie for a dime: When most prescreening responses are honest but most study participants are impostors," *Soc. Psychol. Personal. Sci.* **8**, 500–508 (2017).
- 6 G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design," *J. Econ. Behav. Organ.* **81**, 1–8 (2012).
- 7 N. Coles, P. S. Forscher, J. K. Flake, L. DeBruine, and B. Jones, "Promises and challenges of Big Team psychology," Springer Nat. Res. Commun. (2022) Retrieved from <http://socialsciences.nature.com/posts/promises-and-challenges-of-big-team-psychology>, last accessed 2024/10/30.
- 8 CPI Inflation Calculator, https://www.bls.gov/data/inflation_calculator.htm, last accessed 2024/10/10.
- 9 F. Del Popolo Cristaldi, U. Granzio, I. Bariletti, and G. Mento, "Doing experimental psychological research from remote: How alerting differently impacts online versus lab setting," *Brain Sci.* **12**, 1061 (2022).
- 10 S. Devaraj, M. Fan, and R. Kohli, "Examination of online channel preference: Using the structure-conduct-outcome framework," *Decis. Support Syst.* **42**, 1089–1103 (2006).
- 11 U. Feest, "Why replication is overrated," *Philos. Sci.* **86**, 895–905 (2019).
- 12 S. Fullerton and T. McCullough, "Using quality control checks to overcome pitfalls in the collection of primary data via online platforms," *J. Mark. Anal.* **11**, 602–612 (2023).
- 13 P. Garaizar and U.-D. Reips, "Best practices: Two web browser-based methods for stimulus presentation in behavioral experiments with high resolution timing requirements," *Behav. Res. Methods*, **51**, 1441–1453 (2019).
- 14 T. Girard, P. Trapp, M. Pinar, T. Gulsoy, and T. E. Boyt, "Consumer-based brand equity of a private-label brand: Measuring and examining determinants," *J. Mark. Theory Pract.* **25**, 39–56 (2017).
- 15 A. Glöckner, M. Jekel, R. A. Torras, A. R. Dorough, C. Anderl, N. Franke, D. Mischkowski, S. Fiedler, S. Miketta, and J. Goltermann, Hagen Cumulative Science Project I (2023, July 17). Retrieved from <https://osf.io/d7za8>.
- 16 H. Honing and U.-D. Reips, "Web-based versus Lab-based studies: A response to Kendall," *Empirical Musicol. Rev.* **3**, 73–77 (2008).
- 17 B. Huber and K. Z. Gajos, "Conducting online virtual environment experiments with uncompensated, unsupervised samples," *PLOS ONE* **15**, e0227629 (2020).
- 18 R. Hudson, "Explicating exact versus conceptual replication," *Erkenntnis*, **88**, 2493–2514 (2023).
- 19 J. Hüffmeier, J. Mazei, and T. Schultze, "Reconceptualizing replication as a sequence of different studies: A replication typology," *J. Exp. Soc. Psychol.* **66**, 81–92 (2016).
- 20 E. Jun, G. Hsieh, and K. Reinecke, "Types of motivation affect study selection, attention, and dropouts in online experiments," *Proc ACM Hum-Comput Interact., Volume 1, Issue CSCW* (Association for Computing Machinery, New York, NY, USA, 2017), Article No.: 56, 1–15.
- 21 D. Kahneman and A. Tversky, "Choices, values, and frames," *Am. Psychol.* **39**, 341–350 (1984).
- 22 E. Kaufmann and U.-D. Reips, "Meta-analysis in a digitalized world: A step-by-step primer," *Behav. Res. Methods*, **56**, 1–21 (2024).
- 23 D. A. Kenny and C. M. Judd, "The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication," *Psychol. Methods*, **24**, 578–589 (2019).
- 24 S. L. Koole and D. Lakens, "Rewarding replications: A sure and simple way to improve psychological science," *Perspect. Psychol. Sci.* **7**, 608–614 (2012).
