

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Fidanoski, Filip; Dixit, Vinayak; Ortmann, Andreas

Working Paper Risky Intertemporal Choices Have A Common Value Function, But A Separate Choice Function

I4R Discussion Paper Series, No. 205

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

Suggested Citation: Fidanoski, Filip; Dixit, Vinayak; Ortmann, Andreas (2025) : Risky Intertemporal Choices Have A Common Value Function, But A Separate Choice Function, I4R Discussion Paper Series, No. 205, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/312256

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. 205 I4R DISCUSSION PAPER SERIES

Risky Intertemporal Choices Have A Common Value Function, But A Separate Choice Function

Filip Fidanoski Vinayak Dixit Andreas Ortmann

February 2025



I4R DISCUSSION PAPER SERIES

I4R DP No. 155

Risky Intertemporal Choices Have A Common Value Function, But A Separate Choice Function

Filip Fidanoski^{1,2,3,4}, Vinayak Dixit⁵, Andreas Ortmann⁵

 ¹Université Côte d'Azur, Nice/France
 ²Centre National de la Recherche Scientifique (CNRS), Paris/France
 ³Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Paris/France
 ⁴Groupe de Recherche en Droit, Économie et Gestion (GREDEG), Sophia Antipolis/France
 ⁵University of New South Wales, Sydney/Australia

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	Anna Dreber	l of Economics	Jörg Ankel-Peters	
University of Ottawa	Stockholm Schoo		RWI – Leibniz Institute for Economic Rese	
E-Mail: joerg.peters@rwi-essen.de RWI – Leibniz Institute for Econom	ic Research	Hohenzollernstraße 1-3 45128 Essen/Germany	www.i4replication	.org

Risky intertemporal choices have a common value function, but a separate choice function

Filip Fidanoski¹, Vinayak Dixit², and Andreas Ortmann³

¹ Université Côte d'Azur, Centre national de la recherche scientifique (CNRS), Institut national de

recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Groupe de recherche en droit,

économie et gestion (GREDEG), Valbonne, Sophia Antipolis, France

² University of New South Wales (UNSW), Research Centre for Integrated Transport Innovation

(rCITI) and Travel Choice Simulation Laboratory [TRACSlab], School of Civil and Environmental

Engineering, Sydney, NSW, Australia

³ University of New South Wales (UNSW), School of Economics, UNSW Business School, Sydney,

NSW, Australia

[18.437 words]

Authors' Note

Acknowledgements: We benefited from discussions with Ashley Luckman, Ben Newell, and Chris Donkin. We thank (alphabetically, by first name) Agnieszka Tymula, Alina Ferecatu, Anna Dreber, Antonio Filippin, Ben Greiner, Davide Dragone, Dmitry Ryvkin, Ernan Haruvy, Ennio Bilancini, Ganna Pogrebna, James C. Cox, Kiril Simeonovski, Lana Friesen, Marco Casari, Maria Bigoni, Riccardo Ghidoni, Sandro Ambuehl, Stefania Bortolotti, Stephen Cheung, Vinayak Alladi, Xuiting Wang, and other attendees at the Sydney Experimental and Behavioural Research Group Seminar Series, the 2021 New Institute for Replication

I4R DP No. 205

Zealand Association of Economists [NZAE] Virtual PhD Workshop, the 2022 Virtual Asia Pacific Economic Science Association [ESA] Meeting, the 15th Australia and New Zealand Workshop on Experimental Economics in Sydney, the 2022 Workshop on Model Evaluation and Causal Search: Empirical and Experimental Approaches in Pisa, the 2022 North-American ESA Meeting in Santa Barbara, Economics Seminar at the University of Bologna, Barcelona School of Economics [BSE] Summer Forum Workshop on External Validity, Generalizability and Replicability of Economic Experiments in Barcelona, 2023 ESA World Meeting in Lyon, NiCE Seminar Series at the Nijmegen School of Management in Nijmegen, Masaryk University Economic Seminars [MUES] in Brno, Behavioral and Experimental Lab Research Seminar at École supérieure des sciences commerciales d'Angers [ESSCA] in Lyon, Montpellier Business School/Centre d'Economie de l'Environnement-Montpellier (CEE-M), CNRS-GREDEG at the University of Côte d'Azur, 16th Australia and New Zealand Workshop on Experimental Economics in Melbourne, Max Planck Institute for Human Development in Berlin, and the Laboratory for the Analysis of compleX Economic Systems (AXES) at the IMT School for Advanced Studies Lucca. We are indebted to Domagoj Dorešić for his help in fixing the code for the Bayesian Model Selection and migrating it to Python. Thanks go to Cicy Zhou and Elena Keller for editorial suggestions. We own all errors in the paper, or in the codes.

