INTERNATIONAL REVIEW OF SOCIAL PSYCHOLOGY

Sunk Cost Effects for Time Versus Money: Replication and Extensions Registered Report of Soman (2001)

PEER COMMUNITY IN REGISTERED REPORT

ubiquity press

NIKOLAY B. PETROV
YIN KAN MEGAN CHAN
CHEUK NAM LAU
TIN HO KWOK
LOK CHING ESTELLE CHOW
WAI YAN LO
WENKAI SONG
GILAD FELDMAN

*Author affiliations can be found in the back matter of this article

ABSTRACT

The sunk cost effect is the tendency for an individual's decision making to be impacted by unrecoverable previous investments of resources. Soman (2001) found that sunk cost effect is weaker for time than for money (Studies 1 and 2) and that the facilitation of money-like accounting strengthens the sunk cost effect for time (Study 5). We conducted a Registered Report of a close, high-powered replication and extension of Soman's (2001) Studies 1 and 2 and a conceptual replication of his Study 5 with an online sample of US American Amazon Mechanical Turk (N = 821). We found support for differences between sunk money costs and sunk time costs in Study 1 (original: φ_c = .61 [.43, .78]; replication: φ_c = .38 [.31, .45]), yet not in Study 2, in which we found sunk cost effects for both money and time (original: money – φ_c = .32 [.12, .52], time – φ_c = .02 [.00, .18]; replication: money – φ_c = .23 [.14, .33] , time – φ_c = .32 [.23, .42]). In Study 5, we found no support for facilitation of money-like accounting as strengthening the sunk time cost effect. Materials, data, and code are available on: https://osf.io/pm264/.

CORRESPONDING AUTHOR:
Gilad Feldman
Department of Psychology, University of Hong Kong, Hong Kong SAR, HK
gfeldman@hku.hk

KEYWORDS:
judgment and decision-making; sunk cost; escalation of commitment; time; money; opportunity cost

TO CITE THIS ARTICLE:
Petrov, N. B., Chan, Y. K. M., Lau, C. N., Kwok, T. H., Chow, L. C. E., Lo, W. Y., Song, W., & Feldman, G. (2023). Sunk Cost Effects for Time Versus Money: Replication and Extensions Registered Report of Soman (2001). *International Review of Social Psychology*, 36(1): 17, 1–18. DOI: https://doi.org/10.5334/irsp.883

PCIRR-STUDY DESIGN TABLE

QUESTION	HYPOTHESIS	SAMPLING PLAN	ANALYSIS PLAN	RATIONALE FOR TEST	INTERPRETATION GIVEN DIFFERENT OUTCOMES	THEORY THAT COULD BE SHOWN WRONG BY THE OUTCOMES
Is the sunk cost effect weaker for time than for money?	The sunk cost effect is weaker for time than for money.	Participants recruited online using the US American Amazon platform.	Chi-square test	We follow the statistical methods of the original paper.	Based on the criteria used by LeBel et al. (2018) we will examine the replicability of the findings of Soman (2001).	The sunk cost effect is weaker for time than for money and the facilitation of money-like accounting for sunk time costs strengthens the sunk cost effect.
Does the facilitation of money-like accounting for sunk time costs strengthen the sunk time cost effect?	Facilitation of money-like accounting by using education about economic approaches to time strengthens the sunk cost effect of time		Two-way between-subject ANOVA			

People will often increase time and money investments in a failing course of action to try and recover or justify an initial investment, leading to an escalating commitment to a losing course of action. This phenomenon has been coined the 'sunk cost effect' (Arkes & Blumer, 1985; Thaler, 1980), given that with larger sunk costs there are stronger tendencies to further escalate.

The sunk cost effect has mostly been investigated with the invested resources being either money or time (or both, e.g., Pandey & Sharma, 2019). In the money domain, findings have been largely consistent and in support of sunk money effects (Arkes & Blumer, 1985; Bornstein et al., 1999; Coleman, 2009; Navarro & Fantino, 2005; Soman & Cheema, 2001), though there were several failures, such as that of Friedman et al. (2007).

Compared to sunk money costs, sunk time costs seem more volatile. For example, Navarro and Fantino (2009) found that undergraduate students were susceptible to sunk time effects across various factors, including the difficulty of and enjoyment from the future time investment and personal responsibility. Silva Castillo et al. (2020) also found evidence for sunk time costs in a within-subject study of 46 undergraduate students, also showing that there is a positive linear relationship between the time investment and the subjective value placed on the outcome. Bornstein and Chapman (1995) similarly found evidence of sunk time costs, with the presence and strength of these differences being affected by other factors, such as who the decision maker is in the scenario and how carefully the decision is considered. In comparing money and time for sunk costs, Park and Jang (2014) found that among people from the general population both sunk time and sunk money costs independently predicted intentions to cancel a future hypothetical trip. In a similar vein, Pandey and Sharma (2019), across three vignette experiments, found that found that graduate students were susceptible to sunk time costs both in scenarios when they spend only time and when they spend money and time both, but money can be recovered. However, in this study, the sunk time cost effect only appeared when the time investment exceeded a specific threshold, which raises the question of what other factors affect the different expressions of the sunk money and sunk time effects.

Some research already points to potential candidates that distinguish between sunk money and sunk time effects. For instance, across online and field studies, Soster et al. (2010) showed that the sunk money and sunk time effects are equivalent if the accounting period is the same, but asymmetrical if the accounting periods are different. Another example comes from Okada and Hoch (2004), who showed that both risk aversion and ambiguity in the outcome produce differences in how time and money costs are accounted for.

Another factor that might differentially affect sunk money and sunk time costs is age. Strough et al. (2008) showed that younger adults are less likely to be susceptible to sunk money costs. One way to contextualise this finding is to consider that sunk costs are not taken in their absolute values, but relative to an individual's total available resource (Garland & Newport, 1991), and older adults are generally wealthier and have less available time, compared to younger adults. Another contextualising factor for the age effect is that experience accounting for both time and money has been shown to predict susceptibility to sunk costs, and younger people likely have much less experience (Bornstein et al., 1999; DeVoe & Pfeffer, 2007; Ronayne et al., 2021).

Methods-wise, Rego et al. (2018) showed that although participants are more likely to stay in an unhealthy relationship when money, but not time, was invested, the effect of sunk time costs was stronger when the outcome was measured on a continuum scale (amount of time willing to invest in an unhappy relationship) rather than as a binary choice (whether or not to invest time).

Overall, although these studies hint at some factors that might affect sunk money and sunk time effects differently, the underlying reasons for these differences remain unclear.

To address this question, Soman (2001) focused on three reasons that make accounting for the sunk costs of time more difficult compared to those of money: 1) time cannot be inventoried or replaced, 2) time is not as easily aggregated as money, 3) accounting for money, unlike time, is a routine activity. In this seminal work, Soman (2001) asked participants, across several experiments, to read

scenarios that only differed in whether they were related to time or money and whether there were any sunk costs to be accounted for. Soman's (2001) core finding was that the strength of the sunk cost effect was weaker for time than for money. He further showed that the facilitation of money-like accounting for sunk time costs by highlighting opportunity costs or by educating about an economic approach to time strengthens the sunk time cost effect.

The ubiquity of sunk costs in everyday life and the impact of Soman's (2001) work (680 citations on Google Scholar as of November 2023) suggests the value of revisiting and expanding on this work. To the best of our knowledge, Soman's (2001) research has not been directly replicated.

We aimed to revisit the classic phenomenon and examine the reproducibility and replicability of the classic findings by replicating the studies and improving the design with extensions. Following the recent growing recognition of reproducibility and replicability in psychological science (Brandt et al., 2014; Open Science Collaboration, 2015; Nosek et al., 2022; Zwaan et al., 2018), we embarked on a well-powered pre-registered replication and extensions of Soman (2001).

We focused our replication on Studies 1 and 2 as they provided the baseline test of the core hypothesis to elucidate whether people account for both the magnitude (Study 1) and the presence (Study 2) of sunk costs in each domain. We also targeted Study 5 in a conceptual replication as it suggested a method for potential mitigation of the effect. We summarized the hypotheses and effects for Studies 1, 2, and 5 in Table 1.

STUDIES OVERVIEW: REPLICATIONS OF STUDIES 1, 2, AND 5

OPEN SCIENCE DECLARATION

This project was submitted as a Registered Report (Chambers & Tzavella, 2022; Nosek & Lakens, 2014; Scheel et al., 2021; Wiseman et al., 2019), and received Peer Community in Registered Report Stage 1 in-principle acceptance (https://osf.io/65htv/; https://rr.peercommunityin.org/articles/rec?id=187) after which we created a frozen pre-registration version of the entire Stage 1 packet (https://osf.io/78vgx/) and proceeded to data collection. We provided all materials, data, and code on: https://osf.io/pm264/ .

All measures, manipulations, and exclusions conducted for this investigation are reported, and data collection was completed before analyses.

We reported results after exclusions below, and in the supplementary materials, we detailed a comparison between pre- and post-exclusion findings as well as any deviations from the pre-registered plan ('Comparisons and deviations' subsection), with additional disclosures ('Open science disclosures' subsection).

PROCEDURE

We focused on Soman's (2001) Studies 1, 2, and 5. We combined the three studies into a unified single data collection. This allowed us to maximize our resources and had the added advantage that we can rule out any sample characteristics that might be driving differences in successful versus unsuccessful replications. Additionally, a single unified survey allowed us to conduct additional exploratory within-subjects analyses and explore links between different studies, something that is not possible with the design of the original. Given that the replication of Study 5 involved education about sunk time costs with a scenario that was first introduced in Study 1, we fixed the order so that Study 5 is always last, with randomized order for the replications of Studies 1 and 2.

Participants first provided consent, after which they read an outline for the studies and three questions confirmed participants qualifications as being American, their understanding of the study procedures, and their agreement to pay close attention (Yes/No/Not sure presented in random order, and participants not answering Yes were asked to return the task). Participants then completed three studies: first Studies 1 and 2 in randomized order, followed by Study 5. In each of the studies, participants read a hypothetical scenario presenting them with two alternatives. In all studies, participants indicated their choice between the two alternatives, and in Studies 1 and 5 they also indicated their preference between the two options on a Likert scale (see below). After Studies 1 and 2, they were asked comprehension checks question to check if they understood the critical information in the scenario and afterwards asked if they had seen the scenario before, and if so, where. After completing all studies, participants answered questions inquiring about their seriousness and familiarity with the materials, reported their experience during the survey, and provided demographic information (with no implications for participation or pay). Finally, participants were thanked and debriefed. Throughout the study, participants could not go back to previous pages. Our replication project received ethical approval from the University of Hong Kong (REF ID: EA220438).

A methodological comparison between the original and the current study on key dimensions can be found in Table 2.

MATERIALS

The descriptions of the stimuli in the target article were limited. We reached out to the author and received the materials used in the original, and we are very grateful for the author's support in making these available. We used the same content with the minor exception that we started each scenario with 'Imagine you are a student' to adjust to the different sample (undergraduates vs. general population, see Table 2). We made some minor

HYPOTHESES	STUDY	DESCRIPTION	STATISTICAL TEST	ORIGINAL OR REPLICATION	EFFECT SIZE[a] [95% CI]	REPLICATION OUTCOME[b]
Hypothesis 1: The sunk-cost effect is weaker in the domain of temporal costs than in the domain of monetary costs.	**1 (Theatre and concert tickets)**	Two types of tickets are expressed in two different types of sunk cost domains—either time or money to investigate the relative strength of each domain.	Chi-square; difference between sunk time and sunk money conditions in rate of choosing a ticket	Original	$\varphi_c = .61$ [.43, .78]	signal – inconsistent, smaller
				Replication	$\varphi_c = .38$ [.31, .45]	
	2 (Choosing a project)	The domain (time/money) and the existence of sunk cost (present/absent) are manipulated within a scenario, describing potential projects to work on to test the strength of the sunk cost effects across domains.	Chi-square; difference between sunk time and no sunk time conditions in rate of choosing a project	Original	$\varphi_c = .02$ [.00, .18]	signal – inconsistent, positive
				Replication	$\varphi_c = .32$ [.23, .42]	
			Chi-square; difference between sunk money and no sunk money conditions in rate of choosing a project	Original	$\varphi_c = .32$ [.12, .52]	signal – consistent
				Replication	$\varphi_c = .23$ [.14, .33]	
Hypothesis 2a: If the absence of a sunk time cost effect is due to difficulties associated with the accounting of time, then the facilitation of accounting should cause the effect to reappear. [Alternative hypothesis] **Hypothesis 2b:** If the absence of a sunk time cost effect is due to the fact that individuals behave rationally when evaluating past time investments, then the facilitation of accounting should not cause the effect to reappear. [Null hypothesis]	**5 (Education and opportunity costs)**	The level of opportunity cost (high/low) and education (present/absent) were manipulated to evaluate the strength of sunk cost effects.	ANOVA; opportunity cost main effect	Original	= .09 [.00, .23]	no signal – inconsistent
				Replication	= .00 [.00, .01]	
			ANOVA; education main effect	Original	= .17 [.04, .32]	no signal – inconsistent
				Replication	= .00 [.00, .01]	
			ANOVA; opportunity cost by education interaction	Original	= .00 [.00, .02]	no signal – consistent
				Replication	= .00 [.00, .01]	

Table 1 Soman (2001): Summary of studies and hypotheses and a comparison of original and replication effects.

[a]We provide additional detail regarding the calculation of effect sizes in the supplementary materials "Effect sizes calculation".

[b]We classified each effect using the criteria set out by LeBel et al. (2019).

	ORIGINAL	REPLICATION	REASON FOR CHANGE
Participants	Undergraduate students from Hong Kong University of Science and Technology and University of Colorado.	Participants from CloudResearch/ Amazon MTurk.	Larger more diverse sample. Addressing sample concerns and allowing for exploratory analyses comparing effects across studies.
	Study 1, 2 and 5 were done separately with different participants.	Study 1, 2 and 5 were done in the same survey with the same participant.	
Delivery	Paper questionnaires	Online questionnaire using Qualtrics	
Questions	The original studies did not use any comprehension checks or instructional manipulation checks.	We used comprehension and instructional manipulation checks in our replication.	To ensure that participants read and understood the materials.
Materials	In Study 5, a class on opportunity cost was delivered to those in the education condition.	A passage about opportunity cost along with questions about that passage as instructional manipulation checks were presented.	To adjust to an online sample, we used a passage that participants read instead of a class.
Scale	In Study 5 the preference scale was originally from 1 to 9 and presented as such.	We adjusted the presentation of the scale to 4/0/4 instead of 1 to 9.	Avoid biasing participants in a certain direction.
Order of studies	Study 1 –> Study 2 –> Study 5	Randomized the order of studies 1 and 2 only, but not study 5. Study 5 is presented last at the end of the experiment.	To address potential impact of presentation order.

Table 2 Original versus replication methodological comparison.

stylistic changes to the presentation of the materials (using bold/underline/italics at places). The survey used is available on the OSF, and a summary of the materials and questions used is provided in the supplementary materials ('Materials used' subsection).

POWER ANALYSIS

We used a 'small-telescope' approach in planning our sample size (Simonsohn, 2015). This approach allows us to both achieve the power to reject a zero-effect null hypothesis, assuming there is a true effect, and to detect an effect much smaller than the original could have possibly detected. To achieve this, it is recommended to use a replication sample 2.5 times that of the original. This is an especially powerful approach in conjunction with our implementation of the studies by combining them into a single survey as it means that powering the largest study entails giving even more power to the other ones. Thus, given that Soman (2001) used a sample size of 206 in his Study 2, we calculated a needed sample of at least 515 participants. However, we also wanted to test whether the order in which the studies was presented (Study 1 first vs. Study 2 first) affected the results, thus we doubled that sample and planned for a 15% planned exclusion rate, meaning we aimed to recruit 1212 participants in order to get a total of 1030 participants, with equal numbers completing Study 1 or Study 2 first.

We conducted a sensitivity analysis for both the 515 and 1030 target samples. We found that we had 99%+ power to detect the original smallest original effect sizes in each study and 80% power to detect effect sizes at least half of those of the original—see Table 3.

EXCLUSION CRITERIA

We excluded participants who indicated low proficiency in English and the understanding of our materials (<5 on a 1–7 scale), low seriousness (<4 on 1–5 scale), familiarity with the materials (answered 'Yes' to seeing these materials before either at the end or at any of the two familiarity checks in Study 1 and 2), failure to comprehend the scenarios (inaccurate response on a) a question whether the scenario was about time or money and b) a question about whether their understanding of the materials was accurate, after ensuring they have understood the critical information), and participants who dropped out and failed to complete all three studies. We report the number of people excluded for each criterion and analyze their effect in the 'Pre-exclusion versus post-exclusion results comparison' section in the supplementary materials. Overall, we found no differences in conclusions when comparing analyses run on participants before exclusions and after exclusions.

PARTICIPANTS

Overall, 1348 participants started the survey with 821 participants passing all exclusion criteria[1] and were included in the final analyses (M_{age} = 44.03, SD_{age} = 12.79; 52.01% males, 47.02% females). We provide details of the final sample and a comparison to Soman's (2001) samples in Table 4.

We recruited native English speakers who were born, raised, and located in the US on Amazon Mechanical Turk using the CloudResearch/TurkPrime platform (Litman et al., 2017). Based on our extensive experience of running similar judgment and decision-making replications on MTurk, to ensure high-quality data

STUDY	SMALLEST EFFECT SIZE FROM ORIGINAL	POWER TO DETECT SMALLEST EFFECT SIZE FROM ORIGINAL		SMALLEST EFFECT SIZE DETECTABLE WITH 80% POWER	
		N = 515	*N* = 1030	*N* = 515	*N* = 1030
Study 1	.61	99%+	99%+	.12	.09
Study 2	.32	99%+	99%+	.17	.12
Study 5	.31	99%+	99%+	.13	.09

Table 3 Power analysis.

Note. Effect size for Study 1 and 2 is φ_c and for Study 5 – ; see "Effect sizes calculation" section in the supplementary materials.

	SOMAN (2001)	REPLICATION
Sample size	Study 1: 122 Study 2: 206 Study 5: 72	821
Geographic origin	Study 1: Hong Kong Study 2: US American Study 5: US American	US American Amazon Mechanical Turk workers
Gender	Undisclosed	427 males, 386 females, 8 other/did not disclose
Median age (years)	Undisclosed	42
Average age (years)	Undisclosed	44.03
Standard deviation age (years)	Undisclosed	12.79
Age range (years)	Undisclosed	20-82
Medium (location)	Study 1: Physical survey Study 2: Physical survey Study 5: Physical survey	Computer (online)
Compensation	Study 1: Credit Study 2: Undisclosed Study 5: Undisclosed	Nominal payment
Year	2001	2023
Sample source	Undergraduate students	General population

Table 4 Comparison of the Soman's (2001) and the current sample.

collection, we employed the following CloudResearch options: Duplicate IP Block. Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, Block Low Quality Participants, and the like. We also employed Qualtrics' fraud and spam prevention measures: reCAPTCHA, prevent multiple submissions, prevent ballot stuffing, bot detection, security scan monitor, and relevantID. We provide more details in the 'Additional information about the study' subsection in the supplementary materials.

The assignment pay was calculated based on the federal wage of 7.25USD/hour (though we did not restrict participation based on state-level minimum wage). We first pretested survey duration with 30 participants to make sure our time run estimate was accurate and then adjust pay as needed. The data from the 30 participants was not analyzed separately from the rest of the sample other than to assess survey completion duration and needed pay adjustments. For those pretest participants, if the survey duration was longer than expected, they were paid a bonus as a pay adjustment. Additionally, we used the feedback given by those participants to improve the quality of the survey. Specifically, we found typos in our original implementation, which we fixed, and identified an inconsistency in one of the conditions in Study 1, which we addressed after consulting with the author of the target article (see 'Method' of 'Replication of Study 1' below). We also added a brief note at the beginning to let participants know that failure to pass comprehension and attention checks or answer in a certain way will not result in rejections so they should answer to the best of their ability.

REPLICATION OF STUDY 1

Study 1 was meant to test the first hypothesis that the sunk cost effect is weaker for time than for money. Participants read a hypothetical scenario about having

invested either time or money and needed to decide whether to invest further resources into a preferred or a non-preferred option. We provided more information on the stimuli, procedure and measures in the supplementary materials ("Materials used" subsection).

METHOD

Design and procedure

We employed a between-subject design with random allocation in either time or money sunk cost condition. In both conditions, the dependent variables were the same: a two-alternative forced choice, like the original, and a continuous preference scale, which we added (see next section).

Both the sunk time and the sunk money conditions asked participants to imagine that they had invested more resource (time or money)[2] for a ticket for a theatre performance compared to the resource (time or money) invested for a ticket for a rock concert, but that they preferred going to the rock concert.

MEASURES

Two-alternative forced choice (replication)

Participants then had to decide whether they would prefer to go to the theatre performance or the rock concert.

Preference (extension)

Because Study 5 employed very similar scenarios, we wanted to compare the responses from Study 1 to those of Study 5. To do so, we added in Study 1 the same measure as the original Study 5, which asked participants to indicate their preference on a scale of 1 (rock concert) to 9 (theatre performance). The scale was presented to participants as 4 (*Definitely Rock Concert*) through 0 (*Indifferent*) to 4 (*Definitely Theatre Performance*). A higher score on the scale represents less susceptibility to sunk cost fallacy. As a preliminary insight, in our replication of Study 5, we added the two-alternative forced choice that the current Study 1 had with the same aim of comparing responses across studies.

RESULTS

Two-alternative forced choice (replication)

We conducted a chi-square test and found support for differences in participants' choice of theatre performance versus rock concert ticket between the sunk time cost (13.3% chose theatre performance ticket) and sunk money cost (48.2% chose theatre performance ticket) conditions, $\chi^2(1) = 120.12$, $p < .001$, $\varphi_c = .38$, 95% CI [.31, .45] – see Figure 1A).

Similarly, the original study found that 4.8% of participants preferred the theatre performance ticket in the sunk time condition, and 61.7% in the sunk money condition, thereby also showing a strong effect of sunk cost domain, $\chi^2(1) = 44.68$, $p < .001$, $\varphi_c = .61$ 95% CI [.43, .78].

Preference (extension)

We conducted an independent samples t-test and found support for differences between the preference ratings of people in the time condition ($M = 2.71$, $SD = 2.21$) compared to those in the money condition ($M = 4.61$, $SD = 2.99$), t (665.68) = –10.16, $p < .001$, $d = -.79$ [–.94, –.63] (see Figure 1A).

DISCUSSION

In our replication of the target's Study 1, we found support for differences between sunk money and sunk time, albeit with weaker effects than that of the target article, that the probability of the sunk cost effect is greater for money than for time, thereby supporting H1.

REPLICATION OF STUDY 2

In this study, we further interrogated the first hypothesis, namely that the sunk cost effect is weaker for time than for money, by building on the previous study by adding another condition: whether there is a sunk cost or not. This allowed us to test whether the sunk cost effect would appear when comparing sunk cost versus no sunk cost conditions in each domain (time/money).