- 25 M. Korbacher, F. Azevedo, C. R. Pennington, H. Hartmann, M. Pownall, K. Schmidt, M. Elsherif, N. Breznau, O. Robertson, T. Kalandadze, *et al.* "The replication crisis has led to positive structural, procedural, and community changes," *Commun. Psychol.* **1**, 1–13 (2023).
- 26 A.-V. Lehtimäki, K. B. Monroe, and O. Somervuori, "The influence of regular price level (low, medium, or high) and framing of discount (monetary or percentage) on perceived attractiveness of discount amount," *J. Revenue Pricing Manag.* **18**, 76–85 (2019).
- 27 G. R. D. Levrini and M. Jeffman dos Santos, "The influence of price on purchase intentions: Comparative study between cognitive, sensory, and neurophysiological experiments," *Behav. Sci.* **11**, 16 (2021).
- 28 J. G. Lynch Jr., E. T. Bradlow, J. C. Huber, and D. R. Lehmann, "Reflections on the replication corner: In praise of conceptual replications," *Int. J. Res. Mark.* **32**, 333–342 (2015).
- 29 E. Machery, "What is a replication?," *Philos. Sci.* **87**, 545–567 (2020).
- 30 M. Maity and M. Dass, "Consumer decision-making across modern and traditional channels: E-commerce, m-commerce, in-store," *Decis. Support Syst.* **61**, 34–46 (2014).
- 31 H. Moshontz, L. Campbell, C. R. Ebersole, H. Ijzerman, H. L. Urry, P. S. Forscher, J. E. Grahe, R. J. McCarthy, E. D. Musser, J. Antfolk, *et al.* "The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network," *Adv. Methods Pract. Psychol. Sci.* **1**, 501–515 (2018).
- 32 J. Musch and U.-D. Reips, "A brief history of Web experimenting," in *Psychological Experiments on the Internet*, edited by M. H. Birnbaum (Academic Press, San Diego, 2000), pp. 61–87.
- 33 X. Niu and N. Harvey, "Are lay expectations of inflation based on recall of specific prices? If so, how and under what conditions?," *J. Econ. Psychol.* **98**, 102662 (2023).
- 34 E. Peer, L. Brandimarte, S. Samat, and A. Acquisti, "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *J. Exp. Soc. Psychol.* **70**, 153–163 (2017).
- 35 M.-M. Pittelkow, S. M. Field, P. M. Isager, A. E. van't Veer, T. Anderson, S. N. Cole, T. Dominik, R. Giner-Sorolla, S. Gok, T. Heyman, *et al.* "The process of replication target selection in psychology: What to consider?," *R. Soc. Open Sci.* **10**, 210586 (2023).
- 36 R: The R Project for Statistical Computing, <https://www.r-project.org/>, last accessed 2024/10/31.
- 37 U.-D. Reips, "Das psychologische Experimentieren im Internet (Psychological experimenting on the Internet)," in *Internet für Psychologen*, edited by B. Batinic (Hogrefe, Göttingen, 1997), pp. 245–265.
- 38 U.-D. Reips, "The Web experiment method: Advantages, disadvantages, and solutions," in *Psychological Experiments on the Internet*, edited by M. H. Birnbaum (Academic Press, San Diego, CA, 2000), pp. 89–118.
- 39 U.-D. Reips, "Standards for Internet-based experimenting," *Exp. Psychol.* **49**, 243–256 (2002).
- 40 U.-D. Reips, "Internet experiments: Methods, guidelines, metadata," *Human Vision and Electronic Imaging XIV* **7240**, 724008 (2009).
- 41 U.-D. Reips, "Web-based research in psychology," *Z. Für Psychol.* **229**, 198–213 (2021).
- 42 U.-D. Reips and M. H. Birnbaum, "Behavioral research and data collection via the Internet," in *The Handbook of Human Factors in Web Design*, edited by K.-P. L. Vu and R. W. Proctor (CRC Press, Mahwah, New Jersey, Erlbaum, 2011), pp. 563–585.