Funding: We acknowledge gratefully the partial funding of this research project through an Australian Research Council (ARC) Linkage grant to Ortmann et al. (LP 150100608), the UNSW Scientia PhD Scholarship, and the Australian Government Research Training Program Scholarship (HDR Completion Scholarship) to Fidanoski. Fidanoski's work has also been supported by the French government through the France 2030 investment plan managed by the National Research Agency (ANR), as part of the Initiative of Excellence of Université Côte d'Azur under reference number ANR-15-IDEX-01.

Conflict of interest/ Competing interests: We have no financial or non-financial conflicts of interest to disclose that are directly or indirectly related to the work submitted for publication.

Ethics approval: The experiment has obtained an approval from the Human Ethics Committee Panel [HREAP] F: UNSW Business School, University of New South Wales in Sydney on 28-May-2021 under HC210406.

Consent to participate: The Consent Form was signed/checked by the participants in an electronic form.

Consent for publication: Participants in the experiment signed/checked the Consent Form. There is no identifiable information which could be linked to the experimental data or code.

Data and code availability statement (repository)/Open Science practices statement: This study was pre-registered with the Open Science Framework [OSF] (OSF identifier: zq2ap) before data were collected. Supplemental Materials (appendices) and Online Materials are available at the following link: https://osf.io/wyq7v/

Corresponding author: Correspondence concerning this article should be addressed to Filip Fidanoski, University of Côte d'Azur, CNRS, INRAE, GREDEG, Valbonne, France. Email: Filip.FIDANOSKI@univ-cotedazur.fr.

Author contributions (CRediT statement): Conceptualisation (Filip Fidanoski, Vinayak Dixit, Andreas Ortmann), Data curation (Filip Fidanoski, Vinayak Dixit), Formal analysis (Filip Fidanoski, Vinayak Dixit), Funding acquisition (Andreas Ortmann), Investigation (Filip Fidanoski, Vinayak Dixit, Andreas Ortmann), Methodology (Filip Fidanoski, Vinayak Dixit, Andreas Ortmann), Project administration (Filip Fidanoski, Andreas Ortmann), Resources (Andreas Ortmann), Software (Filip Fidanoski), Supervision (Vinayak Dixit, Andreas Ortmann), Validation (Filip Fidanoski, Vinayak Dixit), Visualisation (Filip Fidanoski), Writing — original draft (Filip Fidanoski), and Writing — review and editing (Filip Fidanoski, Andreas Ortmann).

Highlights

- We tested the reductionist assumption that risky choices and intertemporal choices are simply a special case of risky intertemporal choices.
- In our conceptual replication of Luckman et al. (2018), we found that while a common value function can accommodate both risky and intertemporal choices, the choice function does not.
- We observed significant differences between risky and intertemporal choices at both the group and individual levels.
- At the group level, there is no significant difference between the choices made by participants paid a flat incentive versus those paid under a random incentive system.
- Order effects matter for intertemporal choices, but not for risky choices.

Institute for Replication

14R DP No. 205

Abstract

Luckman et al. (2018) experimentally tested the conjecture that a single model of risky intertemporal choice can account for both risky and intertemporal choices, and under the conditions of their experiment, found evidence supporting it. Given the existing literature, that is a remarkable result which warrants (conceptual) replication. Following a tradition in psychology, Luckman et al. (2018) had firstyear psychology students participate that were rewarded with non-monetary course credits (see also Luckman et al., 2020). Proper incentivisation is a long-standing bone of contention among experimentally working economists and psychologists, last but not least when it comes to the elicitation of preferences of any kind. Another reason to be sceptical is that the experiment was not properly powered up; the no-difference results reported by the authors might be spurious. In our conceptual replication of Luckman et al. (2018), we find significant differences between the risky and intertemporal choices at both the group and individual level. We find further that there is no significant difference between choices made by participants that are paid a flat incentive and participants that are paid under the random incentive scheme, at the group level. We find that order effects matter for intertemporal choices, but not for risky choices. At the individual level, we find evidence in favour of the model that assumes a common value function, but separate choice functions. This result is robust across our incentive systems, and order of presentation, but sensitive to different prior distributions. Keywords: experimental practices, replication, risky intertemporal preferences, risk preferences, time preferences