METHOD

Design

We employed a 2 (*sunk cost domain*: time or money) × 2 (*sunk cost presence*: sunk cost or no sunk cost) between-subjects design with random allocation. In all conditions, the dependent variable was the same two-alternative forced choice.

Procedure

In the sunk cost conditions (regardless of the sunk cost domain), participants were asked to imagine that they had already invested substantial resources in developing a new rocket engine invention for a competition compared to no resource invested in developing a solar-powered pump. To finish either project would require the same resources, but they learn that the winner of last year's competition also worked on a rocket engine design. They are then asked whether they would prefer to continue working on the rocket engine design (on which they have already spent resources) or to complete a solar-powered pump design.

In the no sunk cost condition, participants are presented with the same story, but they are not told that they had already invested resources in either design. We provided additional details in the 'Materials used' subsection of the supplementary materials.

RESULTS

We conducted two chi-square tests to analyze the difference between the sunk cost and no sunk cost conditions in each domain (time and money).

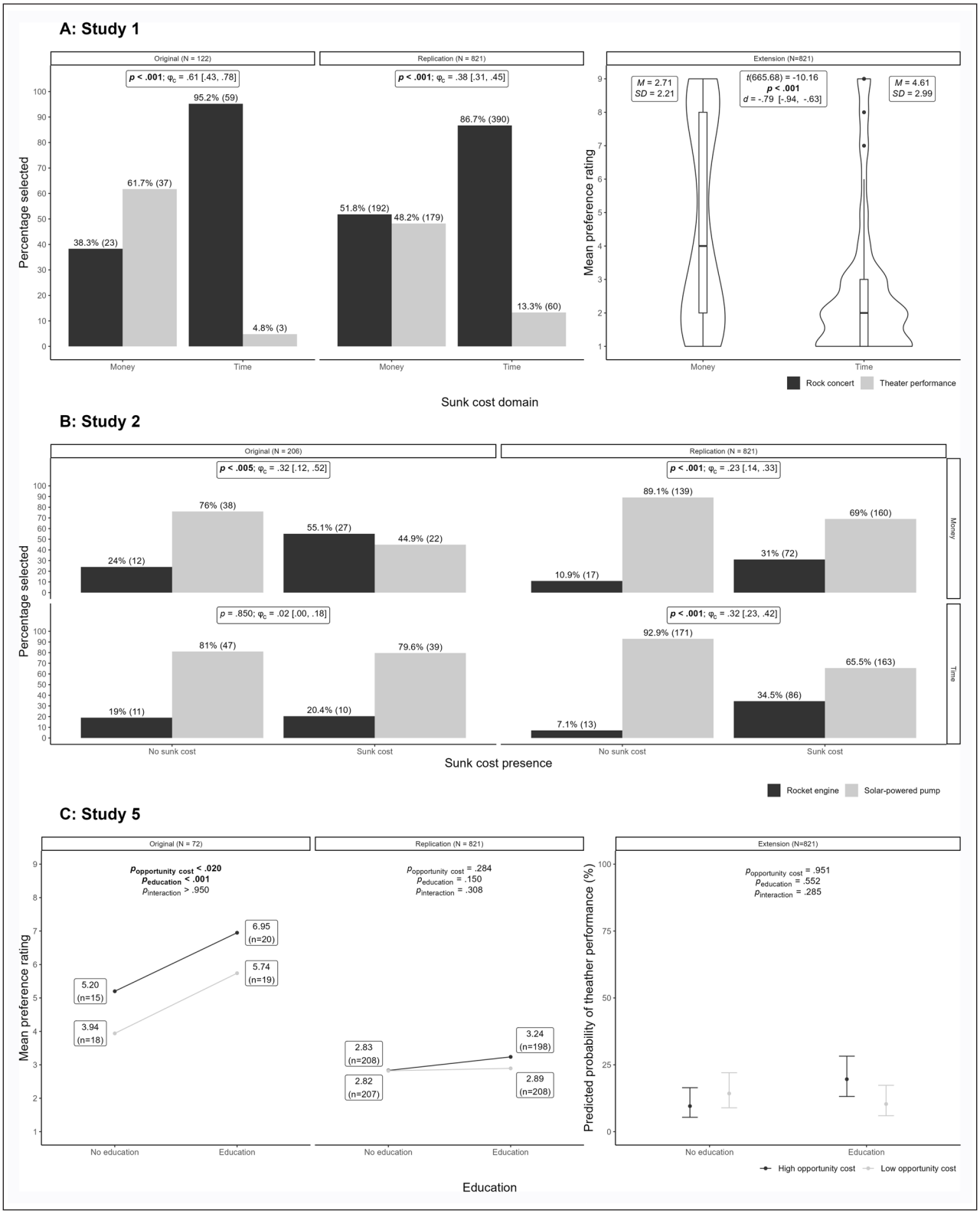

Figure 1 Summary of results comparing Soman's original studies to the current replication effort.

Note. Bold indicates statistically significant correlations.

With time sunk cost, we found support for differences between those that read the sunk cost scenario (34.5% chose the rocket engine) and those who read the no sunk cost scenario (7.1% chose the rocket engine) in choosing which design to work on, $\chi^2(1) = 45.28$, $p < .001$, $\varphi_c = .32$, 95% CI [.23, .42].

With money sunk cost, we also found support for differences between those that read the sunk cost scenario (31.0% chose the rocket engine) and those who read the no sunk cost scenario (10.9% chose the rocket engine) in choosing which design to work on, $\chi^2(1) = 21.40$, $p < .001$, $\varphi_c = .23$, 95% CI [.14, .33] – see Figure 1B).

In comparison, the original study also found support for an effect in the money domain (55.1% chose the rocket engine in the sunk cost, and 24.0% in the no sunk cost, $\chi^2(1) = 10.03$, $p = .002$, $\varphi_c = .32$, 95% CI [.12, .52]), yet, unlike in our replication, found no effect in the time domain (20.4% chose the rocket engine in the sunk cost, and 19.0% in the no sunk cost, $\chi^2(1) = .04$, $p = .852$, $\varphi_c = .02$, 95% CI [.00, .18])

In summary, we found support for differences between sunk cost and no sunk cost conditions in both time and money domains. This conclusion is in line with the original's findings for the money domain, albeit with a smaller effect size, but not in line with the findings for the time domain, which we found comparable to larger effects compared to the money domain.

DISCUSSION

In our replication we found support for sunk cost effect in the money domain, although weaker than that in the target article. However, unlike in the target article, we also found support for sunk costs effect in the time domain. Crucially, the results of the target's Study 2 indicate that the sunk cost effect in the time domain were greater than in the money domain, thereby not in line with H1, not in line with the target's Study 2 results, and diverging from the support for H1 that we found in our replication of Study 1.

REPLICATION OF STUDY 5

In this study, we tested the second hypothesis, namely whether the facilitation of accounting for time strengthens the sunk time cost effect. To do this, we presented participants with a few paragraphs aimed at educating them about economic approaches to time. Additionally, we also varied the magnitude of the opportunity cost, such that it could be either low or high. This setup allowed us to test not only whether the education intervention works, but also the conditions in which that occurs.

METHOD

Design

We employed a 2 (*opportunity cost*: low or high) × 2 (*education*: education or no education) between-subjects design with random allocation. In all conditions, the dependent variables were the same: a continuous preference scale, like the original, and a two-alternative forced choice, which we added (see next section).

Procedure

The scenario was similar to the one used in the replication of Study 1 with two differences. First, in the high opportunity cost condition, participants were told that they were 'badly pressed for time,' while in the low opportunity cost condition they were told that there is 'relative flexibility in your schedule.' Second, an education intervention was implemented: those who received education about opportunity costs were asked to read a short passage, which explained what an opportunity cost is and gave a thorough example. We provided additional details in the 'Materials used' subsection of the supplementary materials.

Measures

Preference (replication)

Participants indicated their preferences on a scale of 1 (*Rock Concert*) to 9 (*Theatre Performance*) which we presented to participants as 4 (*Definitely Rock Concert*), 0 (*Indifferent*) and 4 (*Definitely Theatre Performance*). A higher score on the scale represents less susceptibility to sunk cost fallacy. We note that this is a deviation from the original's measure that ranged from 1 to 9 in presentation. We made this adjustment to avoid biasing participants towards the option presented with larger numbers.

Forced choice (extension)

To be able to compare the findings of Study 1 with that of Study 5 that employed similar stimuli we added the same two-alternative forced choice measure that was used in Study 1. As in Study 1 (above), participants had to decide whether they would prefer to go to the theatre performance or the rock concert.

RESULTS

Preference (replication)

To analyze the effects of opportunity cost and education on preference ratings for one ticket or the other, we selected Type-III ANOVA (to account for any variance in potential interactions; Field 2017). Assumptions for normality, outliers and homogeneity of variances were met, although ANOVA is robust to these violations with large samples (Blanca et al., 2017).

We conducted a 2 (Opportunity cost) × 2 (Education) between-groups ANOVA on preference ratings. We found no support for a main effect of opportunity cost, $F(1, 817) = 1.15$, $p = .284$, $\omega^2 = .00$, $\eta^2_p = .00$, 95% CI [.00, .01] with those in the high opportunity cost condition ($M = 3.03$, $SD = 2.43$) not providing statistically different preference ratings than those in the low opportunity cost condition ($M = 2.86$, $SD = 2.24$). We also found no support for a main effect of education, $F(1, 817) = 2.08$, $p = .150$, $\omega^2 = .00$, $\eta^2_p = .00$, 95% CI [.00, .01] with no support for differences in preference ratings between those in the education condition ($M = 3.06$, $SD = 2.46$) and those in the no-education condition ($M = 2.83$, $SD = 2.21$). We found no support for an interaction effect between the independent variables ($F = 1.04$, $p = .308$ – see Figure 1C).

In comparison, in the original study they found a main effect of opportunity cost, $F(1, 68) = 6.63$, $p < .020$, $\omega^2 =$

.073, η^2_p = .089, 95% CI [.00, .23], with those in the high opportunity cost condition (*M* = 6.20)[3] providing higher preference ratings that those in the low opportunity cost condition (*M* = 4.86). The original study also found a main effect of education, $F(1, 68) = 13.65$, $p < .001$, $\omega^2 = .149$, η^2_p = .167, 95% CI [.04, .32], with those who underwent education (*M* = 6.36) providing higher preference ratings than those who did not undergo education (*M* = 4.52). The interaction between the two factors was not supported in the original study, $p > .950$.

Thus, though we also find no interaction effects, we fail to replicate the main effects of opportunity cost and education.

Forced choice (extension)

To analyze the two-alternative forced choice responses, we built a generalized linear model (GLM). We have already built this model for our later exploratory analyses (see section 'Study 1 versus Study 5: Analysis of within-subject effects' for details of the model building procedure) which included the same two independent variables as the ANOVA, namely opportunity cost, education, and their interaction, as well as an additional independent variable and other interactions. In that generalized linear model, we coded the factors such that we get the results for the current study, therefore we report the results for the two main effects of interest and their interaction.

Specifically, the GLM showed no support for a main effect of opportunity cost (OR = .97 [.38, 2.46], *p* = .951), no support for a main effect of education (OR = 1.21 [.64, 2.30], *p* = .552), and no support for an interaction (OR = 2.00 [.56, 7.23], *p* = .285) – see Figure 1C right-hand side as well as Table 5.

DISCUSSION

Not in line with the results report in the target article, we failed to find evidence in support of facilitation of accounting for time, either using education or in highlighting opportunity costs. This suggests no support for H2a.

SUMMARY OF RESULTS

REPLICATION RESULTS

We summarized our findings in Table 1.

In our replication of Soman's (2001) Study 1, we found support for sunk costs in money larger than sunk costs for time (H1), although with weaker effects. Our extension using a continuous preference variable supported the conclusions of the replication.

In our replication of Study 2, we found mixed evidence relative to the original. On the one hand, our findings are consistent with Soman's (2001) when comparing the sunk cost with the no sunk cost conditions in the money domain. On the other hand, we also found support for an effect in the time domain, inconsistent with Soman's (2001) findings. Moreover, we found that the effect in the time domain was larger than in the money domain.

In our replication of Study 5, we failed to find any support for any of the effects, inconsistent with two main effects of education and opportunity cost in Soman's (2001). We found no indication of benefits from facilitation of accounting for time, not in line with H2a.

ADDITIONAL ANALYSES AND CHECKS

Sunk cost effect stronger for money than for time: Re-analysis using logistic regression

To address H1, Soman (2001) conducted multiple chi-square tests. Specifically, in Study 2, he showed that in the money condition, the chi-square test found support for differences between sunk cost and no sunk cost conditions, whereas the same difference was not supported for the time condition.

A different way to approach H1 is to ask whether the likelihood of picking the option associated with sunk costs (rocket engine in Study 2) is different not only between levels of a single independent variable (sunk cost presence or sunk domain) but also whether there was an interaction between the two variables. To address this question, we conducted logistic regression analyses for Study 2 for both the original and the replication data as it allowed us to test the interaction effect.

PREDICTORS	PREFERENCE			TWO-ALTERNATIVE FORCED CHOICE		
	ESTIMATE	*95% CI*	*p*	*ODDS RATIO*	*95% CI*	*p*
Opportunity cost	.09	–.46, .63	.758	.97	.38, 2.46	.951
Study	.15	–.23, .54	.437	1.37	.75, 2.57	.311
Education	.22	–.17, .60	.273	1.21	.64, 2.30	.552
Opportunity cost × study	–.46	–1.23, .31	.240	.65	.19, 2.23	.497
Opportunity cost × education	.48	–.30, 1.25	.228	2.00	.56, 7.23	.285
Study × education	.08	–.47, .63	.770	1.04	.45, 2.43	.922
Opportunity cost × study × education	.31	–.79, 1.41	.578	1.67	.31, 9.11	.551

Table 5 Results of linear (DV: Preference) and generalized linear (DV: Binary choice) models from the additional within-subjects analysis.

We ran a logistic regression for Study 2. The dependent variable was coded as 0 (solar-powered pump) and 1 (rocket engine). The predictors in the model were sunk domain (money/time) and sunk cost presence (no sunk cost/sunk cost) as well as their interaction. We wanted our model to test whether there was a main effect of sunk domain in the sunk cost present condition (thereby replicating the effect from Study 1) and also whether there was a main effect of sunk presence, regardless of the sunk domain. In order to achieve this, we coded sunk domain as a sum contrast and sunk presence as a treatment contrast, with sunk cost present as its baseline condition. In order to get predicted probabilities from the model for main effects with no baseline condition for remaining factors (in this case the main effect for sunk presence), we applied marginal standardization, which has been reliably shown to be a robust method compared to alternatives (Muller & MacLehose, 2014; Williams, 2012).

The results of the logistic regression for Study 2 on Soman's (2001) original data showed that there was support for a main effect of sunk domain, such that the odds of selecting the rocket engine design in the sunk cost condition went down by 79% in the domain of time compared to money (OR = .21 [.08, .50], p = .001). Soman's data also revealed support for a main effect of sunk cost presence, regardless of sunk domain, such that the odds of selecting the rocket engine were 52% lower in the no sunk cost compared to the sunk cost condition (OR= .48 [.25, .92], p = .027; sunk cost effect differences between the money and time domains: OR = 3.55 [.99, 13.06], p = .053; see Figure 2A).

We conducted a logistic regression for Study 2 in our replication data and found a main effect of sunk presence (OR = .20 [.13, .30], p < .001), but no support for a main effect of sunk domain (OR = 1.17 [.80, 1.72], p = .414), and no support for an interaction (OR = .53 [.22, 1.23], p = .142) – see Figure 2B. See Table 6 for a summary.

Study 1 versus Study 5: Analysis of within-subject effects

We extended the original analyses of H2 by considering an additional within-subject factor: study. Specifically, we took advantage of three of our design choices: 1) the replications of Study 1 and Study 5 both involved the same theatre performance versus rock concert ticket scenario, with the only difference that the design of Study 5 was a 2 × 2 between-subjects; 2) the same participants completed both Study 1 and Study 5 in the same survey; 3) we included both the two-alternative forced choice and the Likert response scales in both Study 1 and Study 5.

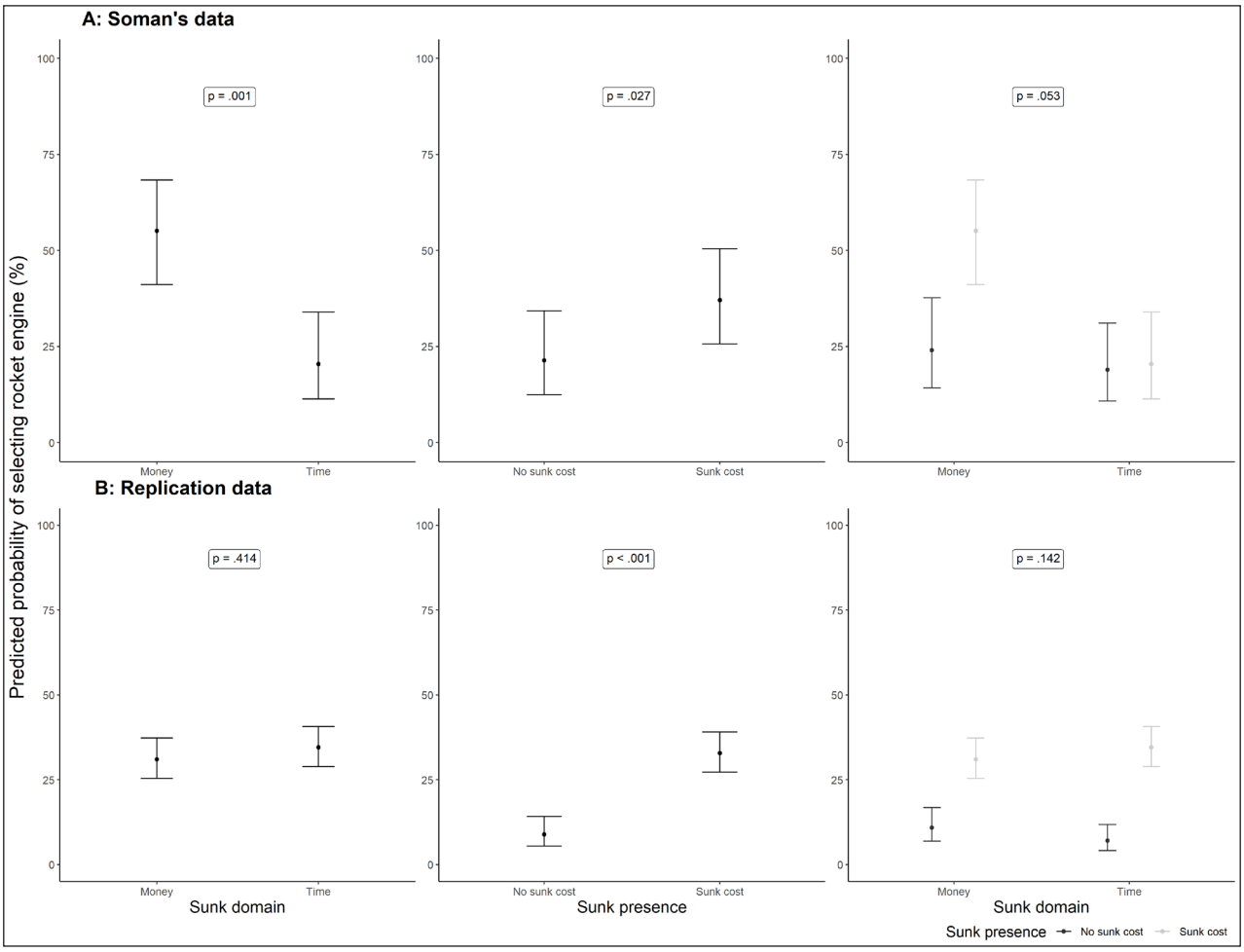

Figure 2 Study 2: Predicted probabilities from logistic regression analyses.

Note. The main effect of sunk domain is plotted based on predicted probabilities from the sunk cost present condition, while the main effect of sunk presence is plotted using marginal standardization across levels of sunk domain.

HYPOTHESIS	QUESTION ADDRESSED	ANALYSIS USED	CONSISTENT WITH REPLICATION ANALYSIS (YES/ MIXED/NO)	DETAILS
The sunk cost effect is weaker for time than for money.	Does the likelihood of picking the option associated with sunk costs (rocket engine in Study 2) vary significantly between levels of one independent variables (sunk cost presence or sunk domain) given a change in the other (i.e., an interaction effect)?	2x2 logistic regression on both Soman's original data as well as the replication data.	Yes	Both re-analyses and replication analyses are not in-line with the hypothesis: the replication analyses showed comparable to larger effect size for time than for money, whereas the re-analyses show no support for domain differences.
Facilitation of money-like accounting by using education about economic approaches to time strengthens the sunk cost effect of time (tested only in the time domain).	What are the differences between Study 1 and the high versus low opportunity cost conditions in Study 5 (i.e., study by opportunity cost interaction, in the no education condition in Study 5)?	Two linear models: one linear model with preference ratings as the dependent variable and one generalized LM with 2-alternative ticket choice as the dependent variable. The models included three independent variables: study (Study 1 vs Study 5), opportunity cost (low vs high), education (no education vs education), and all their interactions.	Yes	Both no support for interactions show that, at least in the time domain, neither the opportunity cost, nor the education manipulations, made a difference. Although this is aligned with the replication analyses in our sample, it is not in-line with Soman's (2001) conclusion.
	Are differences between Study 1 and the high versus low opportunity cost conditions in Study 5 affected by education (study by opportunity cost by education interaction)?		Yes	

Table 6 Summary of additional analyses.

This allowed us to address two additional questions: 1) What are the differences between Study 1 and the high versus low opportunity cost conditions in Study 5 (i.e., study by opportunity cost interaction, in the no education condition in Study 5), and 2) Are those differences affected by education (study by opportunity cost by education interaction). To test both questions, we focused on the time sunk cost domain, as Study 5 only included the vignette version in the time domain.

To address these questions, we constructed two linear models: one linear model (LM) with preference ratings on a continuous scale as the dependent variable and one generalized LM (GLM) with ticket choice (theatre performance coded as 1 and rock concert as 0) as the dependent variable. We included three independent variables: study (Study 1 vs. Study 5), opportunity cost (low vs. high), education (no education vs. education), and all their interactions. The factor variables were coded such that study was set as a treatment contrast, with Study 5 as the baseline condition, opportunity cost was coded as a sum contrast, and education was coded as a treatment contrast, with no education as the baseline condition.

The results of both models showed no two-way interaction between study and opportunity cost in the no education condition, and no three-way interaction between opportunity cost, study and education in either of the models—see Table 5 for outputs of those models and Table 6 for a summary of the results. Our findings for Study 5 indicate that no support for the introduction of opportunity cost, with or without education, as having any impact relative to Study 1.

Order effects between studies

One deviation from the original study is that all participants completed all scenarios. We considered this to be a stronger design with many advantages that we laid out in the 'Studies overview' section above, yet one disadvantage is that answers to one scenario may bias participants' answers to following scenarios (recall that Study 1 and 2 were presented in random order followed by Study 5). To address this is to run all analyses for each of the studies by only focusing on the participants that completed that study first. We found no differences in conclusions—see Table 7.

