

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Brooker, Robin; Lo Iacono, Sergio

Working Paper A Comment on "An Experimental Manipulation of the Value of Effort"

I4R Discussion Paper Series, No. 201

Provided in Cooperation with: The Institute for Replication (I4R)

Suggested Citation: Brooker, Robin; Lo Iacono, Sergio (2025) : A Comment on "An Experimental Manipulation of the Value of Effort", I4R Discussion Paper Series, No. 201, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/311306

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

INSTITUTE for **REPLICATION**

No. 201 I4R DISCUSSION PAPER SERIES

A Comment on "An Experimental Manipulation of the Value of Effort"

Robin Brooker Sergio Lo Iacono

February 2025



I4R DISCUSSION PAPER SERIES

I4R DP No. 201

A Comment on "An Experimental Manipulation of the Value of Effort"

Robin Brooker¹, Sergio Lo lacono¹

¹University of Essex, Colchester/Great Britain

FEBRUARY 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur University of Ottawa Anna Dreber Stockholm School of Economics Jörg Ankel-Peters RWI – Leibniz Institute for Economic Research

E-Mail: joerg.peters@rwi-essen.de RWI – Leibniz Institute for Economic Research Hohenzollernstraße 1-3 45128 Essen/Germany www.i4replication.org

A comment on "An experimental manipulation of the value of effort"

Robin Brooker, Sergio Lo Iacono (University of Essex)

Abstract

Lin et al. (2024) investigates the effect of rewarding effort on task preferences in adult participants using a mixed-design experiment with Bayesian linear regression models. The aim was to recognise whether rewarding effort increases individuals' willingness to choose effortful (harder) tasks and whether this effect persists beyond incentivised contexts. The authors find limited evidence that rewarding effort increases people's willingness to choose more effort task when rewards are no longer offered. Moreover, the authors present mixed-evidence that incentivising effort increased willingness to choose more effortful tasks in a separate unrelated and unrewarded task. We successfully computationally reproduce the main claims of the paper (hypotheses 1-8), but uncover some minor typographical/rounding errors in presentation of the confidence intervals. We further test the robustness of the results to different priors in Bayesian Regression Models using Stan, increasing the number of iterations and independent chains to estimate posterior parameters. Results are robust to these changes to model specification.

1. Introduction

Lin et al. (2024) investigate the effect of reward effort on task preferences in adult participants. The authors sought to test the hypothesis that 'people can learn to value effort and will seek effortful challenges if directly incentivized to do so' (p. 988). The study used a mixed-experimental design with a between-subjects manipulation of reward conditions (rewarded effort vs. rewarded performance vs. no rewards) and a within-subject repeated-measures design. The within-subject repeated measures involved participants engaging in three sequential stages, including a pre-training, training and post-training stage where effort preferences and performance in cognitive tasks were measured. The dataset was collected via Prolific between April 2022 and July 2022. The authors tested the effect of effort-based rewards (X) on preferences for harder tasks (Y) in adult participants (P) using Bayesian linear regression models (M). The authors describe their main findings on p. 994 as follows: the analyses 'provide limited evidence for the idea that people can learn to value effort'.

The authors found that, in rewarded trials, participants in the effort condition were more likely to choose harder tasks than those in the performance condition (H1; b = 0.3495% CI [0.29,0.38], Cohen's d = 1.22, Bayes factor (BF) > 100; P < 0.001) and in the neutral condition (H2; b = 0.26 [0.22, 0.31], d = 0.95, BF > 100; P < 0.001). In testing for near-transfer effects in unrewarded trials, authors argue that preferences were not higher in the effort condition than the neutral (H4; b = 0.05 [0.01, 0.09], d = 0.19, BF = 0.83; P = 0.036) and performance (H3; b = 0.04 [0.00, 0.09], d = 0.17, BF = 0.52; P = 0.066) conditions, as the Bayes Factors (BF) indicate anecdotal support for the null hypotheses. Post-training near-transfer effects were observed on the inhibition task, as participants in the effort condition had higher effort preferences than those in the performance condition (H5; b = 0.08 [0.04, 0.12], d = 0.35, BF > 100; P < 0.001). However, evidence favored the null hypothesis when comparing the effort and neutral condition (H6; b = 0.04 [0.00, 0.09], d = 0.19, BF = 0.92; P = 0.032). Finally, evidence provided some support for far-transfer effects to an updating task. In this instance, participants in the effort condition were more likely to choose harder tasks in the performance condition (H7; b = 0.05 [0.01, 0.08], d =0.26, BF = 6.19; P = 0.004) and neutral condition (H8; b = 0.04 [0.01, 0.08], d = 0.21, BF = 1.24; P = 0.022), though the BF indicates only anecdotal support for the latter effect.