JEL Classification: C11, C52, C91, D01, D81, D90

14R DP No. 205

Risky intertemporal choices have a common value function, but a separate choice function

The relation between risk and time preference remains one of the most topical and important research questions in economics, psychology, and other social sciences. The reason for this ongoing and growing interest potentially lies in the welfare effects produced by risk and time preferences (Harrison, 2011; Charness & Viceisza, 2016). We document below that there is at best mixed evidence in favour of the claim made by Luckman et al. (2018) that pure risky decisions and pure intertemporal decisions are just a special case of risky intertemporal choice (i.e., the reductionist assumption). Considerable evidence suggests that monetary incentives matter especially when it comes to the elicitation of risk preferences. Many studies, notably Holt and Laury (2002) and Harrison et al. (2005a), suggest that outcomes of risky choices are different when participants face real monetary consequences rather than hypothetical payoffs, especially when such monetary consequences are substantial. Meanwhile, the replicability of experimental findings remains a challenge, as demonstrated by what is widely considered a replication crisis in the social sciences (Ortmann, 2021). The replication crisis, and research such as loannidis et al. (2017), have highlighted the need for experiments to be properly powered up following the well-known exhortation of Cohen (1962) that researchers should only conduct studies with sufficient statistical power to detect true effects.

Below we report the results of our attempt to replicate Luckman et al. (2018) – who, as did Luckman et al. (2020) which draws significantly on the earlier study, used non-monetary incentives (credit points) — in both non-incentivised and incentivised laboratory settings, and with a larger sample to boot. We check for the (conceptual) replicability of the original study and test whether a single model that captures both risky and intertemporal decisions by a single value and a single choice parameter outperforms other models. For reasons explained below, a direct replication was not possible.

We address the following research questions: 1) can we replicate the results in Luckman et al. (2018) in a non-incentivised laboratory setting; 2) do monetary incentives shift risky and intertemporal choices; and 3) can a single model account for both risky and intertemporal decisions once choices are incentivised?

We failed to replicate the results from Luckman et al. (2018) at the group level in both incentivised and non-incentivised laboratory settings. Moreover, we found order effects of administering the experimental tasks for intertemporal choices but not for risky choices.

Specifically, we find that there are statistically significant differences between risky and intertemporal choices, as measured by the number of safer choices in the Risky Task and the number of sooner choices in the Intertemporal Task. Contrary to our expectations (as specified in our pre-registration at the OSF), we failed to find (strong statistical) evidence in favour of the hypothesis that monetary incentives matter for risky choices and intertemporal choices at the group level. This is in line with Hackethal et al. (2023) who also report statistically insignificant differences for risky choices between, and within, participants in their incentivised and non-incentivised schemes. Similarly, the main takeaway message from Brañas-Garza et al. (2023) is that paid and hypothetical intertemporal choices are mostly the same. Drawing on the results reported by Holt and Laury (2002) and Harrison et al. (2005a), we conjecture that these previous findings, and our results, are vulnerable to the stakes being used.

At the individual level, we found that the overwhelming majority of participants' decisionmaking can be explained by a semi-nested model with a common value power function but with separate choice (logit) functions. More details are available in the section **Results**. In other words, we found that one common value function accounts for both risky and intertemporal choices but the choice functions for risky and intertemporal choices differ. This result is robust across the two incentive

systems we use, and the order of presentation. However, other semi-nested models are sensitive to incentives and order effects.

It is important to understand under what design and implementation choices data were generated before we can have confidence in evaluations and comparisons of models of risky intertemporal choices. Any such horse race is a function of those design and implementation choices.