Exploratory comprehension questions analyses

Given the large number of participants excluded due to failing a comprehension question (70% of all excluded), we decided to investigate further. There were two questions in Study 1 and 3 questions in Study 2. The first question was about the focus of the scenario which asked, 'What was the cost of the tickets, time or money?' in Study 1 and 'What was the scenario mainly focused on, time or money?' in Study 2. The second question was a self-report on understanding which was phrased 'Did you correctly understand the scenario the first time...?' for both studies. The third question appeared only in Study 2 and was about whether the participant correctly understood if they have already invested something:

ANALYSIS	STATISTICAL TEST AND FACTORS		FULL SAMPLE *N* = 821		STUDY 1 FIRST *N* = 393		STUDY 2 FIRST *N* = 428	
			ES	*p*	ES	*p*	ES	*p*
Replication analyses								
Study 1: Forced choice	Chi-square		.38	< .001	.41	< .001	.36	< .001
Study 1: Preference	Independent samples t-test		–.79	< .001	–.74	< .001	–.84	< .001
Study 2: Time domain	Chi-square		.32	< .001	.36	< .001	.29	< .001
Study 2: Money domain	Chi-square		.23	< .001	.28	< .001	.19	< .001
Study 5: Preference	between-groups ANOVA	opportunity cost	.00	.284	.01	.118	.00	.969
		education	.00	.150	.01	.058	.00	.850
		opportunity cost × education	.00	.308	.00	.359	.00	.527
Study 5: Forced choice	Generalized Linear Model	opportunity cost	.97	.951	1.40	.619	.68	.570
		education	1.21	.552	1.75	.207	.82	.693
		opportunity cost × education	2.00	.285	1.22	.825	3.69	.185
Additional analyses and checks								
Study 2 re-analysis	Logistic Regression	sunk domain	1.17	.414	1.01	.960	1.35	.282
		sunk presence	.20	< .001	.15	< .001	.26	< .001
		sunk type × sunk presence	.53	.142	.49	.291	.55	.297
Study 1 versus Study 5: Analysis of within subject effects	Linear model	opportunity cost	.09	.758	.35	.370	–.23	.566
		study	.15	.437	.03	.901	.29	.308
		education	.22	.273	.44	.118	.05	.852
		opportunity cost × study	–.46	.240	–.38	.486	–.54	.342
		opportunity cost × education	.48	.228	.42	.457	.65	.234
		study × education	.08	.770	.33	.412	–.18	.647
		opportunity cost × study × education	.31	.578	–.01	.986	.62	.424
	Generalized Linear Model	opportunity cost	.97	.951	1.40	.619	.68	.570
		study	1.37	.311	1.23	.650	1.43	.436
		education	1.21	.552	1.75	.207	.82	.693
		opportunity cost × study	.65	.497	.93	.939	.49	.436
		opportunity cost × education	2.00	.285	1.22	.825	3.69	.185
		study × education	1.04	.922	1.24	.722	.94	.928
		opportunity cost × study × education	1.67	.551	1.29	.832	2.02	.597

Table 7 All analyses re-run, split by whether Study 1 or Study 2 was presented first.

Note. Reported effect sizes (ES) are: Chi-square – φ_c, Independent samples t-test – Cohen's *d*, ANOVA – , Generalized Linear Model and Logistic Regression – Odds Ratios, Linear model – β.

'Have you already invested anything in your current design?' The order of these questions was such that the scenario focus and the already invested questions (the latter for Study 2 only) appeared on the same page then on the next page was the comprehension quiz participants had to take and on the following page was the self-report question. The percentage of people who responded incorrectly out of everyone who completed the question (before exclusions) is found in Figure 3.

Further exploratory analyses revealed two notable trends. First, across both studies and all questions, participants failed a comprehension question more often in the money condition compared to the time condition, $\chi^2(1) = 65.71$, $p < .001$, $\varphi_c = .10$, 95% CI [.08, .13]. Second, in Study 2 across all questions, participants failed a comprehension question in the no sunk cost condition more often than in the sunk cost condition, $\chi^2(1) = 170.05$, $p < .001$, $\varphi_c = .21$, 95% CI [.18, .25].

These findings suggest two observations. First, given the high rate at which participants reported some lack of comprehension, we think our study was well-designed to incentivize people to report their understanding truthfully, without limiting their pay and incentivizing them to lie. Second, we believe that the presence (or lack thereof) of comprehension questions in hypothetical scenarios, often used in studies of the sunk cost effect,

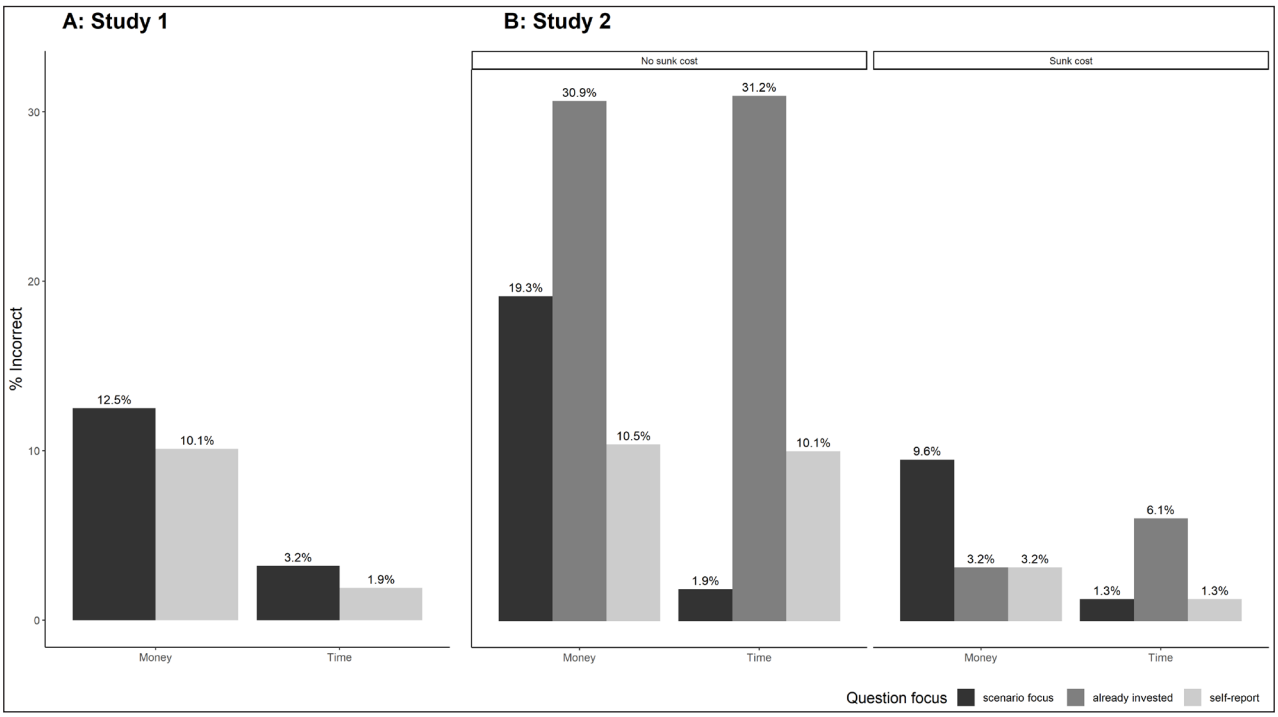

Figure 3 Percent of incorrect responses on comprehension questions across the entire sample.

can have major impact on the data quality, and so we encourage further research to add them to gain further insights into potential comprehension issues.

We therefore believe that it was important to include exploratory comprehension questions after the target's replication scenarios, and that our findings are likely robust given that we ensured that the participants included in the final sample read and understood the scenarios as intended.

GENERAL DISCUSSION

We conducted a close, high-powered, replication and extension Registered Report of Soman's (2001) Studies 1, 2, and 5 and found mixed results.

In line with Soman's (2001) findings, we found that, based on Study 1, participants show a stronger sunk cost effect in money than in time. However, in our replication of Study 2, we found evidence for both sunk money costs and sunk time costs. In our replication of Study 5, we found no evidence suggesting that manipulating either the strength of the opportunity costs or educating people about economic approaches to time had any impact on sunk time cost effect.

Similar to previous studies, we found strong evidence for sunk money costs (Arkes & Blumer, 1985; Bornstein et al., 1999; Coleman, 2009; Navarro & Fantino, 2005; Soman & Cheema, 2001). Our findings also seem to be in line with some of the existing literature suggesting sunk time costs seem volatile, depending on context.

One possible reason for this could be that whereas Soman (2001) recruited undergraduate students across his studies, our sample was from the general population. These groups could be different based on age, which differentially affects susceptibility to sunk costs (Strough et al., 2008), based on the experience they have accounting for time (Bornstein et al., 1999; DeVoe & Pfeffer, 2007; Ronayne et al., 2021), and based on cultural influences, which might make our sample more sensitive to accounting for time, given the greater emphasis placed productivity and time-management in modern discourse. We also note that conducted our study right after the world started coming back to normal after the COVID-19 pandemic, which may have impacted people's appreciation of time.

Although differences abound between our replication and the target's, the inconsistent mixed results of our replications of Studies 1 and 2 in relation to the strength of the sunk time cost effect are more difficult to explain. One of the major strengths of our replication was that all participants completed all studies, thus rendering explanations that refer to between-sample differences unlikely. Still, we find that the sunk money cost effect is stronger than the sunk time cost effect in Study 1, but with no indication for differences between the two in Study 2. This difference could not be explained by the randomized order of the studies, the exclusion criteria, or the comprehension of the scenarios. This likely suggests that there are other factors that further complicate the study of sunk cost effects that would need to be addressed by future research.

LIMITATIONS OF OUR REPLICATION AND DIRECTIONS FOR FUTURE RESEARCH

Our replication had limitations that can be broadly split into two categories: contextual differences and implementation.

The context surrounding how people think about time and money sunk costs might have changed over time. Our replication was conducted more than 20 years after the original thus how people approach time and money may have changed. This is partly why ongoing repeating replications are needed, to keep our knowledge about an important phenomenon up to date. In addition to repeating replications, we call for regular systematic reviews and/or meta-analyses that would help summarize the growing literature, address and help explain seemingly discrepant results (such as our differences in patterns of results across Studies 1 and 2), and would drive the development of better theories and empirical investigations. Another factor that could influence how people think about time and money sunk costs was sample differences: we used a sample from the general population, which might systematically differ on demographic characteristics that are correlated with the sunk cost effect compared to the student sample used in the original. For instance, our replication sample was likely twice as old as that of the original, which might make them more sensitive to any further time investments (e.g., Strough et al., 2008), given that sunk cost calculations are done relative to one's total available resources, rather than taken as absolute (Garland & Newport, 1991). Following similar logic, it is possible that the social class of the different samples could also have an impact. The influence of such demographic factors is speculative at best and further research is needed to disentangle these open questions.

The implementation of our replication also introduced some limitations, especially in our exploratory conceptual replication of Study 5. First, we made adjustments to the opportunity cost manipulation. Second, in the original, the education intervention was implemented by manipulating when the study was conducted—either before a classroom discussion about the economic value of time (control condition) or after (education condition)—whereas in our replication, the intervention was implemented by having participants read information on the screen and complete comprehension checks. These changes were necessary given the change in the medium, yet it may have affected the results. Third, the studies were originally run separately, and in our design, we ran the studies together, with Study 5 always last, given its similarity to Study 1. This allowed us to gain additional insights and conduct a comparison between Study 1 and Study 5, yet this does mean that our adjustments make the replication of Study 5 less direct in comparison to Studies 1 and 2, with higher likelihood of the results being different than that of the target.

CONCLUSION

We conducted a Registered Report of a close, high-powered, replication and extension of Studies 1 and 2 and a conceptual replication of Study 5 in Soman (2001), testing the predictions that sunk cost effect is weaker in the time domain than in the money domain, and that the facilitation of money-like accounting for sunk time costs would strengthen the sunk time cost effect. We concluded mixed support: a) sunk time cost effect was stronger than the sunk money cost in Study 1, yet no support for differences between the two in Study 2, and b) we found no indication of facilitation of money-like accounting as having any impact on the sunk time cost effect. We tested and ruled out several study order or exclusions as the possible explanations and discussed directions for future research.

See recommendation and open peer-review on: https://rr.peercommunityin.org/articles/rec?id=452.

NOTES

1 The ~40% exclusion rate we find is significantly higher than the 15% we anticipated in our preregistration. 70% (373 of 527) of those that were excluded were due to failing a comprehension question. We further investigate this in the "Comprehension question analysis" subsection below.

2 For the money condition, we adjusted the original's absolute values to account for inflation and meet updated minimum wages in the US. Specifically, the original article, published in 2001, promised compensation to the hypothetical student in an imagined scenario equivalent to US$20 for 5 hours of work and US$60 for 15 hours of work. We adjusted these numbers to US$75 and US$225 for 5 and 15 hours of work, respectively.

3 The original did not report standard deviations.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Supplementary Materials.** DOI: https://doi.org/10.5334/irsp.883.s1

ACKNOWLEDGEMENT

We thank Devin Zaixuan ZHANG for his help assisting with feedback. We thank Prof. Dilip Soman for his support and for making the materials available for reproduction.

FUNDING INFORMATION

The project was supported by the University of Hong Kong Teaching Development Grant awarded to Gilad Feldman.

COMPETING INTERESTS

The authors have no competing interests to declare.

CONTRIBUTOR ROLES TAXONOMY

ROLE	NIKOLAY PETROV	YIN KAN CHAN, CHEUK NAM LAU, TIN HO KWOK. LOK CHING CHOW, WAI YAN LO	WENKAI SONG	GILAD FELDMAN
Conceptualization		X		X
Data curation	X	X		X
Formal analysis	X	X		
Funding acquisition				X
Investigation	X	X		
Preregistration verification	X		X	
Data analysis verification	X			
Methodology	X	X		
Project administration				X
Resources	X	X		X
Software	X	X		
Supervision				X
Validation	X			
Visualization	X	X		
Writing – original draft	X	X		
Writing – review and editing	X		X	X

AUTHOR CONTRIBUTIONS

Yin Kan Chan, Cheuk Nam Lau, Tin Ho Kwok, Lok Ching Chow, and Wai Yan Lo designed the study, developed the experimental materials for each study respectively and wrote an initial draft of the Registered Report Stage 1. Nikolay B Petrov took lead over the project, redid the analyses, finalized the materials and analysis scripts, wrote the manuscript for submission, and addressed reviewers' feedback and revisions. Wenkai verified and provided feedback on the analyses and documents. Gilad Feldman supervised the project, collected data, revised drafts, and finalized manuscripts for submission.

Yin Kan (Megan) Chan, Cheuk Nam (Chris) Lau, Tin Ho (Donald) Kwok, Lok Ching (Estelle) Chow, Wai Yan Vivian Lo shared second author.

AUTHOR AFFILIATIONS

Nikolay B. Petrov orcid.org/0000-0002-1305-0547
PhD student at University of Cambridge, UK
Yin Kan Megan Chan orcid.org/0009-0009-7645-7721
Students at the Department of Psychology, University of Hong Kong during the academic year 2021/22, HK
Cheuk Nam Lau
Students at the Department of Psychology, University of Hong Kong during the academic year 2021/22, HK
Tin Ho Kwok orcid.org/0009-0006-0769-836X
Students at the Department of Psychology, University of Hong Kong during the academic year 2021/22, HK
Lok Ching Estelle Chow
Students at the Department of Psychology, University of Hong Kong during the academic year 2021/22, HK
Wai Yan Lo
Department of Psychology, University of Hong Kong, Hong Kong SAR, HK
Wenkai Song orcid.org/0000-0001-9905-7100
School of Entrepreneur and Management, ShanghaiTech University, Shanghai, CN
Gilad Feldman orcid.org/0000-0003-2812-6599
Assistant Professor at the Department of Psychology, University of Hong Kong, HK

REFERENCES

Arkes, H. R., & **Blumer, C.** (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. DOI: https://doi.org/10.1016/0749-5978(85)90049-4

Blanca, M. J., Alarcón, R., & **Arnau, J.** (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552–557. DOI: https://doi.org/10.7334/psicothema2016.383

Bornstein, B. H., & **Chapman, G. B.** (1995). Learning lessons from sunk costs. *Journal of Experimental Psychology: Applied*, *1*(4), 251–269. DOI: https://doi.org/10.1037/1076-898X.1.4.251

Bornstein, B. H., Christine Emler, A., & **Chapman, G. B.** (1999). Rationality in medical treatment decisions: Is there a sunk-cost effect? *Social Science & Medicine*, *49*(2), 215–222. DOI: https://doi.org/10.1016/S0277-9536(99)00117-3

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & **Van't Veer, A.** (2014). The replication recipe: What makes for a convincing replication?. *Journal of Experimental Social Psychology*, *50*, 217–224. DOI: https://doi.org/10.1016/j.jesp.2013.10.005

Chambers, C. D., & **Tzavella, L.** (2022). The past, present and future of Registered Reports. *Nature Human Behaviour*, *6*(1), 29–42. DOI: https://doi.org/10.1038/s41562-021-01193-7

Coleman, M. D. (2009). Sunk Cost and Commitment to Dates Arranged Online. *Current Psychology*, *28*(1), 45–54. DOI: https://doi.org/10.1007/s12144-009-9042-5

DeVoe, S. E., & **Pfeffer, J.** (2007). When time is money: The effect of hourly payment on the evaluation of time. *Organizational Behavior and Human Decision Processes*, *104*(1), 1–13. DOI: https://doi.org/10.1016/j.obhdp.2006.05.003

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th edition). SAGE Publications.

Friedman, D., Pommerenke, K., Lukose, R., Milam, G., & **Huberman, B. A.** (2007). Searching for the sunk cost fallacy. *Experimental Economics*, *10*(1), 79–104. DOI: https://doi.org/10.1007/s10683-006-9134-0

Garland, H., & **Newport, S.** (1991). Effects of absolute and relative sunk costs on the decision to persist with a course of action. *Organizational Behavior and Human Decision Processes*, *48*(1), 55–69. DOI: https://doi.org/10.1016/0749-5978(91)90005-E

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*(3), 389–402. DOI: 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*. DOI: https://doi.org/10.15626/MP.2018.843

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. DOI: https://doi.org/10.3758/s13428-016-0727-z

Muller, C. J., & **MacLehose, R. F.** (2014). Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. *International Journal of Epidemiology*, *43*(3), 962–970. DOI: https://doi.org/10.1093/ije/dyu029

Navarro, A. D., & **Fantino, E.** (2005). The sunk cost effect in pigeons and humans. *Journal of the Experimental Analysis of Behavior*, *83*(1), 1–13. DOI: https://doi.org/10.1901/jeab.2005.21-04

Navarro, A. D., & **Fantino, E.** (2009). The sunk-time effect: An exploration. *Journal of Behavioral Decision Making*, *22*(3), 252–270. DOI: https://doi.org/10.1002/bdm.624

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

Nosek, B. A., & **Lakens, D.** (2014). Registered Reports. *Social Psychology*, *45*(3), 137–141. DOI: https://doi.org/10.1027/1864-9335/a000192

Okada, E. M., & **Hoch, S. J.** (2004). Spending time versus spending money. *Journal of Consumer Research*, *31*(2), 313–323. DOI: https://doi.org/10.1086/422110

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. DOI: https://doi.org/10.1126/science.aac4716

Pandey, S. K., & **Sharma, D.** (2019). Sunk time fallacy with recoverable monetary costs. *Marketing Intelligence & Planning*, *37*(2), 154–167. DOI: https://doi.org/10.1108/MIP-02-2018-0052

Park, J.-Y., & **Jang, S. (Shawn).** (2014). Sunk costs and travel cancellation: Focusing on temporal cost. *Tourism Management*, *40*, 425–435. DOI: https://doi.org/10.1016/j.tourman.2013.08.005

Rego, S., Arantes, J., & **Magalhães, P.** (2018). Is there a sunk cost effect in committed relationships? *Current Psychology*, *37*(3), 508–519. DOI: https://doi.org/10.1007/s12144-016-9529-9

Ronayne, D., Sgroi, D., & **Tuckwell, A.** (2021). Evaluating the sunk cost effect. *Journal of Economic Behavior & Organization*, *186*, 318–327. DOI: https://doi.org/10.1016/j.jebo.2021.03.029

Scheel, A. M., Schijen, M. R. M. J., & **Lakens, D.** (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211007468. DOI: https://doi.org/10.1177/25152459211007467

Silva Castillo, L. H., Cisneros Plazola, M. C., Torres Ceja, C. de J., & **Hernández Rosas, E.** (2020). The sunk-time effect: Effect of time invested and reward magnitude using within-subject design. *Behavioural Processes*, *181*, 104247. DOI: https://doi.org/10.1016/j.beproc.2020.104247

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, *26*(5), 559–569. DOI: https://doi.org/10.1177/0956797614567341

Soman, D. (2001). The mental accounting of sunk time costs: Why time is not like money. *Journal of Behavioral Decision Making*, *14*(3), 169–185. DOI: https://doi.org/10.1002/bdm.370

Soman, D., & **Cheema, A.** (2001). The effect of windfall gains on the sunk-cost effect. *Marketing Letters*, *12*(1), 51–62. DOI: https://doi.org/10.1023/A:1008120003813

Soster, R. L., Monga, A., & **Bearden, W. O.** (2010). Tracking costs of time and money: How accounting periods affect mental accounting. *Journal of Consumer Research*, *37*(4), 712–721. DOI: https://doi.org/10.1086/656388

Strough, J., Mehta, C. M., McFall, J. P., & **Schuller, K. L.** (2008). Are older adults less subject to the sunk-cost fallacy than younger adults? *Psychological Science*, *19*(7), 650–652. DOI: https://doi.org/10.1111/j.1467-9280.2008.02138.x

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

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal: Promoting Communications on Statistics and Stata*, *12*(2), 308–331. DOI: https://doi.org/10.1177/1536867X1201200209

Wiseman, R., Watt, C., & **Kornbrot, D.** (2019). Registered Reports: An early example and analysis. *PeerJ*, *7*, e6232. DOI: https://doi.org/10.7717/peerj.6232

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

TO CITE THIS ARTICLE:

Petrov, N. B., Chan, Y. K. M., Lau, C. N., Kwok, T. H., Chow, L. C. E., Lo, W. Y., Song, W., & Feldman, G. (2023). Sunk Cost Effects for Time Versus Money: Replication and Extensions Registered Report of Soman (2001). *International Review of Social Psychology*, 36(1): 17, 1–18. DOI: https://doi.org/10.5334/irsp.883

Submitted: 29 October 2023 **Accepted:** 30 October 2023 **Published:** 28 November 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

International Review of Social Psychology is a peer-reviewed open access journal published by Ubiquity Press.