In the present report prepared as part of a collaboration between the Institute for Replication and Nature Human Behaviour (Brodeur et al., 2024), we investigate whether their analytical results are computationally reproducible and further test their robustness to different priors in Bayesian Regression Models using Stan (brms). Lin et. al. (2024) use default brms priors (4 Markov chain Monte Carlo chains, 2,000 iterations and 1,000 warm-up samples), arguing that sufficient data makes priors unlikely to alter parameter estimates. We test the robustness of the main hypotheses (H1-H8) to changes in brms priors. Specifically, we adjust the priors by increasing iterations from 2,000 (as in the original article) to 4,000 and 5,000, and by increasing the number of Markov chain Monte Carlo chains from 4 to 6.

We successfully reproduced Lin et al.'s results for hypotheses H1-H8, although there were minor discrepancies between the confidence intervals reported by the authors and those reproduced in our analysis (in several instances, rounding to the second decimal place did not match values for the third decimal place reported in our analysis). Hypotheses tested matched hypotheses registered before data collection and are available <u>here</u>. Consistent with the authors' expectations, changes in brms priors did not affect results for hypotheses H1-H8.

2. Computational Reproducibility

Complete analysis scripts and clean code were here obtained from package on the Open Science Framework: <u>https://osf.io/9unj5/</u>

Table 1. Bayesian Linear Regressions from original study, computationalreproducibility, and robustness checks

Hypothesis	Original Study	Computational Reproducibility	Robustness I [4,000 iterations]	Robustness II [5,000 iterations]
	Beta Estimate [CIs]; p-value, frequentist approach; Bayes Factor (BF)	Beta Estimate (estimated error) [CIs]; p-value, frequentist approach	Beta Estimate (estimated error)	Beta Estimate (estimated error)
H1: Effort preferences will be higher in the effort condition than	0.34 [0.29, 0.38]; P<0.001; BF>100	0.337 (0.024) [0.291, 0.386]; P<0.001	0.337 (0.024)	0.337 (0.024)

5

performance condition				
H2: Effort preferences will be higher in the effort condition than neutral condition.	0.26 [0.22, 0.31]; P<0.001; BF>100	0.261 (0.025) [0.213, 0.310]; P<0.001	0.261 (0.024)	0.261 (0.025)
H3: Effort preferences will be higher in the effort condition than performance condition.	0.04 [0.00, 0.09]; P=0.066; BF=0.52	0.045 (0.024) [-0.003, 0.091]; P=0.066	0.045 (0.024)	0.045 (0.024)
H4: Effort preferences will be higher in the effort condition than neutral condition.	0.05 [0.01, 0.09]; P=0.036; BF=0.83	0.048 (0.023) [0.004, 0.093]; P=0.036	0.047 (0.022)	0.047 (0.022)
H5: Effort preferences will be higher in the effort condition than performance condition.	0.08 [0.04, 0.12]; P<0.001; BF>100	0.083 (0.021) [0.042, 0.124]; P>0.001	0.083 (0.021)	0.083 (0.021)
H6: Effort preferences will be higher in the effort condition than neutral condition.	0.04 [0.00, 0.09]; P=0.032; BF=0.92	0.043 (0.020) [0.005, 0.085]; P=0.032	0.044 (0.020)	0.044 (0.021)
H7: Effort preferences will be higher in the effort condition than performance condition.	0.05 [0.01, 0.08]; P=0.004; BF=6.19	0.047 (0.017) [0.014, 0.080]; P=0.004	0.048 (0.017)	0.048 (0.016)
H8: Effort preferences will be higher in the effort condition than neutral condition.	0.04 [0.01, 0.08]; P=0.022; BF=1.24	0.040 (0.018) [0.005, 0.075]; P=0.022	0.040 (0.017)	0.040 (0.017)

Table 1 presents the beta coefficients, confidence intervals, p-values, and Bayes factors originally reported by the authors for hypotheses 1–8. It also includes the results from our computational reproducibility analysis and two robustness checks for each hypothesis. We successfully computationally reproduce the main claims of the paper (hypotheses 1-8), but uncover some minor typographical/rounding errors in presentation of the

I4R DP No. 201

confidence intervals. We further test the robustness of the results to different priors in Bayesian Regression Models using Stan, increasing the number of iterations and independent chains to estimate posterior parameters. Results are robust to these changes to model specification.