We motivate, and elaborate on, our experimental design and implementation choices in section **Experiment**, and discuss our estimation strategy in section **Estimation Strategy**. We next present results from the estimation. Lastly, we connect our work to the related literature, summarise our findings, discuss the limitations of our replication, and sketch out a research agenda.

Experiment

Following Luckman et al. (2018), we had two main tasks in the experiment: one task involved risky choices [Risky Task] and the other involved intertemporal choices [Intertemporal Task]. Participants were asked to make a choice between two options with different probabilities in the Risky Task and to make a choice between two options with delayed times in the Intertemporal Task. Both Risky Task and Intertemporal Task involved 130 choices followed by comprehension questions and three scales about confidence, effort, and complexity. The latter is particularly important because the reliability of the elicitation procedures hinges heavily on the level of comprehension (Charness et al., 2013).

Going beyond Luckman et al. (2018), we added the standardised version of the Berlin Numeracy Test [BNT] (Cokely et al., 2012) to gauge risk literacy and probability understanding of our participants, and we added questions from the Big-5 test on key financial literacy concepts used in the Household, Income and Labour Dynamics in Australia [HILDA] survey (Wilkins & Lass, 2018) which draws on Lusardi and Mitchell (2011). It involves questions that test the understanding of the compounding of interest

rates, inflation rate, risk diversification, risk and return, and money illusion. We also included an Exit Survey [ES] through which we collected information regarding socio-demographic and personal characteristics that are relevant and might affect risk-preference and the time-preference estimates. We collected, in the background, information about participants' response time (Spiliopoulos & Ortmann, 2018).

The authors of the original study could not retrieve the original instructions, so we reengineered them [**Appendix E**] based on information from Luckman et al. (2018) and our communication with them. The Risky Task and Intertemporal Task, BNT survey, Big-5 survey/test, ES are included in **Appendix F**, **Appendix G**, **Appendix H**, and **Appendix I**, respectively.

Sample Size, Transparency and Openness

Given the importance of effect size in replication projects, we followed the 2.5 rule of thumb proposed by Simonsohn (2015), and considering the treatments in our study, our experiment needed 180 participants per (non-)incentivised group.

The total sample size for the project was 360 participants. This sample size was necessitated by the following samples for each of our two participant groups (and therein for three subgroups, or cells each, to be discussed below): 1) Participant Group I [FLAT/participants were paid under a flat incentive scheme (180 participants)]; 2) Participant Group II [RANDOM/participants were paid randomly selected choices/Random Incentive System (180 participants)]. The sample sizes for each of the cells (and the groups) are sufficient to answer our research questions based on Simonsohn (2015); they also correspond to sample power calculations based on the Bayes Factor Design Analysis (elaborated on in detail in **Appendix J**).

Participants were recruited through Online Recruitment System for Economic Experiments [ORSEE] (Greiner, 2015) from an existing database of volunteers at the Business Experimental Research Laboratory [BizLab] of the UNSW (180 participants) and from the Prolific platform (180 participants).

Following best research practices (Christensen et al., 2019), we pre-registered the main hypotheses of the experiment before data were collected.

Design and Incentives

The design of the experiment is 2x3 factorial: apart from varying the incentive system, following Luckman et al (2018), we vary the order of administering tasks for both incentive systems. **Figure 1** summarises the experimental design.

Figure 1

Experimental design



Participants in the FLAT conditions (second and third level, left) were paid under a flat incentive scheme, in line with UNSW BizLab and Prolific guidelines. Those in the RANDOM conditions (second and

third level, right) were paid based on only one randomly selected choice¹ from one randomly selected task [Risky Task and Intertemporal Task]. The RANDOM conditions correspond to the "pay one" system discussed by Charness et al. (2016), who suggests that it is an adequate and efficient system for the simple pairwise lotteries. The treatment cells on the third level constitute a between-subject design.

The expected value of the RANDOM payments was calibrated to be roughly equal to the payment in the FLAT conditions. Participants paid under the RANDOM scheme could earn up to AU\$57 in addition to a show-up fee of AU\$5. Average earnings in the experiment, as shown in **Table K3** in **Appendix K**, were AU\$27 in the RANDOM condition and AU\$25 in the FLAT condition.