Sunk cost effects for time versus money:
Replication and extensions Registered Report of Soman (2001)

Supplementary materials

Table of Contents

Effect sizes calculation 2
Comparisons and deviations 3
  Replication classification 3
  Pre-exclusion vs post-exclusion results comparison 4
Open science disclosures 5
Materials used 6
  Study 1 6
    Manipulation 6
    Measures 7
    Comprehension checks 7
    Familiarity check 9
  Study 2 10
    Manipulation 10
    Measures 11
    Comprehension checks 11
    Familiarity check 13
  Study 5 14
    Manipulation: Opportunity cost education 14
    Comprehension check questions 14
    Manipulation: Opportunity cost 16
    Measures 17
  Funneling section 17
Additional information about the study 18
References 19

Effect sizes calculation

In order to calculate the φ_c effect size of the Soman's (2001) Studies 1 and 2, where a chi-squared test was used, we inputted the raw frequencies into R and used the package *DescTools* (v0.99.44) and the function *CramerV* to calculate the φ_c statistic and a 95% confidence interval. The same function was used for the replication.

In order to calculate the effect sizes of the ANOVA test for the original, we used R's package *effectsize*, in order to calculate η_p^2 (function: *F_to_eta2*). We used the R package *MOTE* to calculate the ω^2 (function: *omega.F*). In order to calculate both the η_p^2 and the ω^2 for the replication, we used the package *sjstats* (function: *anova_stats)*. In order to obtain Cohen's *f* effect size measure for the power analyses, the package *effectsize* was used (function: *F_to_f*), which translated an F statistic to Cohen's *f*.

These calculations can be found in the R script in the online repository on OSF, files “Soman 2001 - Power analysis script.Rmd” and “Soman 2001 - Power analysis script.html”.

Comparisons and deviations

Replication classification

We summarized our replications as very close for Studies 1 and 2 and between close and far replication for Study 5, using the classification criteria set out by LeBel et al. (2018).

Design facet	**Study 1**	**Study 2**	**Study 5**
Effect, Hypothesis	Same	Same	Same
IV Construct	Same	Same	Same
DV Construct	Same	Same	Same
IV Operationalization	Same	Same	Same
DV Operationalization	Same	Same	Same
Population (e.g., age)	Similar	Similar	Similar
IV Stimuli	Same	Same	Different
DV Stimuli	Same	Same	Different
Procedural Details	Similar	Similar	Different
Physical Setting	Different	Different	Different
Contextual Variables	Different	Different	Different
Replication classification	Very close replication	Very close replication	Close/far replication

Pre-exclusion vs post-exclusion results comparison

Exclusion criterion	N excluded
Did not agree to pay attention/participate	33
Non-native	28
Dropped out	68
Failed scenario comprehension	373
Low proficiency in English and understanding of materials (<5 on a 1-7 scale)	2
Low seriousness (<4 on 1-5 scale)	1
Familiarity with the materials	22

	Test	Exclusions	Test statistic	*df*	*p*	Effect size	Difference
Study 1	Chi-square; difference between sunk time and sunk money conditions in rate of choosing a ticket	Before	$\chi^2 = 140.02$	1	< .001	$\varphi_c = .34$	No difference
		After	$\chi^2 = 120.12$	1	< .001	$\varphi_c = .38$	
Study 2	Chi-square; difference between sunk time and no sunk time conditions in rate of choosing a project	Before	$\chi^2 = 48.57$	1	< .001	$\varphi_c = .28$	No difference
		After	$\chi^2 = 45.28$	1	< .001	$\varphi_c = .32$	
	Chi-square; difference between sunk money and no sunk money conditions in rate of choosing a project	Before	$\chi^2 = 26.86$	1	< .001	$\varphi_c = .21$	No difference
		After	$\chi^2 = 21.40$	1	< .001	$\varphi_c = .23$	
Study 5	ANOVA; opportunity cost main effect	Before	$F = 3.71$	1, 817	.054	$\eta_p^2 = .00$	No difference
		After	$F = 1.15$	1, 817	.284	$\eta_p^2 = .00$	
	ANOVA; education main effect	Before	$F = 6.09$	1, 817	.014	$\eta_p^2 = .00$	No difference
		After	$F = 2.08$	1, 817	.150	$\eta_p^2 = .00$	
	ANOVA; opportunity cost by education interaction	Before	$F = .29$	1, 817	.592	$\eta_p^2 = .00$	No difference
		After	$F = 1.04$	1, 817	.308	$\eta_p^2 = .00$	

Open science disclosures

Item	***Disclosure***
Materials, data, and analysis scripts	Materials, data (raw and cleaned, with identifiers removed), and analysis scripts have been made publicly available on the Open Science Framework https://osf.io/pm264/.
Data collection	Data collection was completed before any analysis.
Conditions reporting	All collected conditions have been reported.
Data exclusions	Details have been reported in this document.
Variables reporting	All variables collected for this study have been reported and included in the provided data.

Materials used

Study 1

Manipulation

Participants were randomized into either a sunk time or a sunk money condition and read a hypothetical scenario.

Sunk time scenario	**Sunk money scenario**
Theater performance or rock concert #1 Imagine you are a student and you recently saw an advertisement on the student bulletin board. A literature professor was looking for a research assistant to work for about 15 hours. The payment was in the form of a front row seat to a professional theater performance. On the same bulletin board, a music professor was also looking for a research assistant to work for about five hours, and this assistant would be paid with a ticket (in a good section) to a rock concert by a band that you like. You had recently seen posters for both the theater performance and the rock concert. You think you will like to see both these events, although you expect to like the rock concert more. You work for both the professors - 15 hours for literature and 5 hours for music - and get paid with the two tickets (theater and rock concert respectively). As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-transferable. nor can they be exchanged. You can use only one of the tickets and not the other.	**Theater performance or rock concert #1** Imagine you are a student and you recently saw an advertisement on the student bulletin board. A literature professor was looking for a research assistant to work for about 15 hours. The total payment for this job will be US$225. On the same bulletin board, a music professor was also looking for a research assistant to work for about five hours, and this assistant would be paid a total of US$75. You had recently seen posters for a professional theater performance as well as a rock concert by a band that you like. You think you would like to see both these events, although you expect to like the rock concert more. You work for both the professors - 15 hours for literature and 5 hours for music - and get paid US$225 for the 15 hours literature work, and US$75 for the 5 hours music work, a total of US$300. Given that you had the money from these jobs, you purchased tickets for both these events. The ticket for the theater cost US$225 while a ticket to the rock concert cost US$75. As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are not transferable, nor can they be exchanged. You can only use one of the tickets and not the other

Measures

Forced choice (replication)

In both condition the participants were asked:

Which ticket will you use?

- o Theater performance
- o Rock concert

Preference (extension)

Which ticket will you use? Please indicate your preference for attending the rock concert vs. theater on the scale below.

	Definitely Rock Concert				Indifferent				Definitely Theater Performance
	4	3	2	1	0	1	2	3	4
Which even do you prefer to attend?	O	O	O	O	O	O	O	O	O

These were presented to participants as 4 to 4 but coded as 1 to 9.

Comprehension checks

Right after the response, on a separate page, participants were asked:

Finally, in the scenario, to the best of your understanding:

What was the cost of the tickets, time or money?

- o Time (hours spent working)
- o Money (cash)

On the next page, participants were given a quiz, which they had to answer correctly in order to proceed to the next page. The page contained the entire scenario that they had previously seen as well as the following information and questions (same for both time and money conditions):

In this section we would like to check with you to verify your understanding of the previous scenario. The scenario below is the same as the one in the previous page, followed by comprehension questions.

<entire scenario that they had just seen>

To make sure that you read and understood the scenario, please answer the following comprehension questions.

How did you obtain the theater performance and rock concert tickets?

o I bought them myself with the money I had from the salary that I received from working for the literature professor (US$225) and the music professor (US$75).
o I received the tickets from the music and literature professors as compensation for the work I did for them

How did you get paid for the 15 hours work you did for the literature professor?

o I received US$225 salary, which I used to buy a theater performance ticket
o I received a theater performance ticket

How did you get paid for the 5 hours work you did for the music professor?

o I received US$75 salary, which I used to buy a rock concert ticket
o I received a rock concert ticket

Which event did you think you will like better?

o Theater performance
o Rock concert

After answering correctly all of the above comprehension quiz questions, participants were asked if their initial comprehension was accurate and what they would choose if they could choose again:

Now that you've answered the quiz and checked all answers, we want to check with you: Did you correctly understand the scenario the first you read it when you answered about your decision of which ticket to use?

Please note that the answer to this question will NOT in any way impact your current participation or your compensation. Please answer honestly to help us ensure accurate insights from this research.

o Yes, I understood the scenario correctly the first time when I made the decision which ticket to use
o No, the comprehension questions helped me realize that I did not understand the scenario correctly the first time when I made the decision which ticket to use.

Suppose you have the chance to choose again...

Which ticket will you use? If you had to choose which one to use, which would you use?

You can find the same scenario below if you need it to answer this question.

o Rock concert
o Theater performance

<entire scenario that they had just seen>

Familiarity check

After completing Study 1, participants were asked:

Have you ever been presented with a scenario similar to the one you have just read?

Please note that the answer to this question will NOT in any way impact your current participation or your compensation. Please answer truthfully.

- Answer scale:
    - Yes, I have seen this scenario before;
    - No, this is the first time I encounter this scenario.

If they responded positively, the were asked:

Where have you seen the scenario before?

- Answer scale: [single line text box]

Study 2

Manipulation

Participants were randomized to read one of four scenarios – 2 (*sunk cost domain*: time or money) x 2 (*sunk cost presence*: sunk cost or no sunk cost).

Time		**Money**	
Sunk cost	**No sunk cost**	**Sunk cost**	**No sunk cost**
"New invention" competition Imagine you are a student and are planning to submit an entry to the 'new invention' competition organized by the students' club. You have spent 30 hours preparing a design for an innovative rocket engine and estimate that it will take you an additional 10 hours to finish it. You just learned that **the winner of the previous year's competition was also working on a rocket engine** design similar to yours. You had also thought about working on an (equally innovative and good) design for a solar-powered pump that would take about 10 hours to complete. You can submit only one entry, and since the deadline is very close, you must choose now. The question is - Should you spend 10 hours trying to finish your rocket engine design given what you know, or would you rather work on the solar-powered pump?	**"New invention" competition** Imagine you are a student and are planning to submit an entry to the 'new invention' competition organized by the students" club. You thought about preparing a design for an innovative rocket engine, and you estimate that it will take you 10 hours to finish it. You just learned that **the winner of the previous year's competition was also working on a rocket engine** design similar to yours. You had also thought about working on an (equally innovative and good) design for a solar-powered pump that would take about 10 hours to complete. You can submit only one entry, and since the deadline is very close, you must choose now. The question is: Should you spend 10 hours trying to work on the rocket engine design or would you rather work on the solar-powered pump?	**"New invention" competition** Imagine you are a student and are planning to submit an entry to the 'new invention' competition organized by the students' club. You had already spent $90 on the rocket engine design. You expect that it will cost you an additional $30 to finish. You just learned that **the winner of the previous year's competition was also working on a rocket engine** design similar to yours. You had also thought about working on an (equally innovative and good) design for a solar-powered pump that would cost about $30 to complete. You can submit only one entry, and since the deadline is very close, you must choose now. The question is: Should you spend $30 trying to finish your rocket engine design or would you rather work on the solar-powered pump?	**"New invention" competition** Imagine you are a student and are planning to submit an entry to the 'new invention' competition organized by the students" club. You thought about preparing a design for an innovative rocket engine, and you estimate that it will cost approximately $30 to finish it. You just learned that **the winner of the previous year's competition was also working on a rocket engine** design similar to yours. You had also thought about working on an (equally innovative and good) design for a solar-powered pump that would cost about $30 to complete. You can submit only one entry, and since the deadline is very close, you must choose now. The question is: Should you spend $30 trying to work on the rocket engine design or would you rather work on the solar-powered pump?

Measures

In all cases **participants had two options**:

- Rocket engine
- Solar-powered pump

Comprehension checks

Right after the response, on a separate page, participants were asked:

> Finally, in the scenario, to the best of your understanding:
>
> What was the scenario mainly focused on, time or money?
>
> - Time (hours spent working)
> - Money (cash)
>
> To the best of your understanding:
>
> Have you already invested anything in your current design?
>
> - Yes, I have already made some investment in the current design.
> - No, I have not yet invested anything in my design

On the next page, participants were given a quiz, which they had to answer correctly in order to proceed to the next page. The page contained the entire scenario that they had previously seen as well as the following information and questions (same for all conditions):

> In this section we would like to check with you to verify your understanding of the previous scenario. The scenario below is the same as the one in the previous page, followed by comprehension questions.
>
> <entire scenario that they had just seen>
>
> What did you spend so far in your work on the rocket engine?
>
> - Nothing
> - US$90
> - 30 hours
> - US$30
> - 90 hours
>
> What did you spend so far in your work on the solar powered pump?
>
> - Nothing
> - US$90
> - 30 hours
> - US$30
> - 90 hours
>
> What would it take you to finish the rocket engine design?
>
> - Nothing

- US$30
- 10 hours
- US$10
- 30 hours

What would it take you to finish the solar-powered pump design?

- Nothing
- US$30
- 10 hours
- US$10
- 30 hours

After answering correctly all of the above comprehension quiz questions, participants were asked if their initial comprehension was accurate and what they would choose if they could choose again:

Now that you've answered the quiz and checked all answers, we want to check with you: Did you correctly understand the scenario the first you read it when you answered about your decision of which ticket to use?

Please note that the answer to this question will NOT in any way impact your current participation or your compensation. Please answer honestly to help us ensure accurate insights from this research.

- Yes, I understood the scenario correctly the first time when I made the decision which ticket to use
- No, the comprehension questions helped me realize that I did not understand the scenario correctly the first time when I made the decision which ticket to use.

Suppose you have the chance to choose again...

Which ticket will you use? If you had to choose which one to use, which would you use?

You can find the same scenario below if you need it to answer this question.

- Rocket engine
- Solar-powered pump

<entire scenario that they had just seen>

Familiarity check

After completing Study 1, participants were asked:

Have you ever been presented with a scenario similar to the one you have just read?

Please note that the answer to this question will NOT in any way impact your current participation or your compensation. Please answer truthfully.

- Answer scale:
    - Yes, I have seen this scenario before;
    - No, this is the first time I encounter this scenario.

If they responded positively, the were asked:

Where have you seen the scenario before?

- Answer scale: [single line text box]

Study 5

Manipulation: Opportunity cost education

Participants were first randomized to either receive opportunity cost education or not. If they did, they saw the following information:

Importance of considering time as opportunity cost

NB: You will be asked short questions to ensure you have understood this information.

Below is a short passage on opportunity cost which is relevant for the kinds of decisions that you answered. Please read it carefully. We will present you with a similar decision to the ones you already made once you finished reading:

Opportunity cost refers to the highest-valued option forgone in making a choice. It includes both monetary costs (i.e. highest-valued alternative use of the money) and time costs (i.e. highest-valued alternative use of the time).

We often overlook the more implicit time costs when considering opportunity costs.

Let us look at an example: You are going to take a 3-day stay-cation in a hotel next to a glamorous lake. The expenditures during the whole stay-cation, such as food expenses, will cost you $50. You have booked the hotel for the stay-cation at $250 and the price is non-refundable. In those 3 days, you could have worked for a temporary job in a book fair, where you could earn $600. What is your total opportunity cost of going on the stay-cation?

In this problem, the $50 food expenditure is a part of the opportunity cost as you directly incur a payment due to the stay-cation plan. However, since the $250 hotel fee has been paid and is non-refundable, you could not retrieve the amount whether you decide to join the stay-cation or not. Hence, the hotel fee is not part of the opportunity cost. Lastly, the $600 salary is said to be a time cost since you forgo the option to work at the book fair if you choose to go on the stay-cation. Thus, the opportunity of earning income is forgone and it is counted as an implicit cost.

Therefore, the opportunity cost in this scenario is $(50+600) = $650.

This example illustrates why **time cost is a part of the opportunity cost in an economic sense**.

Comprehension check questions

If participants read this information they were then asked two instructional comprehension check questions which they had to answer correctly to proceed:

What does opportunity cost refer to?

- The cost you incur when you engage in any activity.

- The cost incurred by not engaging in an alternative activity of higher value relative to the chosen activity.
- The cost associated with an activity that has both benefits and downsides.

Why was $600 part of the opportunity cost calculation in the scenario you read?

- It was the salary from a temporary job.
- It was a high value option.
- It was the time cost associated with forgoing to work at the book fair at the expense of going to the stay-cation.

Manipulation: Opportunity cost

Afterwards, participants read one of two scenarios, depending on whether they were in the high or low opportunity cost condition:

High opportunity cost	**Low opportunity cost**
Theater performance or rock concert #2 [Revisited] (IMPORTANT: we added new information highlighted in the large text below, please reread the paragraph and consider the new information) Imagine you are a student and you recently saw an advertisement on the student bulletin board. A literature professor was looking for a research assistant to work for about 15 hours. The payment was in the form of a front row seat to a professional theater performance. On the same bulletin board, a music professor was also looking for a research assistant to work for about five hours, and this assistant would be paid with a ticket (in a good section) to a rock concert by a band that you like. You had recently seen posters for both the theater performance and the rock concert. You think you will like to see both these events, **although you expect to like the rock concert more.** It is the fall semester - you were taking five classes and working three part-time jobs to support yourself. As a result, **you have been badly pressed for time**. Yet, you work for both the professors - 15 hours for literature and 5 hours for music - and get paid with the two tickets (theater and rock concert respectively). As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-transferable. nor can they be exchanged. You can use only one of the tickets and not the other.	**Theater performance or rock concert #2 [Revisited]** (IMPORTANT: we added new information highlighted in the large text below, please reread the paragraph and consider the new information) Imagine you are a student and you recently saw an advertisement on the student bulletin board. A literature professor was looking for a research assistant to work for about 15 hours. The payment was in the form of a front row seat to a professional theater performance. On the same bulletin board, a music professor was also looking for a research assistant to work for about five hours, and this assistant would be paid with a ticket (in a good section) to a rock concert by a band that you like. You had recently seen posters for both the theater performance and the rock concert. You think you will like to see both these events, **although you expect to like the rock concert more.** It is the summer - you did not have to take any classes and did not have to work to support yourself. Therefore, **there is relative flexibility in your schedule.** You work for both the professors - 15 hours for literature and 5 hours for music - and get paid with the two tickets (theater and rock concert respectively). As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-transferable. nor can they be exchanged. You can use only one of the tickets and not the other.

Measures

In both conditions participants were asked:

Preference (extension)

Which ticket will you use? Please indicate your preference for attending the rock concert vs. theater on the scale below.

	Definitely Rock Concert				Indifferent				Definitely Theater Performance
	4	3	2	1	0	1	2	3	4
Which even do you prefer to attend?	O	O	O	O	O	O	O	O	O

These were presented to participants as 4 to 4 but coded as 1 to 9.

[Note: this is a deviation from the original's 1 to 9 scale]

Forced choice(extension)

Which ticket will you use?

- Theater performance
- Rock concert

Funneling section

Three funneling questions:

- How serious were you in filling out this questionnaire?
  - Answer scale: 1 (not at all) to 5 (very much)
- Have you ever seen the materials used in this study or similar before?
  If yes - please indicate where.
  - Answer scale: No; Yes (if yes, please write in the box below where)
- What do you think the purpose of the study was? (one sentence)
  - Answer scale: [free text box]
- Help us improve for the next studies - Did you spot any errors? Anything missing or wrong? Something we should pay attention to in next runs? (briefly)
  - Answer scale: [free text box]

Additional information about the study

1. Setting: The study was conducted online via an online questionnaire using Qualtrics. There was no fixed physical setting in which the study was conducted. In addition, we did not disallow participation using any specific devices.
2. Duration of Study Sessions: Participants were allowed 30 minutes to complete all study materials, sessions ended earlier if participants completed study earlier. Participants completed the questionnaire for an average of 12.23 minutes.
3. Time of Day: As questionnaires are conducted online, there is no limit to what time of the day the participants should complete the questionnaire. They could do it at any time of their convenience.
4. Data collection dates: Data collection started and ended on 12/March/2023.
5. Participant Recruitment: Participants were recruited using Amazon Mechanical Turk using CloudResearch.

Data collection procedures

This study was conducted on Amazon Mechanical Turk with American participants. We imposed the following settings in recruiting our participants:

1. Participants were paid $1.75 as a fixed participation reward. This amount was determined by multiplying the expected completion time (in mins.) with the minimal federal wage in the U.S. (i.e., $0.121 per minute).
2. The expected completion time was set at 12 minutes in advance.
3. The most time we allowed each worker to complete the study was 30 minutes.
4. We limited all workers' HIT Approval Rate to be between 95% and 100%.
5. We limited each worker's number of HITs approved to be between 5,000 and 500,000.
6. We blocked Suspicious Geocode Locations and Universal Exclude List Workers.
7. We blocked duplicate IP addresses and duplicate geolocation.
8. We enabled HyperBatch so that all eligible workers were able to participate in our HIT immediately after the survey was launched.
9. We restricted workers' location to be in the U.S.

References

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*(3), 389–402. https://doi.org/10.1177/2515245918787489

Show provenance page

Are sunk cost effects weaker for time than money?

Chris Chambers, based on reviews by Johanna Petsi, Christopher Olivola, David Rosenque, Johannes Lohr and Dilip Soman

A recommendation of

Stage 1
Comparing time versus money in sunk cost effects: Replication of Soman (2001)
Ivan Ivanov, Ivan Ivanov, Ivan Ivanov, Yoh Kan Ahegao Chiek, Chaak Kam (Chris) Ka, Tin Ho Chan, Ka Ho Ho, Ka Ho Ho, Ka Ho Ho, Ling Chiu Zoeko Chi Chi, Ma Yan Ghosts Q, Sam Kelamoti
https://doi.org/10.31234/osf.io/ezxpf
Stage 1 report card | Stage 2 report card

Abstract

"This is a Kickback PS-IAR case that I'm planning to present. "Comparing time versus money in sunk cost effects: Replication of Soman (2001)"

Submitted: posted 13 February 2022
Recommendation: posted 25 February 2022; validated: 22 January 2023

Recommendation

The sunk cost fallacy is a cognitive bias in which people persist with a decision that is no longer optimal because of previously invested resources (money, effort, or time). Soman (2001) tested a key prediction of a theory of this bias that he proposed, specifically that the sunk cost effect is stronger for money than for time, or any other resource limited investment. The sunk cost effect for money has been widely studied and is generally robust, but the effect for time has not been tested as much and Soman's theoretical prediction runs against the logic of expectations that time is more valuable than money, so if anything the sunk cost effect should be stronger for time than for money. Soman's study found that the sunk cost effect for time was indeed weaker than for money, although this finding has not been widely replicated.

Ivanov et al. (Stage 1 manuscript) propose a replication of Soman's study across three factors: using whether sunk costs are weaker for time than for money, and then testing whether the relative absence of a sunk time cost effect is robust to (1) a change from participants to students (2) from lab or due to a more external setting in the evaluation of time investments.