Regarding computational reproducibility, as shown in Table 1, the beta coefficient for H3, when rounded to two decimal places (i.e., from 0.045 to 0.05), differs from the value reported in the original paper (0.04), with the inconsistency occurring at the second decimal place. Similarly, when rounding the confidence intervals to two decimal places in our reproducibility analysis, discrepancies were observed for H1, H2, H4, and H6 compared to the original study. We speculate that these differences arise from either (a) rounding errors or (b) variations in analytical software producing slightly different results. However, these inconsistencies have no substantive impact on the study's results or findings. Notably, all other beta coefficients, confidence intervals and all p-values remain consistent between the computationally reproduced analysis and the original results. Additionally, no coding errors were identified in the original coding script.

3. Robustness Reproduction and Replication with New Data

To assess robustness through reproduction, we adjusted: a) the number of iterations per chain, and b) the total number of Markov Chain Monte Carlo (MCMC) chains. Specifically, we increased the number of MCMC chains to 6 and estimated the model using 4000 and 5000 iterations per chain, respectively. The original study employed the default settings in the brms package of 4 chains with 2000 iterations per chain. Subject to convergence, increasing the number of iterations can increase the precision of posterior estimates by reducing Monte Carlo error. Moreover, increasing the number of chains can show convergence issues.

Apart from adjustments to iterations and chains, to check robustness, other model specifications remain consistent with the original study. The decision to conduct these two robustness checks was taken after reading the original paper and having checked the computational reproducibility of the main effects reported in the original study.

Posterior estimates and errors for robustness checks are summarised in Table 1. As shown, the Beta estimates remain approximately equal to a) the original study, and b) the

I4R DP No. 201

computationally reproduced study for hypotheses 1-8. The posterior Beta estimates are consistent across models with 4000 and 5000 iterations per chain. Beta estimates are approximately consistently with the computationally reproduced models and the models reported in the original paper for hypotheses 1-8, with some minor discrepancies at the third decimal point for some of the models, which may result from rounding issues. Estimated errors are approximately equal between robustness models, the computationally reproduced model and original study for all hypothesis (agian, notwithstanding slight discrepancies at the third decimal place, which may result from rounding rounding error).

4. Conclusion

We successfully computationally reproduced hypotheses 1-8 from Lin et al.'s (2024) study investigating the effect of rewarding effort on task preferences in adult participants using a mixed-design experiment with Bayesian linear regression models. The original study found limited evidence that rewarding effort increases willingness to choose more effortful tasks when rewards are no longer offered. Full analytical code and clean data were available for analysis, available through the Open Science Framework. The beta estimates, confidence intervals and frequentist p-values were in line with those reported in the original paper. As robustness checks, we increased the number of chains from 4 to 6, and the number of iterations from 2000 to 4000 and 5000. The results from the robustness checks were consistent with the original study. While we computationally reproduced the study, there were some minor rounding discrepancies at 3-decimal places, which may result from software differences or rounding inconsistencies.

Other empirical exercises that can be conducted by other replicators include checking the computational reproducibility with the raw, uncleaned data and estimating models with further adjustments to model parameters and inclusion of other covariates.

This report was done during the UKRN community project's replication games on December 5th, 2024.

References

Brodeur, A. *et al*. Reproduction and replication at scale. *Nat Hum Behav* **8**, 2–3 (2024). https://doi.org/10.1038/s41562-023-01807-2

Lin, H., Westbrook, A., Fan, F. *et al.* An experimental manipulation of the value of effort. *Nat Hum Behav* **8**, 988–1000 (2024). <u>https://doi.org/10.1038/s41562-024-01842-7</u>