In the replication project, we consider both horizontal [blocked] and vertical randomisation [mixed]. For both the FLAT conditions and the RANDOM conditions, the first group of randomly chosen participants made choices in the *Risky Task first* [RTF] while, to control for order effects, the second group of randomly chosen participants made choices in the *Intertemporal Task first* [ITF]. This corresponds to the blocked condition² from the original study of Luckman et al. (2018), explained in

¹ We programmed three random devices. The first random device chooses one choice from the two tasks, the second random device chooses the pay-off relevant option from each of the two tasks and determined the potential earnings, and the third random device chooses the specific task – potential earnings from only one task has been selected – that determined the final earnings.

² Note that Luckman et al. (2018) ultimately collapsed the data. Nevertheless, we argue that the order effects are inextricably linked to the decision-making in the Risky Task and in the Intertemporal Task, particularly if we bear in mind that the participants are asked to make a large number of choices [130 pure risky choices and 130 pure intertemporal choices]. Moreover, since Luckman et al. (2018) presented both blocked choices [36 participants] and mixed choices [36 participants], they had 6 participants per order [36 participants in the blocked condition divided by 6 orders] which we believe to be insufficient to detect order effects if indeed they exist.

detail in his dissertation (Luckman, 2016) [**Experiment 2.2**]. Our *Mixed* [M] condition corresponds to Luckman et al. (2018, p.788) where the mixed section "contained a mixture of all choice types". We recruited 120 participants each [60 participants in RANDOM condition and 60 participants in FLAT condition] for both the RTF and the ITF condition. Following Luckman et al. (2018), we call this the blocked or horizontal randomisation. For the M condition, we also recruited 120 participants [60 participants in the RANDOM condition and 60 participants in the RANDOM condition and 60 participants.

To summarise, all participants faced two tasks [Risky Task and Intertemporal Task] in all treatments but there were three conditions³:

- RTF condition: 2 Tasks ([130 pure risky choices in the Risky Task] and [130 pure intertemporal choices in Intertemporal Task]). The list of choices is fixed and static for both tasks. The only difference is that the risky choices are the first task, and the intertemporal choices are the second task.
- ITF condition: 2 Tasks ([130 pure risky choices in the Risky Task] and [130 pure intertemporal choices in Intertemporal Task]). The list of choices is fixed and static for both tasks. The only difference is that the intertemporal choices are the first task, and the risky choices are the second task.

³ We initially considered varying the information set as well. Harrison et al. (2002), Harrison et al. (2005b), Andersen et al. (2008) and Andersen et al. (2014) provided participants with information about the annual implied rate in order to reduce comparison errors. They cite Coller and Williams (1999) who found that providing that information has a significant negative effect on discount rates. Therefore, we considered enriching the information set for our participants and to create an INFO and NOINFO condition. But such a design would have required us to recruit 720 participants, which was not feasible due to the very low response rate in our laboratory during the COVID pandemic as well as our financial constraints. We did include comprehension questions and inserted some information from our initially planned INFO treatment into the instructions.

M condition: 2 Tasks ([TASK 1 - 65 intertemporal & 65 risky choices] and [TASK 2 - 65 intertemporal & 65 risky choices]). In other words, both tasks contain a mix of both risky and intertemporal choices. For example, in Task 1, the first choice is a pure risky choice, the second choice is a pure intertemporal choice, and so on. Similarly, in Task 2, the first choice is a pure intertemporal choice, the second choice is a pure intertemporal choice, the second choice, and so on. The list of choices is fixed and static for both tasks.

Since we conducted the experiment online, we used PayID, a payment mechanism which all major banks in Australia offer, for all the payments of participants recruited through the ORSEE pool. Participants that were recruited from the Prolific pool were paid using the Prolific platform as required by the platform. The Prolific platform subsequently transferred these earnings through the PayPaI system. Participants received their payoffs within hours – and on the same day – after the end of the (online) experiment in the FLAT treatment. Participants in the RANDOM condition were likewise paid within hours after the end for the Risky Task and based on the delay chosen in the randomly selected option in the Intertemporal Task.

Tasks

In the original study, Luckman et al. (2018) tested the reductionist assumption that pure risky choices and pure intertemporal choices are special cases of risky intertemporal choices. They did this by presenting a group of participants with both risky choices and intertemporal choices. They then compared the performance of a model fitted to both choice types simultaneously, with the performance of separate models fitted to the risky choice data and the intertemporal choice data, respectively.