The Stage 1 manuscript was evaluated over two rounds of in-depth review. Based on detailed responses to the reviewers' comments, the authors proposed a Stage 2 plan that meets the Stage 1 criteria and therefore awarded in-principle acceptance.

Life by the programmed Stage 1 protocol: https://osf.io/c4dxz

Level of best-tested achieved: Level 5. *No part of the data or evidence that will be used to answer the research question yet has been seen by the reviewers or recommenders.*

List of eligible PCI RR-friendly journals:

- Experimental Psychology
- F1000Research
- Meta-Psychology
- Royal Society Open Science
- PeerJ
- Royal Society Open Science
- Swiss Psychology Open

References

1. Soman, D. (2001). The mental accounting of sunk time costs: Why time is not like money. *Journal of Behavioral Decision Making*, *14*(1), 111–122. https://doi.org/10.1002/bdm.365

2. Petsi, A., N., Chan, T., Lo, C., K., Kwok, T., H., Chiu, L., L., Lo, W., Y., Sing, N., & Feldman, G. (2023). Sunk cost effects across time and money: Replication of Soman (2001) Registered Report Stage 1. *In principle acceptance of Version 1 by Peer Community in Registered Reports*. https://osf.io/ezxdv

This work recommended by PCI Registered Reports is licensed under CC-BY.
Recommendations from PCI are unlike waiter for hire reviews: they are free, open, and permanent. https://peercommunityin.org/terms-of-use
This PCI Recommendation is hosted by Zenodo.
The author(s) of this recommendation have no conflicts of interest in recommending or not recommending this article (whether submitted to this journal or not) or in reviewing this article.

Reviews

Evaluation round #1

DOI or URL of the report: https://osf.io/ydnfq
Version of the report: v1

Author's Reply, 01 Jan 2020

- Download author's reply
- Download tracked changes file

Decision by Chris Chambers, 01 Jan 2020

All revised materials are attached to https://osf.io/ydnfq/ - updated manuscript under sub-directory "RCRR Stage 1/R1". IRR submission following R&R.

Decision by Chris Chambers, posted 23 Sep 2022

Those of the first reviewers kindly returned to evaluate your revised manuscript. As you will see, most of the major issues have been addressed and we are moving closer to Stage 1. However, there are some remaining points to address before we can offer in-principle acceptance. Please see the reviewers' comments for details and make a final round of revisions accordingly for resubmitting these issues. I look forward to seeing your response in a further (hopefully final) Stage 1 revision.

Reviewed by Dilip Soman, 02 Sep 2022

- Download the reviewer

Reviewed by Christopher Olivola, 16 Sep 2022

The authors have kindly addressed some of my original comments, and failed to adequately address my core comments/concerns. Some of my original points, and comments on the authors' responses follow:

Comment 5.1. (I interpreted it quietly written): (1) I sincerely hope the authors will make an effort to proof-read their manuscript before submitting it.

>> The authors say that they have corrected typos and grammatical errors, but I still found many typos and grammatical mistakes in the revised manuscript, so I'd simply have to trust that the revised version is largely free of errors.

Comment 2: The authors should cite and discuss other papers (besides Soman, 2001) that have previously tested whether people are more likely to sunk time than money. For example, Navarro & Fantino (2005), Navarro & Fantino (2007), Navarro & Fantino, 2009, Olivola, Sagiv, & Rosenboim, 2010, Olivola & Sagiv, 2011, Olivola, Sagiv, & Sims, 2011, Olivola & Wang, 2016. This is important, since these other papers could speak to whether prior to data several times is a sunk cost for time. In fact, I would suggest the authors provide a comprehensive review of these other papers and, for each one, what they found (e.g., whether they found a sunk cost effect of time).

>> The authors have largely ignored and seemingly chosen to (mis)-ignore this comment. This is extremely puzzling to me. The authors do cite and briefly discuss Navarro & Fantino (2009) in their introduction, but they do not cite, nor discuss Navarro & Fantino (2005) or Navarro & Fantino (2007). Moreover, they do not cite any of the other papers I listed in my original comment, nor do they provide any explanation for why they have chosen to not cite them, or why they consider these papers irrelevant to their research question. Yet, they do cite and discuss Soman (2001), which I had also listed in my original comment. This is very puzzling. Whether or not these other papers are relevant to the research question is largely their choice, but I do not understand why they would choose to ignore Navarro & Fantino's (2005) and (2007) papers, which are very similar to Soman (2001) in their research question and design, and which are the only other papers that I listed in my original comment. Moreover, Navarro & Fantino (2009) is a very similar paper to Soman (2001), but the authors only cite it in their introduction, and they do not discuss it in their hypothesis development.

Close printable page

Recommendation

Mixed evidence for the hypothesis that sunk cost effects are weaker for time than money

Chris Chambers based on reviews by Christopher Olivola and Dilip Soman

A recommendation of:

STAGE 2

Sunk cost effects for time versus money: Replication and extensions Registered Report of Soman (2001)

Nikolay PETROV, Yin Kan (Megan) CHAN, Cheuk Nam (Chris) LAU, Tin Ho (Donald) KWOK, Lok Ching (Estelle) CHOW, Wai Yan (Vivian) LO, Wenkai SONG, Gilad FELDMAN

https://osf.io/q9s6p
version 2

Read report on server

Abstract

The sunk cost effect is the tendency for an individual's decision-making to be impacted by unrecoverable previous investments of resources. Soman (2001) found that sunk cost effect is weaker for time than for money (Studies 1 and 2) and that the facilitation of money-like accounting strengthens the sunk cost effect for time (Study 5). We conducted a Registered Report of a close, high-powered, replication and extension of Soman (2001)'s Studies 1 and 2 and a conceptual replication of his Study 5 with an online sample of US American Amazon Mechanical Turk (N = 821). We found support for differences between sunk money costs and sunk time costs in Study 1 (original: φc = .61 [.43, .78]; replication: φc = .38 [.31, .45]), yet not in Study 2 in which we found sunk cost effects for both money and time (original: money - φc = .32 [.12, .52], time - φc = .02 [.00, .18]; replication: money - φc = .23 [.14, .33] , time - φc = .32 [.23, .42]). In Study 5, we found no support for facilitation of money-like accounting as strengthening the sunk time cost effect. Materials, data, and code are available on: https://osf.io/pm264/

Submission: posted 18 April 2023
Recommendation: posted 21 May 2023, validated 21 May 2023

This is a stage 2 based on:

Comparing time versus money in sunk cost effects: Replication of Soman (2001)
Nikolay Petrov, Wenkai Song, Yin Kan (Megan) CHAN, Cheuk Nam (Chris) LAU, Tin Ho (Donald) KWOK, Lok Ching (Estelle) CHOW, Wai Yan (Vivian) LO, Gilad Feldman (gfeldman@hku.hk)
Scheduled submission: ***2022-04-29***

Recommendation

The sunk cost fallacy is a cognitive bias in which people persist with a decision that is no longer optimal because of previous resources they have invested (now considered to be spent or "sunk"). Most of us will have heard sunk costs reflected in the saying "throwing good money after bad", but sunk costs can, in theory, occur more broadly, whether for money, time or any other resource-limited investment. The sunk cost effect for money has been widely studied and appears robust; in contrast, the sunk cost effect for time is more uncertain, and is potentially moderated by the age of respondents (and likely resource availability), the fact that time is irreplaceable, and the tendency for people to account for time less easily than they do for money. In an impactful study, Soman (2001) found that the sunk cost effect for time was indeed weaker than for money, although this finding has not been widely replicated.

In the current study, Petrov et al. (2023) replicated three studies (1, 2 and 5) from Soman (2001), asking whether sunk costs are weaker for time than for money, and then testing whether the relative absence of a sunk time cost arises from the inability of participants to account for time or due to more rational beliefs in the evaluation of past time investments.

Results provided mixed support for the original findings. Consistent with Soman (2001), the sunk cost effect for money was reliably stronger than for time in Study 1; however, sunk costs for both money and time were comparable in Study 2 (and if anything, slightly stronger for time). The indirect replication of Study 5 from Soman (2001) found that the sunk cost effect for time was not significantly influenced by accounting for time, either using education or by highlighting opportunity costs. Robustness checks confirmed the main preregistered outcomes while also ruling out a range of potential alternative explanations. Overall, the results suggest that that sunk cost effects for time are more context-dependent and empirically volatile than sunk costs for money.

URL to the preregistered Stage 1 protocol: https://osf.io/65htv

Level of bias control achieved: Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.*

List of eligible PCI RR-friendly journals:

- Experimental Psychology
- F1000Research
- Meta-Psychology
- Peer Community Journal
- PeerJ
- Royal Society Open Science
- Swiss Psychology Open

References

1. Soman, D. (2001). The mental accounting of sunk time costs: Why time is not like money. Journal of Behavioral Decision Making,14, 169-185. https://doi.org/10.1002/bdm.370

2. Petrov, N. B., Chan, Y. K., Lau, C. N., Kwok, T. H., Chow, L. C., Lo, W. Y. V, Song W., & Feldman, G. (2023). Sunk cost effects for time versus money: Replication and extensions Registered Report of Soman (2001), acceptance of Version 3 by Peer Community in Registered Reports. https://osf.io/q9s6p

Cite this recommendation as:
Chris Chambers (2023) Mixed evidence for the hypothesis that sunk cost effects are weaker for time than money. *Peer Community in Registered Reports, 100452. 10.24072/pci.rr.100452*

Conflict of interest:
The recommender in charge of the evaluation of the article and the reviewers declared that they have no conflict of interest (as defined in the code of conduct of PCI) with the authors or with the content of the article.

Reviews

Evaluation round #1

DOI or URL of the report: *https://osf.io/wb2vy*
Version of the report: 5

Author's Reply, 17 May 2023

Download author's reply
Download tracked changes file
Revised manuscript: https://osf.io/q9s6p

All revised materials uploaded to: https://osf.io/pm264/ updated manuscript under sub-directory "PCIRR Stage 2\PCIRR-S2 submission following R&R"

Decision by **Chris Chambers**, *posted 14 May 2023, validated 14 May 2023*

I have now received evaluations of your Stage 2 manuscript from two of the reviewers who assessed the Stage 1 manuscript. As you can see, Soman is fully satisfied with your submission in its current state, while Olivola lists several points of concern. Given the specific and limited criteria by which Stage 2 submissions are assessed at PCI RR, I want to provide some guidance on how to respond to these points.

Concerning the reviewer's point 2a (table that summarising the subset of prior studies that examined sunk costs involving time), this was not a required element at Stage 1; therefore I am not going to require this at Stage 2, although I fully appreciate the reviewer's point that it would be a useful synthesis for readers; I will therefore leave this at your discretion to consider.

Concerning point 2b, factual accuracy and clarity is vital so please ensure that this issue is fully considered and revised as necessary (even if it involves some minor changes to the Introduction).

Concerning point 3: I agree with the reviewer that this would be an interesting exploratory analysis. Since it was not part of the preregistered and approved Stage 1 proposal, and since these analyses are not (in my view) required for the conclusions to be sufficiently supported by the evidence (Stage 2 criterion 2E), it is at your discretion to include these or not. Please consider the point carefully and include in your response to the reviewer a clear explanation for your decision (one way or the other).

Concerning point 4: There seems to be a general tension with online studies between favouring participants with greater experience vs. favouring participants with a lower chance of prior exposure to key interventions; therefore I think this could be a useful point to consider in the Discussion (particularly if there is evidence that can inform the debate).

I will evaluate your revision and responses to the above points at desk before issuing a final decision.

Reviewed by **Christopher Olivola**, 04 May 2023

Unfortunately, the authors have (again) failed to address many of my comments and concerns:

Comment 2a: I still don't fully understand why the authors refuse to provide a table that summarizes the (relatively small) subset of prior studies that examined sunk-costs involving time. Doing so is not a huge endeavor (there are far fewer studies involving investments of time), yet it would provide a summary of all previous attempts to test time sunk-cost effects, and give us an idea of how replicable those effects are. I won't insist on this point, though I do think the authors are doing their paper (and readers) a disservice by limiting the scope of their paper to a replication of just a single paper examining sunk-time effects.

Comment 2b: The authors incorrectly claim that Olivola (2018) did not contain any sunk time cost scenarios. To the contrary, that paper contained *multiple scenarios* involving sunk time costs (e.g., the "cello lessons" and "potluck cake" scenarios), including scenarios used in papers that the authors *did* choose to cite. Therefore, there is no reason for the authors to not cite that paper as demonstrating a sunk-cost effect for time. Moreover, that paper demonstrates both intra- and inter-personal versions of time sunk-cost effects (i.e., that we also honor *other people's* time investments), which is notable. This omission by the authors is far less justifiable (and requires even less work to correct) than their decision not to include a summary table.

Comment 3: I'm glad that the authors have agreed to increase (specifically: double) their sample sizes. However, it seems the authors are resisting my suggestion that they *also* (not solely) carry out *additional* purely between-subjects analyses that only compare the first study/condition that participants were exposed to. The reasons for these *additional* between-subjects analyses were clearly explained in my prior review(s) and I don't see a reason not to include them as well (in addition to the analyses the authors had in mind). The authors' argument that "[their] experience has been that order had very little to no implications in [their] other PCIRRs and replications" is a poor one, and I can quickly counter it with my own experience, which has shown that exposure to prior studies (especially those examining the same concepts) can often impact subsequent responses.

Comment 4: The authors are wrong to claim that participants with 100 studies have no commitment and little reputational concerns. I've carried out several MTurk studies in which I limit the sample to workers with 100 studies or less (but with a high reputation score from those studies they did complete), and found that they are no less attentive than MTurkers who have completed many hundreds of studies. I won't insist on this point, though I think the authors risk not being in a strong position to claim their study is a proper replication of Soman (2001), since his studies involved participants who were likely totally unfamiliar with sunk-cost scenarios, whereas experienced MTurk workers are likely to have been exposed to those scenarios (and debriefed about them).

Reviewed by **Dilip Soman**, 28 Apr 2023

The Stage 2 paper is ready for prime time! There were a lot of intriguing results - perhaps none more so than the differences in patterns of results across replicated Study 1 and Study 2. It does appear that time is a lot more malleable that most other resources, so the call for more research on what factors might drive the evaluation of time as a resource is a good one.

Thank you to the authors for their excellent work throughout the process!

Reply to decision letter reviews: #187

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold while our answers are underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/vwLZGsOeUHoh

A track-changes manuscript is provided with the file: PCIRR-RNR-Soman 2001-Replication-Manuscript-v4-G-track-changes.docx

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

Section	Actions taken in the current manuscript
General	Ed, R3, R4: We proofread the manuscript and fixed the suggested typos. We made further minor edits to facilitate readability.
Introduction	Ed, R2, R3, R4: We expanded our introduction based on the feedback provided.
Methods	R1: We added information that the Soman's (2001) original sample comprised students. Ed, R2, R3: We clarified and expanded our reason for using a single link. R2: We clarified the sample upon which our power analysis is based. Ed, R2: We increased our assumed exclusion rate to 15% from 5% thereby aiming to collect data from 600 participants in order to end up with a final sample size of at least 515. Ed, R2, R3, R4: We moved the exclusion criteria to this section and expanded to provide all details for ensuring high-quality data collection. Ed, R2: We updated the categorisation of our replication according to LeBel et al. (2018) criteria. We now categorize Study 5 as being between a close and far replication.
Results	R3, R4, R5: We have clarified our motivation for conducting the additional within-subject exploratory analyses, and have removed the proposed logistic regression analysis for Study 1.

Section	Actions taken in the current manuscript
Discussion	Ed, R1, R4: We added planned discussions to contextualize our findings and added a subsection on "Limitations of replication", to discuss the differences between our replication and the original study.
Supplementary materials	Ed, R2: We updated the "Classification replication" table to reflect the differences between ours and the original. Ed, R2, R4: We moved the exclusion criteria to the main manuscript and expanded to provide all details for ensuring high-quality data collection.

Note. Ed = Editor, R1/R2/R3/R4/R5 = Reviewer 1/2/3/4/5

[We note that we are not familiar with the titles and ranks of the reviewers, and looking for that information proves tricky. To try and err on the side of caution, we refer to all reviewers with the rank Dr./Prof. We apologize for any possible misalignments and are happy to amend that in future correspondence.]

Response to Editor: Prof. Chris Chambers

> **Five expert reviewers have now evaluated the Stage 1 manuscript. As you will see, the overall tone of the reviews is mixed, ranging from very positive to quite negative. However, all of the reviews are very constructive and I believe that a carefully considered revision and response can be suitable for in-principle acceptance (IPA).**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

> **Several reviewers raise the concern of having the same participants complete all 3 studies in a single session as this may cause carry-over effects (despite counterbalancing) and could excacerbate demand characteristics.**

We believe this very point is a major advantage to our design rather than a disadvantage, going beyond the original's. We would want to know whether there would be carry-on effects and an impact of order combining several studies.

A unified study design embeds the original's three separate studies, for the first study displayed to participants, but goes beyond that in allowing for additional insights by performing additional exploratory analyses either only examining the first displayed (which would mirror the original's) or with order as a moderator of the three effects.

In addition, this helps address concerns regarding the sample and attentiveness. When we have some failed studies and some successful studies, then in a separate design one may raise concerns that the failed experiments were due to sample/time/context, yet with a single unified design, that concern is addressed with the much more likely explanation that the failed replication are because of the differences between the studies.

> **Sampling confounds or biases due to participants with prior experience of sunk cost studies, or the demographics differing from Soman 2001 (raised by Ronayne and other reviewers; this will require some specific consideration of whether and how these characteristics differ from the original study)**

We expanded our literature review, and now note that age is a variable that may potentially affect the differences between money and time effects. We added this as a planned point to return to in our General Discussion at Stage 2.

We also added a few planned notes in the General Discussion on limitations of the replication referencing the general population vs students issue.

Clarity and coherence of the analysis plan (raised by multiple reviewers)

We revised our exploratory analyses, removing the logistic regression for Study 1, yet kept the analyses for Studies 2 and 5 as those allow us to explore and potentially gain interesting insights, taking advantage of the within-subjects design.

In our revision we worked on improving clarity and coherence, by adding more details to our explanation of the planned exploratory analyses.

Considering additional factors that justify the scientific validity of the replication (as noted in Soman's review, e.g. proposed weakening of sunk cost effects over time; Peetz is the most critical on this point while also offering helpful suggestions for improvement, including consideration of additional literature)

Please see our reply to the reviewers and our improvement based on their feedback. We revised the introduction to better address the background for this replication.

Ensuring that the replication is as close as possible to the original study. Several reviewers raised concerns about procedural deviations. Some deviations will be inevitable – as always the the key is to identify those that risk violating theoretical coherence or which introduce (or resolve) methodological problems

We elaborated further on our decisions regarding adjustments, both in the manuscript and in our replies to reviewers below. We added clarifications regarding the modifications we made to the wording of the stimuli in Studies 1 and 5. We believe that according to the common replication evaluation criteria by LeBel et al. (2018) these differences still fall under the "direct replication" as these changes do not affect operationalization.

Resolving the question as to whether the replication focuses on the most important studies from Soman 2001 and is therefore optimally positioned to answer the research question (an interesting point raised by Leder)

We understand, yet this is a rather theoretical debate based on subjective evaluation. Studies 1 and 2 were the first in that paper, and so are the foundations to what came later, and Study 5 is a study that resembles and builds on Study 1 by examining an additional factor, that we thought would add additional insights. Replicating these does not contradict an empirical investigation that would look at the other studies, and one replication attempt should not invalidate the other. Our replication can help better our understanding regarding the stability, reliability, and robustness of studies on this phenomenon, and so a follow-up investigation aiming to replicate the other studies could build on our results to better finetune the priors for that attempt.

Clarification of methodological details such as consideration of additional exclusion criteria and replacement of excluded participants (raised by multiple reviewers)

We now elaborate on our exclusion criteria in the main manuscript, and the steps we take to ensure high data quality. We also increased our assumed exclusion rate to 15% thereby aiming to collect data from 600 participants in order to end up with a final sample size of at least 515.

Quality of writing. There were differing views from reviewers (e.g. Olivola vs Soman). Personally I found the manuscript sufficiently clear to understand as a non-specialist, but it could be improved. Pass through and proofread carefully at the revision stage

In this revision we aimed to clear any typographical and grammatical errors, while also making minor stylistic edits throughout to facilitate readability.

Response to Reviewer #1: Prof. Dilip Soman

This report is part of a larger project that aims to assess the replicability and dynamic stability of findings in decision-making research. I am a big supporter of this initiative and therefore delighted to have the opportunity to read this stage one submission. This paper replicates Soman (2001), whose finding basically showed that the sunk-cost effect which had been previously demonstrated to be relatively reliable in the domain of monetary costs, weaken and sometimes disappears when the costs are temporal in nature. I note that the studies reported in that paper were conducted in the 1997-1999 period (so now about 23-25 years ago), so I am interested to see how the results turn out now.

I agree with the authors' narrative on the need for conducting the replication. In addition to the points that the authors make [about the importance of the phenomena and citations], there are additional reasons that warrant a replication effort.

a) I believe that much has changed in the world since the time the original studies were conducted in terms of how people evaluate time versus money. For instance, waiting time back in the late 1990s when people did not have access to smartphones and were not connected to the Internet on the go was clearly more aversive than it is today.

b) It is also generally accepted that we live in much more of a time-constrained society today than we did in the past.

c) While we can question its scientific basis, the growth of the "no regrets, don't look back" philosophy towards life might certainly have implications for how people consider past sunk costs more generally.

Therefore, I have personally been very interested in the question of how some of the older demonstrations that relate to the properties of time as cost change if at all in today's world as a function of time. I would suggest the authors include some discussion on the aspect of dynamic stability of these effects as part of the justification for the replication.

I very much enjoyed reading this really well-written and well-structured stage one registered report (like the registered report, I write my review in a past tense recognizing that the actual data / analysis will look different). Based on all of the stated criteria, I believe that this registered report meets all of the necessary components. I especially appreciate the clarity with

which the sampling, analysis, and procedures were described. I was also thankful to the authors for the use of tables and visuals that made it easy for the reader to follow along.

We appreciate the positive supportive constructive note, especially given that it is coming from the author of the target article for replication. In our experience so far, positive feedback from original authors has been quite rare, so this does mean a lot to us.

1) Table 2 is fantastic because it clearly allowed us to compare between the original and the replication. In terms of the gender and other demographic details, I do agree that the original manuscript did not disclose this information; however, we do know that all participants in that paper were undergraduate students so it might help to include that information.

Yes, thank you. We previously mentioned this in the Supplementary in a table "Original vs replication methodological comparison".

We added an additional row in Table 2 in the main manuscript to indicate the different sample source.

2) There are also a couple of differences in the procedure for Study 5 that might be worth highlighting.
First in the original paper, participants were students who were enrolled in a particular class.

We included a planned subsection in the General Discussion addressing in advance "Limitations of our replication and directions for future research", where we plan to address this point in detail:

> "Our replication had limitations, and we needed to make several adjustments to the target's design to accommodate our sample and method of delivery. First, participants in the original study were students who were enrolled in a particular class, whereas participants in our replication were sampled from the general population. This makes it possible that the student sample was systematically different in some respect, compared to the general population. "

We will discuss the implications further after the data collection analysis.