- 43 U.-D. Reips, T. Buchanan, J. Krantz, and K. McGraw, "Methodological challenges in the use of the Internet for scientific research: Ten solutions and recommendations," *Stud. Psychol. Theor. Pract.* **15**, 139–148 (2015).
- 44 U.-D. Reips, T. Blumer, J. Caffier, C. Neuhaus, and J. Simson, WEXTOR, <https://wextor.eu>, last accessed 2025/01/23.
- 45 S. J. Ritchie, R. Wiseman, and C. C. French, "Failing the future: Three unsuccessful attempts to replicate Bem's 'Retrospective facilitation of recall' effect," *PLOS ONE* **7**, e33423 (2012).
- 46 J. M. Rodd, "Moving experimental psychology online: How to obtain high quality data when we can't see our participants," *J. Mem. Lang.* **134**, 104472 (2024).
- 47 H. L. Roediger III, "Psychology's woes and a partial cure: The value of replication," *APS Obs* **25**, 12–20 (2012).
- 48 M. Sauter, D. Draschkow, and W. Mack, "Building, hosting and recruiting: A brief introduction to running behavioral experiments online," *Brain Sci.* **10**, 251 (2020).

- ⁴⁹ M. Sauter, M. Stefani, and W. Mack, "Equal quality for online and lab data: A direct comparison from two dual-task paradigms," *Open Psychol.* **4**, 47–59 (2022).
- ⁵⁰ S. Schmidt, "Shall we really do it again? The powerful concept of replication is neglected in the social sciences," *Rev. Gen. Psychol.* **13**, 90–100 (2009).
- ⁵¹ K. Semmelmann and S. Weigelt, "Online psychophysics: Reaction time effects in cognitive experiments," *Behav. Res. Methods.* **49**, 1241–1260 (2017).
- ⁵² D. J. Simons, "The value of direct replication," *Perspect. Psychol. Sci.* **9**, 76–80 (2014).
- ⁵³ S. Sperandei, "Understanding logistic regression analysis," *Biochem. Medica.* **24**, 12–18 (2014).
- ⁵⁴ A. Steckler and K. R. McLeroy, "The importance of external validity," *Am. J. Public Health.* **98**, 9–10 (2008).
- ⁵⁵ W. Stroebe and F. Strack, "The alleged crisis and the illusion of exact replication," *Perspect. Psychol. Sci.* **9**, 59–71 (2014).
- ⁵⁶ R. H. Thaler, "Mental accounting matters," *J. Behav. Decis. Mak.* **12**, 183–206 (1999).
- ⁵⁷ R. Ulrich and J. Miller, "Questionable research practices may have little effect on replicability," *eLife* **9**, e58237 (2020).
- ⁵⁸ J. J. Van Bavel, P. Mende-Siedlecki, W. J. Brady, and D. A. Reinero, "Contextual sensitivity in scientific reproducibility," *Proc. Natl. Acad. Sci.* **113**, 6454–6459 (2016).
- ⁵⁹ G. Vogel, "Psychologist accused of fraud on 'Astonishing scale'," *Science* **334**, 579–579 (2011).
- ⁶⁰ E.-J. Wagenmakers, R. Wetzels, D. Borsboom, and H. L. J. van der Maas, "Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem," *J. Pers. Soc. Psychol.* **100**, 426–432 (2011).
- ⁶¹ J. R. Wagge, C. Baciú, K. Banas, J. T. Nadler, S. Schwarz, Y. Weisberg, H. Ijzerman, N. Legate, and J. Grahe, "A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect," *Collabra Psychol.* **5**, 5 (2019).
- ⁶² R. A. Zwaan, A. Etz, R. E. Lucas, and M. B. Donnellan, "Making replication mainstream," *Behav. Brain Sci.* **41**, e120 (2018).