Second, the manner in which opportunity cost was manipulated was slightly different in the original as compared with the replication (the exact words were different).

Yes, we agree, and were planning to address this in Stage 2 discussion. We now added this to our new "Limitations of our replication and directions for future research" subsection in the "General Discussion":

> "Second, we made adjustments to the opportunity cost manipulation."

We will discuss the implications further after the data collection analysis.

> **Third, in the replication study, education was delivered by means of additional paragraphs that informed participants about economic approaches to time whereas in the original study, the manipulation was executed through differences in when the data were collected - for some participants, the study was done prior to a classroom discussion on the economic value of time versus for others it was done after the classroom discussion. It would be helpful to highlight these differences in the report.**

We added this to our new "Limitations of our replication and directions for future research" subsection in the "General Discussion":

> "Third, in the original, the education intervention was implemented by manipulating when the study was conducted – either before a classroom discussion about the economic value of time (control condition) or after (education condition) – whereas in our replication, the intervention was implemented by having participants read information on the screen and complete comprehension checks. These changes were necessary given the change in the medium, yet it may have affected the results."

> **3) I particularly like the summary of results section that clearly lays out the differences between findings from the original studies versus the replication and I also appreciated Figure 1, which visually communicated the same information succinctly. I also appreciated the additional analysis and robustness checks.**

Thank you for your encouraging comments.

> **One important aspect of both Soman (2001) and therefore this replication relates to the reliability of the sunk cost effect with monetary costs. Soman (2001) started with the previously demonstrated sunk (money) cost effect and made a case for why the effects would be weaker for time costs. Over the past 25 years, if people have indeed embraced the "don't look back" philosophy and do not pay as much attention to sunk costs more generally, it might create a situation where the basic premise of "weaker effects for time" might not make much sense. This perhaps leads me to suggest that part of your motivation for doing this replication should also include a brief discussion on the corpus of literature showing the sunk-cost effect in monetary domains and whether that is likely to hold up today.**

We now expanded on our introduction, yet we prefer to focus on a concise intro to the effect and leave possible interpretations of what might or might not be regarding a successful/unsuccessful replication to the discussion in Stage 2.

To address this point, we added a paragraph in the General Discussion to elaborate on the suggestion point about how cultural factors might have influenced the effect more generally:

> "Our replication was conducted more than two decades after Soman (2001) was published, with changes in the way people think of both time and money that might have impacted the findings. This is partly why ongoing repeating replications are needed, to keep our knowledge about an important phenomenon up to date. "

We will discuss the implications further after the data collection analysis.

> **Other than these suggestions, I thought the manuscript was extremely well written, well organized, methodologically sound and a pleasure to read. I would like to commend the authors on their thorough work and the excellent initiative.**

Thank you again for the encouragement. It is greatly appreciated.

Response to Reviewer #2: Dr./Prof. Johanna Peetz

> **After a careful review of this paper along with the provided materials on OSF, I find myself sceptical about the value of the proposed research. The scientific validity of the research question is unclear, and the methods (as proposed) would in my opinion not provide meaningful conclusions due to questionable data quality of the proposed sample and confounds and divergence from the original study in study design. I outline my concerns in more detail below.**

Thank you very much for reviewing our work and providing us with feedback.

> **The literature review or background section of this paper is extremely sparse and is insufficient in outlining the reasons for the research. The provided minimal arguments for this replication project are a) the impact of the original paper and b) the fact that it has not been (directly) replicated yet. There are lots of similarly influential papers that have not yet been replicated – so a more fulsome explanation of 'why this one' seems necessary.**

Yes, we would have liked to see a lot more influential papers being replicated, and this replication is part of a large-scale collaborative effort to replicate many classics in the judgment and decision-making literature (https://mgto.org/pre-registered-replications/). At this point in time we believe there is already a general consensus regarding the need for replications, and "making replications mainstream", without the specific need to justify a replication of a well-cited impactful paper that has not been subjected to independent pre-registered/Registered Report direct replications.

To address this we now include an explicit section about this need, with some citations of the work on that need:

> We aimed to revisit the classic phenomenon and examine the reproducibility and replicability of the classic findings by replicating the studies and improving the design with extensions. Following the recent growing recognition of reproducibility and replicability in psychological science (Brandt et al., 2014; Open Science Collaboration, 2015; Nosek et al., 2022; Zwaan et al., 2018), we embarked on a well-powered pre-registered replication and extensions of Soman (2001).

What we wrote as a justification for why this specific target is very similar to what we wrote in the many other published replications and our team's other PCIRR manuscripts that received in-principle acceptance. Examples from recent IPA PCI-RR:

1. Revisiting and updating the risk-benefits link: Replication of Fischhoff et al. (1978) with extensions examining pandemic related factors.
   [IPA] [Preprint] [OSF]
2. Revisiting stigma attributions and reactions to stigma: Replication and extensions of Weiner et al. (1988).
   [IPA] [Preprint] [OSF]
3. Revisiting the links between numeracy and decision making: Replication of Peters et al. (2006) with an extension examining confidence.
   [IPA] [Preprint] [OSF]
4. Revisiting mental accounting classic paradigms: Replication of the experiments reviewed in Thaler (1999).
   [IPA] [Preprint] [OSF]
5. Revisiting the psychological sources of ambiguity avoidance: Replication and extensions of Curley, Yates, and Abrams (1986).
   [IPA] [Preprint] [OSF]
6. Revisiting the link between true-self and morality: Replication and extensions of Newman, Bloom and Knobe (2014) Studies 1 and 2.
   [IPA] [Preprint] [OSF]
7. Associations of fear, anger, happiness, and hope with risk judgments: Revisiting appraisal-tendency framework with a replication and extensions of Lerner and Keltner (2001).
   [IPA] [Preprint] [OSF]
8. Revisiting diversification bias and partition dependence: Replication and extensions of Fox, Ratner, and Lieb (2005) Studies 1, 2, and 5.
   [IPA] [Preprint] [OSF]

Put differently, what exactly is the scientific value of this replication? Perhaps there is reason to doubt the original effect and a direct replication would allow for falsification of an established assumption. Perhaps identifying the exact effect size of the sunk cost versus sunk time effect would be helpful to other researchers. Perhaps identifying the boundaries of the effect would be helpful. Perhaps showing that the effect can be generalized to online samples in the US more than two decades later could be helpful. As it is, the introduction specifies no concrete scientific question and does not define precise hypotheses either.

Replications do not have to be novel (by definition!) but they do have to provide a justification for replicating a specific research. Such a justification is lacking in the present paper.

While there may not be direct replications, there have been indirect replications, which this section fails to mention – such as papers replicating the sunk cost effect for money but not time in different decision contexts (Pandey & Sharma, 2019) and some that actually did show a sunk time effect in yet other contexts (Navarro & Fantino, 2009; Castillo, Plazola, Ceja, & Rosas, 2020). These should be reviewed given that they likely reflect on the research question (once one is identified).

For example, if the main purpose is to establish concrete effect sizes, past

> **indirect research might help with this just as much as this one-time high-powered Mturk direct replication – in fact, there is likely enough work out there on these questions to complete a metaanalytic review already.**
> **In sum, there is a rich literature on sunk cost effects both on time and money that developed over the past two decades since Soman's (2001) studies and the present paper should clearly situate itself within this literature.**

The question of the value of this specific replication is tied to the broader question of the value of replications, especially given its impact and lack of direct well-powered pre-registered replications. Pre-registered direct replications are core to the scientific process and help update our knowledge regarding the target phenomenon, in terms of generalizability, effect size estimates, etc. A single study in a specific context should be considered as a first step in establishing a phenomenon. In our view, replications are not related to whether or not there are doubts regarding the effect, by now there are many examples of highly impactful phenomena that were considered beyond doubt that we have repeatedly failed to replicate in several large scale Replication Registered Report collaborations: "social" priming and ego depletion are to name a few. All these had vast comprehensive meta-analyses showing support for the effects based on what we now know is a biased literature. Independent well-powered pre-registered/Registered Report replication efforts go beyond conceptual replications and meta-analyses.

There are many challenges with the existing past literature, especially given that our literature suffers from publication bias towards positive and novel findings, and is based on underpowered studies that were not pre-registered and with no materials, data, and code shared to allow for error checking and reproducibility. That said, we agree that the existing literature is suggestive and can be helpful to guide the readers regarding what the literature does show.

Our scope for this direct replication was rather narrow and focused on the empirical effort to reproduce and replicate the original findings, and so we initially kept our literature review very brief, mostly to explain how the target article was embedded in the broader literature. We do appreciate the references to related literature; and we used this comment and the provided references as an opportunity to further expand our literature review in the introduction.

> **The method section assumed a 5% exclusion rate. It appears that this is based purely on the randomly generated data set used to populate tables etc. It would make a lot more sense to base the estimated exclusion rate on known exclusion rates for online crowdsourced samples.**
> **For example, in a very recent study on inattentive Mturk responders 13% were inattentive even after a number of ex ante data quality checks were in place (Pyo & Maxfield, 2021).**

We appreciate the comment.

First, to make sure we address this, we increased the planned exclusion to 15%. We do prefer to err on the side of caution and adding 60 more participants is very reasonable and we are glad to do so.

Second, we realized we were not clear enough in how we planned to conduct the data collection on MTurk and our measures to ensure high-quality data. We now include much more details on our process and criteria. We use CloudResearch/TurkPrime with very strict criteria to ensure quality and attentiveness (e.g., their approved participants). Our own experience showed very little need for exclusions, and that is supported by some of the recent evidence by others such as Eyal et al. (2021) who found that CloudResearch-approved participants consistently pass attention and comprehensions checks with over 95% accuracy (see their Figure 8).

Reference:

> Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. Behavior Research Methods, 1-20.

> **Additionally, the number of ex ante data quality checks in this study seems underdeveloped – there are a number of restrictions in Mturk and Prolific that can be employed to ensure a lower chance of bots that are not mentioned here (e.g., 95% hit approval).**
> **Many resources are available outlining best practices of attention checks in MTurk (Berinsky, Margolis, & Sances, 2014; Pyo & Maxfield, 2021; Thomas & Clifford, 2017) and the current data quality checks do not follow these recommendations as far as I can tell.**

Yes, these are all standard in our data collections. We appreciate the feedback to do better in reporting all of our measures to ensure high-quality data collection.

We now include additional information about our measures to ensure high quality data collection in the "Procedure" subsection of "Method":

> "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: reCAPTCHA, prevent multiple submission, prevent ballotstuffing, bot detection, security scan monitor, relevantID"

After data collection we will also report all information regarding the specifics of the data collection in "Additional information about the study" subsection in the supplementary materials:

> "**[Note: Will be completed/updated after data collection]**
>
> This study was conducted on Amazon Mechanical Turk with American participants. We imposed the following settings in recruiting our participants:
>
> 1. Participants were paid $1.25 as a fixed participation reward. This amount was determined by multiplying the expected completion time (in mins.) with the minimal federal wage in the U.S. (i.e., $0.121 per minute).
> 2. The expected completion time was set at 10 minutes in advance.
> 3. The most time we allowed each worker to complete the study was 30 minutes.
> 4. We limited all workers' HIT Approval Rate to be between 95% and 100%.
> 5. We limited each worker's number of HITs approved to be between 5,000 and 100,000.
> 6. We blocked Suspicious Geocode Locations and Universal Exclude List Workers.
> 7. We blocked duplicate IP addresses and duplicate geolocation.
> 8. We enabled HyperBatch so that all eligible workers were able to participate in our HIT immediately after the survey was launched.
> 9. We restricted workers' location to be in the U.S."

> **The switch from real in-person surveys to online surveys comes with a chance of high inattention or even 'bots' producing random noise. In a good faith replication (especially one where original in-person collection is changed to online sample pools with notorious attention problems), the steps taken to make sure that participants are real and are actually reading the questions are extremely important. The present data collection plan would not make me feel confident that any potential null effect is not actually just due to poor quality, inattentive participants.**

First, we realized an oversight. To be absolutely clear, the sample will come only from a data collection of US Americans on Amazon MTurk (using CloudResearch). We had some mention of Prolific Academic in one of the tables in the submitted Supplementary document, which we now fixed.

Second, using the checks mentioned in the above comment and CloudResearch-approved participants has been shown to be robust, with successful passing of attention and comprehension checks of over 95% consistently (Eyal et al., 2021) and has also been successfully implemented in our past replication efforts. Thus, we do not believe that inattentive participants or bots are going to impact our results in any tangible way.

We have a lot of evidence to show the reliability of our target sample. We receive this comment quite often from reviewers that we are in the process of writing a manuscript aimed to address this specific issue and help others use the platform and achieve high-quality data collections. In our manuscript, we cited and referred to many of our other completed replication projects using this very approach. We will try and summarize our experience in short below.

We completed over 80 replications of classic findings in judgment and decision making using MTurk online samples (see https://mgto.org/pre-registered-replications/), and our experience has been that these samples are very reliable, at least for replications in judgment and decision making.

There is much that we can share on that but briefly:

1. Our successful replication rate is currently at 68% (+12% mixed/inconclusive), higher than most other replication rates in other domains. Even in the ones that are mixed/inconclusive or seemed to have failed we identified reasons that are not related to the samples.
2. When conducting 8 replications in two different online samples, Americans on MTurk and British on Prolific, we found the results highly consistent across the two samples.
    1. See summary tweet: https://twitter.com/giladfeldman/status/1215175786543534090?s=20
    2. Browse the reports: http://mgto.org/hkureplications2019
3. In a number of replications, when we conducted replications on both students samples and online on Mturk, we found the findings consistent across the two samples.
    1. Example 1: https://www.researchgate.net/publication/331431431_Agency_and_self-other_asymmetries_in_perceived_bias_and_shortcomings_Replications_of_the_Bias_Blind_Spot_and_extensions_linking_to_free_will_beliefs
    2. Example 2: https://journals.sagepub.com/eprint/MVTW3KE2MXN2SRRKDGYE/full
4. When we ran the exact same replications on MTurk in two time periods, with a time gap of several months to two years, ensuring different participants from the same online platform, we found highly consistent results.
    1. Example 1: https://www.researchgate.net/publication/326548295_The_impact_of_past_behavior_normality_on_regret_Replication_and_extension_of_three_experiments_of_the_exceptionality_effect
    2. Example 2: https://www.researchgate.net/publication/339167597_Revisiting_status_quo_bias_Replication_of_Samuelson_and_Zeckhauser_1988

5. We have already collected and analyzed the data for most of the IPA-ed replication and extension PCI-RR mentioned above with similar unified designs, and the majority of those turned out to be successful replications. This is further reassurance that our measures for high quality data collections and our method of unifying several similar studies works well. Examples (theses with data collection completed, to be submitted as PCI-RR Stage 2 soon):
    1. Thesis: Revisiting the link between true-self and morality: Replication and extensions of Newman, Bloom, and Knobe (2014) Studies 1 and 2.
       [IPA] [OSF]
    2. Thesis: Revisiting diversification bias and partition dependence: Replication and extensions of Fox, Ratner, and Lieb (2005) Studies 1, 2, and 5.
       [IPA] [OSF]
    3. Thesis: Revisiting stigma attributions and reactions to stigma: Replication and extensions of Weiner et al. (1988).
       [IPA] [OSF]
    4. Thesis: Revisiting mental accounting classic paradigms: Replication of the experiments reviewed in Thaler (1999).
       [IPA] [OSF]

References:

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. Behavior Research Methods, 1-20.

Bots do not seem to be an issue with CloudResearch-approved participants and with all our checks in place. Some related readings:

- Moss, A., & Litman, L. (2018). After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it. Retrieved February, 4, 2019.
- Hauser, D., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2021, August 20). Evaluating CloudResearch's Approved Group as a Solution for Problematic Data Quality on MTurk. https://doi.org/10.31234/osf.io/48yxj
- Litman, L., Rosenzweig, C., & Moss, A. (2020). New Solutions Dramatically Improve Research Data Quality on MTurk.

The current data collection plan also does not outline whether excluded participants will be replaced and whether the power calculation refers to the final sample or simply to the number of slots posted on Mturk.

All analyses and power calculations are based on the post exclusion sample of $N = 515$. We clarified this further in the "Power analysis" subsection to make that point more explicit.

> **Further, is there a planned point of percentage of discarded data at which the study would be deemed failed? Would you consider data even if, say, 30% of respondents have to be excluded?**

We do not have a plan for such a contingency and based on our experience, the exclusions are generally minor and have rarely had impact on the findings. We will include details about how many participants were excluded and also an analysis of pre- and post-exclusion results, as detailed in our supplementary ("Comparisons and deviations" section).

> **The planned design of running all three studies on the same sample of participants is not true to the original study designs. Since participants are being paid according to the time it takes to do the survey, it would require the same resources to run these three studies separately, so the reason for this divergence from the original study is unclear. Note that no reason is given for this considerable change from the original procedure - not even in the table outlining original vs replication methodological comparison that includes a 'reason for change' column (Appendix B).**

We laid our logic for this change in our Procedure section in the main manuscript. To address the comment, we now expanded that section (marked in bold):

> "We combined the three studies in a single online survey. This allowed us to maximize our resources and had the added advantage that we can rule out any sample characteristics that might be driving differences in successful versus unsuccessful replications. **Additionally, a single survey allowed us to conduct additional exploratory within-subjects analyses and explore links between different studies, something that is not possible with the original's design.**"

This was also elaborated in the "Original vs replication methodological comparison" table.

> **For example, if participants thought about and responded in line with a sunk time effect in an earlier question, they might then respond consistent with earlier responses in Study 5 and would be less likely to be swayed by the education condition. Giving Study 5 always after Study 1 or 2 stacks the deck against replicating the Study 5 effect because of these consistency biases in responding.**

The combined design was meant to be able to address specifically these kinds of questions. It is difficult to tell whether this change would stack the deck against the replication, increase its chances, or have no impact. Many judgment and decision-making paradigms are much stronger in within-subject designs than in between-subject design (e.g., omission bias, action-effect), though there are paradigms that only work in between-subject designs (e.g., less is better).

We are interested in whether education is capable of doing what it was hypothesized to do: change reactions. This unified design allows us to test those against baselines, and get a more accurate understanding of what is taking place.

We will examine whether there are differences in responding in Study 5, compared to Study 1, across the education and opportunity cost conditions - see subsection "Study 1 versus Study 5: Analysis of within-subject effects" in the main manuscript.

We also accept the point that the changes made to Study 5 in both the way we conducted the manipulation and in running after a similar Study 1 means that our replication criteria should be adjusted from "very close" to "between close and far replication". This study is somewhere between a direct and a conceptual replication. We adjusted our replication classification in the supplementary to reflect that change.

Above, we mentioned several examples of our other replications with a similar unified design that were successful with the unified design providing valuable insights.

> **Even if all studies were administered in the proposed way (Study 1 and 2 counterbalanced, then Study 5), authors should check for order effects of the counterbalancing in all analyses. In the current proposed analyses, potential order effects are not tested.**

Good suggestion, we agree.

If we fail to find support for the original, then we will add additional order effect analyses. We added that to the "Additional analyses and robustness checks" subsection.

> **I disagree with the authors' characterization of the IV materials in S3 as being the "same" as materials in the original study (according to Appendix B, Table on replication classification). In the original study, students were listening in person to a university lecture on opportunity cost. This information is coming from a source they take seriously (a professor at the university they attend) and is of considerable length and depth. This is not in any way the 'same' as reading 266 words about opportunity cost in an online experiment. Even if they show comprehension of this information on 2 followup questions, the information is not likely to be processed anywhere in near the same depth.**
>
> **As a followup on point 3, authors do not even plan to exclude people who did not answer one or both of the comprehension questions correctly (this is neither mentioned explicitly nor reflected in the simulated data). So, someone who skims the brief paragraph on opportunity cost and answers at random would still be included as bona-fide participant. In my opinion,**

> **this proposed experiment could not conclude whether opportunity cost education has a moderating effect on the sunk time cost effect because there is no evidence that participants read the or processed the information. Of course, the original experiment cannot be sure participants processed the information either, but in that experiment, participants were present for an hour+ long lecture – the likelihood that something 'sticks' is much higher than for 266 words in an online survey. I'm not saying that a meaningful replication of Study 5 cannot be done online but the proposed way of doing it is in no way equivalent to the original study.**

Good point, thank you for that feedback. We agree that our classification of Study 5 should be adjusted.

We updated our table in the supplementary materials subsection "Replication classification" table to "Different" for IV and DV stimuli, and the procedure to "different" given that it is always last. Thus we concluded our replication classification as "Close/far replication" instead of "Very close replication", as per LeBel et al. (2018) criteria.

> **A minor inconsistency. In the method description, the central outcome scale is labelled 1-9 (in line with the original scale). However, in the appendix (as well as in the OSF survey) the scale anchors are changed and the scale is now labelled 4 – 0 – 4 (not even -4 to +4). I would recommend the authors stay true to the original materials (as they said they did) in all aspects of the materials.**

Thank you very much for catching this and providing us with this feedback. Much appreciated!

First, we realized there were several oversights regarding our description of the measures where we could have done better and laid out things more clearly. We now adjusted the reference to the scale, explained our deviation (more on that below), and expanded on the measures in all of the studies. We also clearly label which measures are a direct replication and which measures were added as extensions.

Second, we appreciate the feedback that we should better document and explain our deviation from the original's scale of 1 to 9 to our adjusted scale of 4/0/4. We made this adjustment to try and avoid the possibility of biasing participants towards the larger number option. We thought that the 4/0/4 is more intuitive for participants to grasp than the 1 to 9. We felt that this was a needed adjustment, yet we would gladly readjust this to the original's scale given clear editorial guidelines to do so.

This has been added to the deviations table, and is now also noted explicitly in the text. We also better explain the differences between how this is presented to participants (4/0/4) and how this is coded (1 to 9).

Response to Reviewer #3: Dr./Prof. Christopher Olivola

> **I am happy to see replications of classic effects, and the sunk-cost effect (for time) is no exception. Therefore, I commend the authors for carrying this out. That said, I do have some comments and concerns about the current plan and/or manuscript:**

Thank you for your encouraging note and the helpful feedback.

> **First off, the current manuscript is poorly written. In particular, I see a lot of grammar errors that could (and should) have been checked and corrected (e.g., using grammar checks in Word). I sincerely hope the authors will make an effort to proof-read their word before they submit it for review.**

In our revision, we went over the manuscript again and tried to address grammar/spelling better.

> **The authors should cite and discuss other papers (besides Soman, 2001) that have previously tested (and found) sunk-cost effects for time (e.g., Bornstein & Chapman, 1995; Bornstein, Emler, & Chapman, 1999; Frisch, 1993; Navarro & Fantino, 2009; Olivola, 2018; Strough et al., 2008). This is important, since these other papers also speak to whether (and to what extent) there is a sunk-cost for time. In fact, I would suggest the authors provide a table that summarizes these other papers and, for each one, what they found (e.g., whether they found a sunk-cost effect of time).**

Thank you for suggesting relevant literature. We used some of these as the basis for expanding our literature review in the introduction. We do note that the intended scope for this replication is rather narrow, and we did not want to shift focus away from that with a review of the vast literature. Please also see our reply to the other reviewers on this point.

> **Having the same participants complete all 3 studies in a single session is problematic, as it may cause spillover effects, amplify demand effects, etc. The authors should consider randomly assigning participants to one of the 3 studies (not all 3). Or, at the very least, the main analyses should only focus on the first study that each participant is assigned to (and subsequent analyses can look at all 3 studies within-participant).**

Thank you, we appreciate this feedback and suggestion. Similar comments were made by the other reviewers, please see our detailed reply above.

In addition, to address your suggestion we added the following to our summary of the results:

> **Between subject studies and order effects (exploratory)**
>
> If we fail to find support for the target's findings, we will conduct additional exploratory analyses examining order effects and controlling for order.
>
> If we fail to find support for the target's findings, we will conduct additional exploratory analyses examining Studies 1 and 2 only when they were the first study presented to participants. This would address possible confounds between the studies, resembling running two separate studies.

> **Another concern, which may lead to a failure to replicate the effect, is that experienced MTurk participants may have been exposed (and some repeatedly) to sunk-cost studies, and this may hinder the effect. The authors should therefore consider limiting the study to MTurk participants who have had relatively little experience (e.g., fewer than 100 MTurk studies completed).**

We have an item in the Funnelling section (last in the survey) which asks participants "Have you ever seen the materials used in this study or similar before?", which we previously included in the supplementary materials in the "Materials used" subsection, and indicated those who responded "Yes" as part of the exclusions criteria that was in the supplementary.

We realized that this was difficult to follow and appreciate the feedback to make this clearer. We moved the exclusions criteria to the main manuscript, and provided additional information about our funneling section in the procedure subsection in Method.

> **On p. 20, the authors write: "In order to address H1, Soman (2001) conducted multiple chi-square tests. Specifically, in Study 2, he showed that in the money condition, the chi-square test found difference between sunk cost and no sunk cost conditions, whereas the same difference was not found for the time condition. A different way to approach H1 is to ask whether the likelihood of picking the option associated with sunk costs (theater performance in Study 1 and rocket engine in Study 2) is different across conditions. To address this question, we conducted a logistic regression analysis for Studies 1 and 2 for both the original and the replication data."**
> **=> I don't understand the distinction that the authors are trying to draw, here. Chi-Square tests also evaluate whether likelihoods vary across conditions, so the authors are mistaken if they suggest otherwise. I suspect they meant something else, but that it did not come across clearly in their writing.**

We agree that both chi-squares and logistic regression (LR) assess likelihoods, but chi-squares cannot test interaction effects. We are using LR, as per recommendations from R5 (Johannes Leder), which allows us to test an interaction effect in Study 2.

With that in mind, we agree with the reviewer that our writing was not clear in the original manuscript. Our goal was to communicate the importance of testing the interaction, which was the goal of our LR analysis. We have now revised the sentence in an attempt to make our point clearer:

> "A different way to approach H1 is to ask whether the likelihood of picking the option associated with sunk costs (rocket engine in Study 2) is different not only between levels of a single independent variable (sunk cost presence or sunk type) but also whether there was an interaction between the two variables."

To that effect, we also removed our LR analysis on Study 1. We agree, the chi-square analysis is sufficient.

Response to Reviewer #4: Dr./Prof. David Ronayne

> **The authors aim to replicate 3 studies from Soman (2001).**
> **The sunk cost effect is important. Distinctions between different types of sunk costs are important. Replicating studies related to that, e.g., those of Soman (2001), are therefore worthwhile contributions.**

Thank you for the positive opening note.

> **This is a replication paper; rationale and plausibility are clear. Below I suggest some wording changes from those presented in Table 1.**
> **Hypothesis 1: "More generally" does not make sense - domain is not a generalization of size... Perhaps splitting it into separate hypotheses would make more sense.**
> **Hypothesis 2b: I would not write "Rational" - you expect to be dealing with subjects who exhibit the sunk cost effect, at least when money is sunk (making their choices inconsistent with some textbook "rational" actor).**

Thank you. We stripped Hypothesis 1 down to the essential information in Table 1:

> "The sunk-cost effect is weaker in the domain of temporal costs than in the domain of monetary costs."

We also removed references to "Difficulty" and "Rational" next to Hypotheses 2a and 2b.

> **Order effects. Did Soman randomize the order of studies 1 and 2, as you will? If yes, did they find an effect? Please test for an order effect.**

Yes, we received similar advice from the other reviewers, and therefore added planned order related analyses in case we fail to find support for the original's findings.

> **Please report the average completion time and the lump sum offered for completion, not only the goal of 7.25/hour. Also, 7.25 is the federal minimum, but it differs by state (https://en.wikipedia.org/wiki/List_of_US_states_by_minimum_wage). I imagine it will affect your sample demographics, e.g., education level or employment experience, which could correlate with your outcome measures of interest. If you can, I would pay more. If you cannot, it would at least seem worth discussing.**

Thank you, appreciated. Yes, we do not restrict access based on state, but based on country. We do understand that this means that those from higher pay minimum wage states will not take part, or that participants from higher pay minimum wage states somehow decide to take part regardless.

We added the following to the planned limitations and future directions subsection in the General Discussion:

> [Based on feedback from peer review: potentially discuss how sample demographics, such as education level or employment experience, and our pay based on a federal minimum wage rather than by state, may correlate with outcome measures.]

In addition we now clarify in the methods section (underlined added):

> The assignment pay was calculated based on the federal wage of 7.25USD/hour (<u>though we did not restrict participation based on state-level minimum wage</u>).

> **Was Soman's sample undergraduate students? MTurkers are a different crowd and their different demographics may be a driver of the results you find. You cannot do a detailed comparison, as you report Soman did not disclose detailed demographic information, but there are systematic differences between college students and Mturkers e.g., age, experience, incentives, etc. I encourage a discussion of this potential source of differences, and analysis of how your subjects' demographics are associated to the treatments (see next points).**

We added the following the "Limitations of our replication and directions for future research" subsection of the General Discussion:

> "Our replication had limitations, and we needed to make several adjustments to the target's design to accommodate our sample and method of delivery. First, participants in the original study were students who were enrolled in a particular class, whereas participants in our replication were sampled from the general population. This makes it possible that the student sample was systematically different in some respect, compared to the general population. "

> **Logistic regression. You already conducted Chi-squared tests for Studies 1 and 2 and you have predicted proportions from the raw data. Why run logistic regressions with only the treatment (dummy?) variables on the RHS? The value I see in regression analysis would be to see if there were some interesting covariates of sunk cost behavior not picked up in Soman, that may explain your data e.g., subject demographics. Please revise or justify why the regressions you propose add value.**

Based on your and Dr./Prof. Christopher Olivola's (R3) feedback we improved on our reporting of the logistic regression (LR) analysis. We are using LR, as per recommendations from Dr./Prof. Johannes Leder (R5), to allow for the testing of an interaction effect in Study 2.

We revised our manuscript to reflect this:

> "A different way to approach H1 is to ask whether the likelihood of picking the option associated with sunk costs (rocket engine in Study 2) is different not only between levels of a single independent variable (sunk cost presence or sunk type) but also whether there was an interaction between the two variables."

We also removed our LR analysis on Study 1, as the chi-square analysis seemed sufficient.

> **Can you explicitly confirm whether the within-subject (design and) analysis was also done by Soman? (I guess it was not - but if yes, a comparison is needed)**

Soman (2001) did not run the studies together and therefore there was no within-subject analysis. We placed this analysis under a section "Additional analyses and robustness checks" in order to make it explicit that this is different from the replication.

> **Overall, except potentially the last point above, this is a pure replication paper. That is of course a great goal in itself. But do you want to explain your and/or Soman's results? If nothing else than by discussing how they may vary with demographics or other variables? I did not see any analysis aimed at this, yet it would seem straightforward to add (e.g., via multivariate regression - see point above) and potentially enlightening.**

We aimed this as a replication paper, and we do go beyond the exact replication by adding additional analyses and extensions. Our scope is well-defined and complex enough, and so we do not plan to make any additional exploratory analyses to examine demographics or other variables, this would complicate the investigation and shift the focus from the main aim needlessly.

We are making our materials, data, and code publicly available, so anyone interested in exploring further, is very welcome to do so.

Please list the details of the Qualtrics implementation somewhere, e.g., availability of a "back" button, time limits, forced responses, etc. The idea being that someone could fully replicate your work with all the same options selected in the software.

We fully support others being able to fully replicate our work. To that effect we already shared an exported .docx, .pdf, and .qsf of our Qualtrics survey. These documents contain all relevant details that one might need to replicate our work. The .qsf is easiest to use but can only be used within Qualtrics (anyone can sign up for a free account, that allows navigating the QSF), while .docx and .pdf can be used by anyone. These are contained in the online OSF repository under the "Qualtrics survey" folder.

Please make your data and analysis code available ex post.

All materials, data, and code will be made available, this is standard practice in all our projects (https://osf.io/5z4a8/).

The "Open Science Declaration" section in the studies overview reads:

> "This replication is submitted as a Registered Report (Chambers & Tzavella, 2022; Nosek & Lakens, 2014; Scheel et al., 2021; Wiseman et al., 2019).
>
> We will pre-register the experiment on the Open Science Framework (OSF) and data collection will be launched shortly after pre-registration. Pre-registrations and all materials used in these experiments are available in the supplementary materials. We provided all materials, data, code, and pre-registration on: https://osf.io/pm264/.
>
> All measures, manipulations, exclusions conducted for this investigation will be reported, all studies will be pre-registered with power analyses, and data collection will be completed before analyses. We reported results after exclusions below, and in the supplementary materials, we detailed a comparison between pre- and post-exclusion findings as well as any deviations from the pre-registered plan ("Comparisons and deviations" subsection), with additional disclosures ("Open science disclosures" subsection)."

> **The exclusion criteria seem good, but I think some other good ones are not listed. First, a criterion based on how quickly the main pages were submitted. Too fast means it was infeasible that they read the text. Second, on a page following the main text (and assuming there is no ability to go back) it would be good to ask questions to check subjects paid attention (questions they would only know the answer to if they read the text). (You ask some questions in Study 5 which require subjects to understand some conceptual info, but those are part of the education treatment.)**

We appreciate these suggestions, which helped us realize that we should do better in explaining the measures we took to ensure high-quality responding.

Before the participants embark on the studies they must indicate their consent, qualifications, and agreement, and we deliberately randomize the choices in these questions which requires participants attention. Those who fail to indicate "Yes" to these questions are asked to return the task. This helps ensure attentiveness.

We now explain this in greater detail in the "Procedure" section:

> Participants first provided consent, after which they read an outline for the studies and three questions confirmed participants qualifications as being American, their understanding of the study procedures, and their agreement to pay close attention (Yes/No/Not sure presented in random order, and participants not answering Yes were asked to return the task).

We added additional information about our recruitment criteria in how we ensure high-quality data collection, please see our reply to Dr./Prof. Johanna Peetz above.

The subsection "Additional information about the study" in the supplementary also details all the criteria used in the recruitment (which will be updated after data collection if there were any changes), please see our reply to Dr./Prof. Johanna Peetz above.

> **Related, you expect only 5% of their sample to be excluded by their criteria (p11). My sense is that that may be conservative. I believe the present study would be significantly boosted by stronger exclusion criteria and a correspondingly larger initial sample.**

Similar comments were made by other reviewers, thank you for the feedback. We now make clearer our measures for ensuring high-quality data collection, our exclusion criteria, and we updated our estimates for the exclusions to 15%, please see our reply to Dr./Prof. Johanna Peetz above.

> **I recommend the authors check the following references, especially the last one which looks to some extent at the distinction between time and money**

in a sunk cost context.
Augenblick, 2016, "The sunk-cost fallacy in penny auctions"
Olivola, 2018, "The interpersonal sunk-cost effect"
Ronayne, Sgroi, and Tuckwell, 2021, "Evaluating the Sunk Cost Effect"

Thank you for suggesting relevant additional literature. We used it to expand our literature review in the introduction.

I believe it would be better to qualify the introductory definition of the sunk cost effect as relating to *irreversible* or *unrecoverable* investments of resources, e.g., line 1 of Abstract.

We now added "unrecoverable" in our definition of sunk costs in the Abstract:

> "The sunk cost effect is the tendency for an individual's decision-making to be biased based on unrecoverable previous investments of resources."

p7 "found that the sunk cost effect was ... not [present] for time": you cannot prove a null. Please re-phrase e.g., "no evidence for an effect of..."

Thank you for this feedback, much appreciated, we agree completely.

We rewrote the sentence to the following:

> "Soman's (2001) core finding was that the strength of the sunk cost effect was weaker for time than for money. He further showed that the facilitation of money-like accounting for sunk time costs by highlighting opportunity costs or by educating about an economic approach to time strengthens the sunk time cost effect."

Two examples of unclear writing below from early in the manuscript. I shall refrain from further comments about the writing, but recommend you have the final manuscript proofread.
a. I do not understand the phrase at the end of the first sentence of the first paragraph of the Introduction"given that with larger sunk costs are stronger tendencies to further escalate". I would avoid the speculation over the ("vicious cycle of") consequences of the SCE, and just talk about the SCE itself, except for the discussion.
b. p1 "yet evidence is sometimes inconsistent with weak effects" does not read well. There are at least two different possible meanings.
c. p7 "appeared" rather than "re-appeared"

Thank you for the helpful suggestions.

We revised the manuscript as follows:

a. We removed "vicious cycle" from the first sentence:
   "...leading to an escalating commitment to a losing course of action"
b. We removed "with weak effects":
   "...yet evidence is sometimes inconsistent…"
c. Changed "re-appeared" to "appeared"

I am not sure that 420 citations in 21 years is a huge amount. Also, some people think Google Scholar is a poor citations counter. I think it distracting and unnecessary and would remove it.

This is a typical section that we included in all our replication submissions, including in-principle accepted preprints by PCIRR, and is a result of requests by some reviewers to demonstrate impact, commonly measured by number of citations. Based on our experience in replicating classics, 420 citations in 21 years is definitely unusual and far above average in the judgment and decision-making domain. Google Scholar, with all its weaknesses, is considered one of the most comprehensive databases, and - unlike other resources - is completely free.

If needed, we provide some citations which demonstrate that 420 citations for 21 years ranks at top 1% of journal articles across top Psychology/Cognitive journals:

- Cho, K. W., Tse, C.-S., & Neely, J. H. (2012). Citation rates for experimental psychology articles published between 1950 and 2004: Top-cited articles in behavioral cognitive psychology. Memory & Cognition, 40(7), 1132–1161. https://doi.org/10.3758/s13421-012-0214-4
- Kurilla, B. (2015). How Many Citations Does a Typical Research Paper in Psychology Receive? – Geek Psychologist. https://geekpsychologist.com/how-many-citations-does-a-typical-research-paper-in-psychology-receive/

Typos I spotted:
Authorship declaration: "is" in line 1.
p10. "have possible detected" should be "possibly".
p15. "we found was", remove "was".11
p22. You write "no support for a main effect of sunk type" and then an effect with significance p=.001…
p22. "Soman found a main effect of sunk presence

Much appreciated. All the listed typos have been fixed.

Response to Reviewer #5: Dr./Prof. Johannes Leder

> **The study seeks to replicate the hypothetical scenarios used in the experiments 1,2 and 5. I am not sure if these are the experiments that the resources should be focused on. The study seeks to replicate a sunk cost effect for time and money DECISIONS – the replication proposed now only seeks to replicate the effects for INTENTIONS. Here is a severe mismatch. Soman (2001) used experiment 6 to validate his previous findings, for this reason, experiment 6 seems to be the most crucial experiment for his argument and not study 1,2, and 5. As he states: "Experiment 6 involved real choices made by individuals who had made real investments of time. The results validated Hypotheses 1 and 2a, namely that the sunk-cost effect was not detected in the domain of temporal investments, but it reappeared when the accounting of time was facilitated."**

Thank you for the feedback. We appreciate you sharing your view, and we agree that there is value in examining real choices made by individuals. We also see value in examining intentions. Both intentions and decisions are important, and then it is a matter of priorities. Soman has shown that the phenomenon seems to extend to both intentions and decisions, and so we felt it important to first address intent before we embark on the more costly and complex real decisions. In our view, this is a necessary first step, and we would like to first establish the phenomenon demonstrated in the initial studies.

Also, effects are typically much larger and easier to detect for intent than they are for decisions, and so this increases the likelihood that we will be able to detect effects, especially if the study design involves interactions. If we are successful in replicating Studies 1, 2, and 5, we hope that this will pave the path for pursuing a replication of Study 6 and real-life behavior in future studies.

> **I appreciate the detail and care the authors have taken in simulating the data and showing the results in an adequate statistical framework (logistic regression). The analysis of the preference ratings should be done with a cumulative logit or probit regression and not ANOVA as the measure is not truly continuous -see:**
> **Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? Journal of Experimental Social Psychology, 79, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009**

Thank you for the valuable feedback and suggestions on an alternative analysis. We also see much value in running additional analyses using logistic regression for a more robust

test of Soman's predictions, which is why we incorporated this type of analysis in the manuscript, in addition to replicating the original's analyses.

In running replications it is important that we atleast try and run the same analyses as the target's, and compare the effects using the same analyses as the ones conducted back in the day. We prefer to err on the side of doing and reporting too much than doing too little.

We added a planned brief discussion of this point in our general discussion in the "Limitations of the original study: Directions for improvement" subsection.

Reply to decision letter reviews: #187 revise and resubmit 2

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold while our answers are underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/pYhszjIqSBTo

A track-changes manuscript is provided with the file: "PCIRR-RNR2-Soman 2001-Replication-Manuscript-v4-G-trackchanges.docx"

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

Section	Actions taken in the current manuscript
General	N/A
Introduction	R2: expanded the introduction (3rd paragraph)
Methods	R1: added a table summarising methodology differences (Table 2) R2: doubled the sample (1212 recruitment with a target of 1030 final N) in order to be able to run order effects analyses. Updated power analysis section. R2, R3: Implemented additional familiarity questions and comprehension check questions, which we use as exclusion criteria
Results	R2: committed to order effects analysis and included a table (Table 7) summarising our findings
Discussion	N/A
Reporting	N/A
Supplementary materials	R2, R3: added the familiarity questions and comprehension check questions

Note. Ed = Editor, R1/R2/R3 = Reviewer 1/2/3

Response to Editor: Prof. Chris Chambers

> **Three of the five reviewers kindly returned to evaluate your revised manuscript. As you will see, most of the major issues have been addressed and we are moving closer to Stage 1 IPA. However, there are some remaining points to address concerning the literature review, exclusion criteria, and planned statistical analyses, and the reviewers offer a range of constructive suggestions for resolving these issues. I look forward to seeing your response in a further (hopefully final) Stage 1 revision.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

Response to Reviewer #1: Prof. Dilip Soman

> **My review of the revised manuscript is short because the authors have addressed the issues I had raised in the last round.**
>
> **The only other suggestion I have is merely a suggestion (but one that I've found to greatly increase the readability of replication papers) is to include an additional table that compares the procedures and analysis plans across the studies (basically, the equivalent of Table 2, but for procedures and analysis). Rows might include: a) Design, b) Method of recruitment, c) Whether study was done independently or as part of a package, d) Mode of data collection (online, paper-pencil), e) Implementation of Manipulations etc. All of this information is currently in the manuscript and it might be good to pull it all together into a table. Also, this might be a good template for others in the field to follow!**
>
> **All the best, and I look forward to seeing the project move further along.**

Thank you for the overall positive feedback and constructive comments in both iterations of the feedback. We added a new table, namely Table 2 (note that table numbers have been changed now), that covers the key dimensions across which the studies differ to allow the reader to get the methodological context at a glance.

Response to Reviewer #2: Dr./Prof. Christopher Olivola

> **The authors have only partially addressed some of my original comments, and failed to adequately address my core comments/concerns. Below, I go through my original points, and comment on the authors' responses to each one**
>
> **Comment 1) "[…] manuscript is poorly written […] I sincerely hope the authors will make an effort to proof-read their word before they submit it for review."**
> **=> The authors say that they have addressed grammar/spelling errors. I certainly hope so, but I didn't go through the trouble of proof-reading the manuscript, so I'll simply have to trust that the revised version is largely free of errors.**
>
> **Comment 2) "The authors should cite and discuss other papers (besides Soman, 2001) that have previously tested (and found) sunk-cost effects for time (e.g., Bornstein & Chapman, 1995; Bornstein, Emler, & Chapman, 1999; Frisch, 1993; Navarro & Fantino, 2009; Olivola, 2018; Strough et al., 2008). This is important, since these other papers also speak to whether (and to what extent) there is a sunk-cost for time. In fact, I would suggest the authors provide a table that summarizes these other papers and, for each one, what they found (e.g., whether they found a sunk-cost effect of time)."**
> **=> The authors have largely ignored and seemingly chosen to "dodge" this comment.**

We did try and address the suggestions in our previous revision, and we understand that we could have done a better job explaining what we did. We are glad to include citations that we feel are relevant and make the manuscript stronger.

We previously went through the citations you provided in detail and considered each of them. Three of the citations have already been included, and two of them did not have any scenarios involving sunk time costs and so we did not think those were relevant to include.

We added the one suggested citation that we seemed to have missed, the Bornstein and Chapman (1995), now included in paragraph 3 of the introduction.

> **…This is extremely puzzling to me, since this comment is (i) easy to address, and (ii) important for reasons I previously tried to explain, but will now elaborate on further: The authors argue that "the intended scope for this replication is rather narrow", but this argument fails on several grounds. First, replicating an effect does not absolve the authors of their responsibility to properly acknowledge and discuss the relevant literature. In this case, the relevant literature clearly concerns a (possible) sunk-cost effect for time (vs. money). Therefore, the authors should cite (and summarize the findings of) any/all prior papers that have tested such an effect, even if it was not the only/main goal of a given paper. Such a discussion is important, as it gives readers a proper sense of what the existing evidence is concerning time sunk-cost effects (or lack thereof), as well as possible boundary conditions, etc.**
>
> **Second, I'm not sure how "narrow" the authors want the scope to be, but I am sure that the goal of a replication (and science, more generally) is not merely to replicate a particular study, but rather to examine the (in)existence of an effect (or set of effects) more generally. Therefore, it is not the Soman (2001) studies and findings per se that are interesting and important, but rather the establishment of a sunk-cost effect for time (or failure to do so). Consequently, the authors cannot argue that discussing other prior studies that tested and found or failed to find time sunk-cost effects is beyond the scope of their manuscript.**

We appreciate this suggestion, and we see the importance of a summary of the literature. We followed what we saw as common in similar other replication manuscripts we and others conducted on PCI-RR and social psychology journals (which we provided examples for), which focuses the scope of a replication on the target article for replication. We understand that we may differ in our opinion on what our role and responsibility should be in a replication project. We see our main responsibility to be about conducting a rigorous empirical close replication of one target article.

We think it important to include several of the key citations in the follow-up literature, and we included several of those in our introduction, yet the scope of a replication as we defined it for our investigation is and has been encouraged by PCIRR is far more narrow than a comprehensive review of the literature covering any/all papers published since and analyzing their effect sizes. Furthermore, the task of summarizing a literature is anything but easy, and should have a dedicated systematic review/meta-analytic effort to address it comprehensively and make sure it lives up to its importance. Doing that in a replication manuscript runs the risk of distracting readers from what the manuscript was meant for - an empirical replication of one classic article in that literature.

We think it is important to align expectations here. We understand that you have a much wider aim of verifying and summarizing the entire literature beyond Soman (2001), yet we cannot address that aim in this effort. Our aim is much humbler and very specific in regards to focusing on the Soman (2001) article. Regardless of our findings in our replication of Soman (2001), we make no claims regarding the broader literature. We feel the focus on this article is interesting and important enough.

We also addressed the issue about the importance of replications of a single target in our previous reply:

> At this point in time we believe there is already a general consensus regarding the need for replications, and "making replications mainstream", without the specific need to justify a replication of a well-cited impactful paper that has not been subjected to independent pre-registered/Registered Report direct replications.

To address this point we now also include an explicit section about this need, with some citations of the work on that need:

> We aimed to revisit the classic phenomenon and examine the reproducibility and replicability of the classic findings by replicating the studies and improving the design with extensions. Following the recent growing recognition of reproducibility and replicability in psychological science (Brandt et al., 2014; Open Science Collaboration, 2015; Nosek et al., 2022; Zwaan et al., 2018), we embarked on a well-powered pre-registered replication and extensions of Soman (2001).

Our introduction in this manuscript, our focus on a single target rather than an entire literature, and our justifications for the importance of replications is similar to other PCIRR manuscripts from our team that received in-principle acceptance. Examples:

1. Yiu, S., & Feldman, G. Revisiting the psychological sources of ambiguity avoidance: Replication and extensions of Curley, Yates, and Abrams (1986).
   [In-principle Acceptance] [Preprint] [Open materials/data/code]

2. Lee, S., & Feldman, G. Revisiting the link between true-self and morality: Replication and extensions of Newman, Bloom and Knobe (2014) Studies 1 and 2.
   [In-principle Acceptance] [Preprint] [Open materials/data/code]

3. Lu, S. & Feldman, G. Associations of fear, anger, happiness, and hope with risk judgments: Revisiting appraisal-tendency framework with a replication and extensions of Lerner and Keltner (2001).
   [In-principle Acceptance] [Preprint] [Open materials/data/code]

4. Li, M., & Feldman, G. Revisiting diversification bias and partition dependence: Replication and extensions of Fox, Ratner, and Lieb (2005) Studies 1, 2, and 5. [In-principle Acceptance] [Preprint] [Open materials/data/code]

5. Yeung, K., & Feldman, G. Revisiting stigma attributions and reactions to stigma: Replication and extensions of Weiner et al. (1988). [In-principle Acceptance] [Preprint] [Open materials/data/code]

6. Zhu, M. & Feldman. G. Revisiting the links between numeracy and decision making: Replication of Peters et al. (2006) with an extension examining confidence. [In-principle Acceptance] [Preprint] [Open materials/data/code]

7. Li, M. & Feldman, G. Revisiting mental accounting classic paradigms: Replication of the experiments reviewed in Thaler (1999). [In-principle Acceptance] [Preprint] [Open materials/data/code]

> **Third, were the authors serious (but mistaken, in my opinion) in trying to argue that their only goal should be a very narrow/specific attempt at a replication of Soman's studies, then their proposal would have to be considered a failure, since they are using a different design, different subject population, etc. In other words, the authors can't have it both ways: they cannot claim to be narrowly focused on specifically replicating Soman's originally studies (as a supposed license to ignore other prior papers that tested time sunk-cost effects), while at the same time carrying out a replication that alters (i.e., departs from) many features of Soman's original studies, and thus would generalize the testing of those findings beyond the paper.**

We see these points as separate and we see no contradiction in our aims. What we are attempting is a very common replication effort. Replications will deviate from the original, in time, sample, context, and sometimes in design. That is not a failure or a weakness, that is a strength. We are using best practices and most up to date categorizations regarding replications (e.g., LeBel et al., 2018 and LeBel et al., 2019) to transparently document deviations and adjustments and reflect on those with humility and caution. As you've seen yourself, in this case we are fortunate to have the author of the target article agree with these aims and claims.

Using these criteria, we categorized and justified our categorization as "close replication" and explained that in detail (see "Replication classification" supplementary as well as main text).

> **In sum, the authors can either (i) carry out a very narrow (i.e., exact) replication of Soman's studies, which means they would need to replicate the original design features and collect data from a similar population OR (ii) they can generalize their replication to be about a time sunk-cost effect more**

> **generally, which allows them to depart somewhat from the original study but also requires that they acknowledge (and not ignore) other prior studies (including the ones I cited, above, as well as those provided by Review #4) that tested time sunk-cost effects. In my opinion, the latter (ii) is far more interesting and will make a greater contribution. Regardless of what they decide to do, the authors cannot have it both ways. Just to be clear, I was not suggesting that the manuscript should shift to becoming a thorough and detailed review of time sunk-cost effects. However, I think it is well within the scope of the paper to at least offer one table (and maybe even a graph plotting effect sizes) summarizing the results of prior studies involving time sunk-cost effects, and at least one paragraph summarizing what those prior studies found (overall).**

Our aim is the first of these two, and what we set out to do best we can using best practices for replications. Our replication includes and strengthens the original's methods, and we document all adaptations and will interpret all findings in accordance. All of our adaptations are typical to this literature and for replications, and we conducted those and published after extensive peer-review with similar adjustments and properties dozens of times before. We believe that they offer an important contribution, both revisiting the original findings with a reproduction and replication, and with several additional extensions that go beyond the original.

We hope that this will inspire more interest in this domain and would be happy to see others follow up on the directions you outlined.

> **Comment 3) "Having the same participants complete all 3 studies in a single session is problematic, as it may cause spillover effects, amplify demand effects, etc. The authors should consider randomly assigning participants to one of the 3 studies (not all 3). Or, at the very least, the main analyses should only focus on the first study that each participant is assigned to (and subsequent analyses can look at all 3 studies within-participant)."**
> **=> It looks like--though this is not entirely clear from their response--that the authors agree to carry out analyses on only the first set of studies/conditions that participants see. This should be done regardless of what the within-subject analyses find. In fact, as I previously indicated, the between-subjects analyses (on the first scenario subjects see) should really be the *main* analyses that they carry out (i.e., the within-subject analyses should be the additional/bonus ones; not the other way around).**

We have now committed to re-running all analyses in the paper split by whether Study 1 or Study 2 were presented first - see "Order effects between studies" section as well as Table 7. We have doubled our target sample size from 600 to 1212 (515*2 + 15% planned exclusion rate).

We wish to note that the request for running analyses only on the first displayed study is a far more conservative and strict test of the original's findings, because of the implications on power, reducing the likelihood of detecting the effect, rather than increasing it. Our experience has been that order had very little to no implications in our many other PCIRRs and replications. In the 7 in-principle accepted PCIRR we conducted and listed above using a similar unified design, we found support for most of the findings in all 7 replications. Order can be easily tested for as moderator, and if there are no order effects, there is little reason to discard doubling the sample.

> **Also, it is critical that participants not know, in advance, that they will be completing multiple studies/conditions on the same question/domain. That is, while participants may be told that they will be completing various studies, they should approach the first sunk-cost scenario without expecting to see additional sunk-cost scenarios (i.e., these should come as a surprise). Otherwise, there could still be spillover effects on the first scenario if subjects know they will be asked about similar things in subsequent parts of the survey. Also, every sunk-cost scenario should appear on a separate (web) page, so that participants' response to the first scenario is completely unaffected by the next one they are presented with.**

We shared a preview to our Qualtrics in the manuscript, and the Qualtrics export files are also made available in the shared Qualtrics in the OSF folder.

To make it clearer, participants are given an outline of the studies at the start which:

> "This survey involves 3 decision-making tasks, the first 2 presented in a random order.
> Please read the scenarios and answer questions carefully.
> In each decision-making task, you will be presented with a short hypothetical scenario.
> There are no "right" or "wrong" opinion answers, so please state your opinion as honestly as possible.
> After that, there you will be asked some questions about your participant and provide demographic information. "

Each study is implemented in separate Qualtrics blocks, which appear in separate pages.

> **Finally, the authors need to make sure that they increase their sample sizes so that they have sufficient power (even) for the between-subjects analyses that only focus on the first sunk-cost scenario that each subject sees.**

We were already well powered to detect weak effects, far weaker than those reported in the original. Aiming to alleviate any further concerns, we went ahead and doubled our planned sample again.

> **Comment 4) "Another concern, which may lead to a failure to replicate the effect, is that experienced MTurk participants may have been exposed (and some repeatedly) to sunk-cost studies, and this may hinder the effect. The authors should therefore consider limiting the study to MTurk participants who have had relatively little experience (e.g., fewer than 100 MTurk studies completed)."**
> **=> The exclusion question/item that the authors propose to use seems far too vague ("Have you ever seen the materials used in this study or similar before?"), especially since it appears at the very end of the survey. The authors should (also) ask more specific recognition questions, along the lines of "Have you participated in any other study/studies that presented a scenario involving a decision to select a project that someone had already invested time or money in? Have you participated in any other study/studies that presented a scenario involving a decision to select an item that someone had already spent time or money to obtain?"**

MTurk participants who have done fewer than 100 studies are too new to the system, and lack experience not only with sunk cost, but in general in being MTurk workers and taking part in studies. Also, they still lack the commitment to high quality work on the platform (with little implications for their reputations), which raises a host of other quality issues. The issue of familiarity can be address by other means.

We added an additional "familiarity check" after Studies 1 and 2, in which we ask participants if they have encountered the scenario before. Those who answer that they have will be excluded.

Details can be found in the supplementary materials as well as on the Qualtrics preview link and Qualtrics exported files.

> **Comment 5) "I don't understand the distinction that the authors are trying to draw, here. Chi-Square tests also evaluate whether likelihoods vary across conditions, so the authors are mistaken if they suggest otherwise. I suspect they meant something else, but that it did not come across clearly in their writing."**
> **=> My confusion stemmed from the fact that the authors seemed to (unintentionally) imply that Chi-Square and Logistic Regression modeled different outcome variables. To avoid confusion, the authors can add (to the revised sentence they propose) that they propose testing an interaction, and that this can only be done with Logistic Regression (or, at least, not with a standard Chi-Square test).**

We revised the manuscript to make this more explicit:

> "To address this question, we conducted logistic regression analyses for Study 2 for both the original and the replication data as it allowed us to test the interaction effect."

Response to Reviewer #3: Dr./Prof. David Ronayne

> **I thank the authors for their responses.**
> **Below I detail some residual issues and suggest remedies. Other than those, I am happy for the authors to proceed with their study.**

Thank you for your valuable feedback in both iterations of revisions.

> **Outstanding issue 1: exclusion criteria. Your response to my and Johanna Peetz's comments clarify things greatly (which was needed - and overviews should be in the final paper even if summarized/footnoted). I now see the practices of the services you use (CloudResearch/Turkprime) help improve data quality, which is of course very welcome (as is your explicit description thereof).**
> **However, I fail to see how that means you should not include explicit and credible criteria within your experiment (i.e., those that allow for how inattention/poor quality may interact with your specific study). The only thing I see that responds to this concern is that you highlight you ask subjects if they will pay attention (yes/no response options), and then ex post whether they were serious (Likert scale) and understood what they read (Likert scale). Asking someone if they intend to pay attention or understood is not the same as checking they did pay attention or did understand, and is not convincing at all.**
> **It would be convincing to include simple and explicit criteria based on hard data such as: a minimum time spent on the experiment's crucial pages (e.g., excluding those who spend less than X seconds on a page); or explicit attention checks (questions subjects know the answer to only if they paid attention - these should be standalone questions aiming to assess attention only, not those that are also part of the experiment, you could also mention at the start that they should pay attention because you will ask comprehension questions).**
> **If your data are guaranteed to be of high quality as you argue strongly in your responses, then they will not bite: your subjects will all pass such criteria and none will be excluded. But your work would be stronger and your results more trustworthy.**

Thank you for the detailed suggestions. We have already implemented some comprehension checks for the education manipulation, which were not present in the original, and already informed participants that comprehension checks will be conducted and they should be paying attention.

Although we are reluctant to include any hard criterion regarding time spent on a page, given its arbitrary nature (e.g. we consider a participant to be legitimate even if they have went to the bathroom in between studies or during the reading of a scenario), we agree that we can include further comprehension checks to test whether participants understood the critical information in the scenarios. We have implemented these checks after each of the studies in Studies 1 and 2 and before asking them if they are familiar with the scenario (see our response to Reviewer 2) as we do not want to detract their attention from the main response. If we were to insert these checks between the scenario and the eventual response, some would consider that to be a major deviation from the original. Instead, we opt to exclude participants post-hoc if they have not understood the scenario. Although that could potentially strain our resources as some participants might be rejected, we are already overpowered (given our new target of 1212 PPs, see our response to Reviewer 2) so that should not be an issue.

The content of the comprehension checks can be found in the "Comprehension checks subsection" under "Materials used" in the supplementary or on the preview link here: https://hku.au1.qualtrics.com/jfe/preview/previewId/af9ed4a0-f30e-4088-8408-07e4b05a8ecd/SV_bNSYF5TiOnh8VrU?Q_CHL=preview&Q_SurveyVersionID=current .

Participants who answer any comprehension check incorrectly were excluded.

> **Outstanding issue 2: Please state explicitly (instead of requiring the reader to go and open your Qualtrics project file): can subjects go back to previous pages of the experiment? In your case it would seem important that they cannot go back (including for explicit attention checks).**

We do not have a previous arrow. We made that explicit in the manuscript:

> "Throughout the study, participants could not go back to previous pages."

Reply to PCIRR decision letter #452:
Soman (2001) replication extension [Stage 2]

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/LNFPRXmuneyU

A track-changes manuscript is provided with the file:
"PCIRR-S2-RNR-Soman 2001-Replication-Manuscript-v1-trackchanges.docx"
(https://osf.io/7y5hk)

Reply to Editor: Dr./Prof. Chris Chambers

> **I have now received evaluations of your Stage 2 manuscript from two of the reviewers who assessed the Stage 1 manuscript. As you can see, Soman is fully satisfied with your submission in its current state, while Olivola lists several points of concern. Given the specific and limited criteria by which Stage 2 submissions are assessed at PCI RR, I want to provide some guidance on how to respond to these points.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

> **Concerning the reviewer's point 2a (table that summarising the subset of prior studies that examined sunk costs involving time), this was not a required element at Stage 1; therefore I am not going to require this at Stage 2, although I fully appreciate the reviewer's point that it would be a useful synthesis for readers; I will therefore leave this at your discretion to consider.**
>
> **Concerning point 2b, factual accuracy and clarity is vital so please ensure that this issue is fully considered and revised as necessary (even if it involves some minor changes to the Introduction).**

We responded to the reviewer on both points in detail below.

> **Concerning point 3: I agree with the reviewer that this would be an interesting exploratory analysis. Since it was not part of the preregistered and approved Stage 1 proposal, and since these analyses are not (in my view) required for the conclusions to be sufficiently supported by the evidence (Stage 2 criterion 2E), it is at your discretion to include these or not. Please consider the point carefully and include in your response to the reviewer a clear explanation for your decision (one way or the other).**

This planned analysis was added during the Stage 1 revisions and has already been implemented in our first Stage 2 submission. As expected, we found no indication for order effects. We expanded on this below.

> **Concerning point 4: There seems to be a general tension with online studies between favouring participants with greater experience vs. favouring participants with a lower chance of prior exposure to key interventions; therefore I think this could be a useful point to consider in the Discussion (particularly if there is evidence that can inform the debate).**

Please see our detailed response to the reviewer below.

Reply to Reviewer #1: Prof. Christopher Olivola

> **Unfortunately, the authors have (again) failed to address many of my comments and concerns:**
>
> **Comment 2a: I still don't fully understand why the authors refuse to provide a table that summarizes the (relatively small) subset of prior studies that examined sunk-costs involving time. Doing so is not a huge endeavor (there are far fewer studies involving investments of time), yet it would provide a summary of all previous attempts to test time sunk-cost effects, and give us an idea of how replicable those effects are. I won't insist on this point, though I do think the authors are doing their paper (and readers) a disservice by limiting the scope of their paper to a replication of just a single paper examining sunk-time effects.**

We understand that we have diverging goals for the manuscript. We explained our views in our reply to the Stage 1 review. We see the value in a summary of the literature yet do not see this fit within the scope of our project. Such a summary should follow a systematic process with clear inclusion and exclusion criteria to avoid situations like the one you mentioned in your next comments indicating a disagreement on whether a study should be included or not. Given that our aim is not to summarize this literature, we prefer to have to engage in the process of having to carefully consider what articles should or should not be included in such a summary.

To address this comment, we added a call for a systematic review of the literature as a future direction in the general discussion in the “Limitations of our replication and directions for future research” subsection:

> In addition to repeating replications, we call for regular systematic reviews and/or meta-analyses that would help summarize the growing literature, address and help explain seemingly discrepant results (such as our differences in patterns of results across Studies 1 and 2), and would drive the development of better theories and empirical investigations.

We are open to changing this if given clear editorial guidelines.

> **Comment 2b: The authors incorrectly claim that Olivola (2018) did not contain any sunk time cost scenarios. To the contrary, that paper contained *multiple scenarios* involving sunk time costs (e.g., the "cello lessons" and "potluck cake" scenarios), including scenarios used in papers that the authors *did* choose to cite. Therefore, there is no reason for the authors to not cite that paper as demonstrating a sunk-cost effect for time. Moreover, that paper demonstrates both intra- and inter-personal versions of time sunk-cost effects (i.e., that we also honor *other people's* time investments), which is notable. This omission by the authors is far less justifiable (and requires even less work to correct) than their decision not to include a summary table.**

Apologies, we were not careful and comprehensive enough in our reply. We assumed these were suggestions for us to consider, and did not realize that a failure to implement these suggestions or explain those in detail would be a point of contention. We do very much appreciate the feedback, and we are open to the possibility of being wrong.

To try and better explain our previous decision to not add Olivola (2018), we note that what we meant was that: 1) it did not compare sunk money vs sunk time costs, which is what we were focused on, and 2) it did not include scenarios that were solely focused on sunk time costs. We will try and elaborate on the second point.

Specifically, in Olivola (2018) (article: https://osf.io/vg9t8/; scenarios available on: https://osf.io/tae9y), experiments 1a (basketball game; free ticket vs $200 ticket), 1b (tennis-club; free membership vs $900 membership), 1c (hotel-TV-movie; free movie vs $19.95 fee), 2 (weekend-trip; cost of tickets), and 3b (airline-investment; $1M investment) were solely focused on sunk money costs.

The other two scenarios: potluck-cake (experiments 1d and 4) and cello lessons (experiment 3a) were more complicated. In the potluck-cake scenario, participants read that "had found the cake on sale for $15 at a bakery located less than 5 min away (low sunk cost) or could only find it for $60 at a bakery located more than 45 min away (high sunk cost)" and in the cello lessons scenario participants read that "[someone] had bought a $100 cello and paid $40 for 1 month of lessons (low sunk cost) or had bought a $1,000 cello and paid $200 for 3 months of lessons (high sunk cost)". Thus, these two scenarios integrated sunk money and some aspect of time investment, rendering it unclear what the interpretation of the findings are for sunk time costs alone.

We are open to changing our handling of this issue if given clear editorial guidelines and further clarifications.

> **Comment 3: I'm glad that the authors have agreed to increase (specifically: double) their sample sizes. However, it seems the authors are resisting my suggestion that they *also* (not solely) carry out *additional* purely between-subjects analyses that only compare the first study/condition that participants were exposed to. The reasons for these *additional* between-subjects analyses were clearly explained in my prior review(s) and I don't see a reason not to include them as well (in addition to the analyses the authors had in mind). The authors' argument that "[their] experience has been that order had very little to no implications in [their] other PCIRRs and replications" is a poor one, and I can quickly counter it with my own experience, which has shown that exposure to prior studies (especially those examining the same concepts) can often impact subsequent responses.**

We have already conducted and reported these analyses in our previous Stage 2 submission, as agreed upon during Stage 1 revisions and in the IPA-ed manuscript.

This is detailed in the "Order effects between studies" subsection in the main manuscript, which points to Table 7 where we summarized these analyses. The relevant code for these analyses can be found in the Rmarkdown in the OSF repository (https://osf.io/qnr2t; subsection "10 Order effects analyses").

Overall, order had no impact on our conclusions.

> **Comment 4: The authors are wrong to claim that participants with 100 studies have no commitment and little reputational concerns. I've carried out several MTurk studies in which I limit the sample to workers with 100 studies or less (but with a high reputation score from those studies they did complete), and found that they are no less attentive than MTurkers who have completed many hundreds of studies. I won't insist on this point, though I think the authors risk not being in a strong position to claim their study is a proper replication of Soman (2001), since his studies involved participants who were likely totally unfamiliar with sunk-cost scenarios, whereas experienced MTurk workers are likely to have been exposed to those scenarios (and debriefed about them).**

We appreciate the feedback and sharing your experience with MTurk. It is an interesting point of debate, but not a crucial one, especially at the current point in time for our manuscript. Data has been collected, the results are rather clear.

We agree that our 100 studies threshold is rather subjective, and we came to that rule of thumb based on our experience with running similar replication studies, many of those already accepted at PCIRR and journals.

We included prompts in every section asking participants if they are familiar with the materials and excluded them if they indicated that they are (N excluded = 22). We explained in our "Exploratory comprehension questions analyses" subsection that we have implemented several methods to incentivise truthful thoughtful responding.

Furthermore, we already reported results in our Stage 2, with one successful replication and one study showing effects for both time and money. Even if somehow someone familiar with the scenarios and the study purpose were to slip through, it is unclear to us what the possible argument is regarding how familiarity would explain the pattern of results. We find it unlikely that participants took the time to consider which condition they're in, that they recalled what pattern of findings experimenters were expecting from their hundreds of studies conducted on MTurk, and to then spent time (costing them money) to devise a response pattern that would meet to explain our mixed results.

Overall, we implemented best practices to try and address familiarity, and we see no concern regarding this point or the relevance to our currently reported findings.

Reply to Reviewer #2: Prof. Dilip Soman

> **The Stage 2 paper is ready for prime time! There were a lot of intriguing results - perhaps none more so than the differences in patterns of results across replicated Study 1 and Study 2. It does appear that time is a lot more malleable that most other resources, so the call for more research on what factors might drive the evaluation of time as a resource is a good one.**
>
> **Thank you to the authors for their excellent work throughout the process!**

We are deeply grateful for your assistance, feedback and encouragement during this process. We are glad that we could contribute to this literature and look forward to more studies to help better elucidate the nature of how people think about time.