Participants were asked to make choices between two options [the LEFT option and the RIGHT option] with different probabilities in the Risky Task and between two options with different delays in

Institute for Replication

the Intertemporal Task [the LEFT option and the RIGHT option]. Both Risky Task and Intertemporal Task involved 130 choices and were related to each other, as follows (Luckman et al., 2018, p.787):

- **Risky Task**: "The probability of gaining money ranged from 0.05 to 1 in increments of 0.05, and the amount of money ranged from \$50 to \$475 in \$25 increments. The probability of receiving \$0 therefore ranged from 0 to 0.95 in increments of 0.05. Choices were generated by crossing 5 probability levels (very low [0.05 - 0.2], low [0.25 - 0.4], moderate [0.45 - 0.6], high [0.65 - 0.8], very high [0.85 - 1]) with 6 amount levels (very low [\$50 - \$100], low [\$125 - \$175], moderately low [\$200 - \$250], moderately high [\$275 - \$325], high [\$350 - \$400], very high [\$425 - \$475]) for each of the two gambles, yielding 435 unique choice pairings. The specific outcome amount and probability for each option was then determined by drawing a value randomly from the appropriate range. Similar to Rieskamp (2008), we removed all choices in which one option either dominated the other or had twice the Expected Value of the other. This process resulted in 125 choices. Five dominated choices were then added to act as check questions."

— Intertemporal Task: "To create intertemporal choices, rather than generate a new choice set, we replaced each probability level in the risky choices with a matching delay and removed the \$0 outcomes. Because we used Expected Value to reduce our risky choice set, we decided to match our risky and intertemporal choice sets by ensuring that the Discounted Value of the intertemporal options approximately matched the Expected Value of their corres ponding risky choice option. This required the setting of an exponential discount rate. We set the discount rate by fitting an Exponential Discount function to the mean intertemporal choice data from the experiments reported in Luckman, Donkin, and Newell (2015), where we had used a similar group of participants and similar ranges of amounts. The fitted rate was 0.053."

We attempted to closely replicate the Luckman et al. (2018) study by replicating their Intertemporal Task and Risky Task. However, their time frame and incentivisation strategy, as well as our participant pool and lab guidelines required adjustments that are explained in more detail in the following subsection and in the **Supplementary Materials**.

Adjustment Procedures

Figure 2a and Figure 2b

In the Intertemporal Task constructed by Luckman et al. (2018), matching delays were linked to probabilities in the Risky Task. In the original version, the sooner option ranged from 0 [immediate] months to 42 months and the later (delayed) option ranged from 3 months to 54 months. The distribution of the delays in the options in the original experiment is shown in **Figure 2a** and **Figure 2b**. We tried to replicate the delays from the original settings, but for reasons elaborated in this section below, we had to make adjustments.



Delays in the LEFT option and in the RIGHT option in the Intertemporal Task

Luckman et al. (2018) used the classical Exponential Model (Samuelson, 1937) with continuous compounding (see Doyle, 2013, for the review of discounting models), that accommodates consistent

preferences⁴, $P = Ae^{-\rho t}$, where ρ and t denote the discount rate and the delay, A is the delayed amount at the second option [the RIGHT option], and P is the delayed amount at the first option [the LEFT option]. In Luckman et al. (2018) a delay of 27 months was not converted to $\frac{27}{12} = 2.25$ years but was expressed in months. Following Luckman et al. (2018), we calculated the net-present value [NPV] difference between the RIGHT and the LEFT option in this way:

$$\Delta NPV_1 = A_2 e^{-\rho t_2} - A_1 e^{-\rho t_1} \tag{1}$$

where A_2 and A_1 , and t_2 and t_1 , are the amounts and the delays in the RIGHT option and the LEFT option, respectively. Based on their previous work (Luckman et al. 2015), Luckman et al. (2018) set the discount rate $\rho = 0.053$.

Further, they calculated the expected values of the corresponding options:

$$\Delta EV_1 = p_2 A_2 - p_1 A_1 \tag{2}$$

where A_2 and A_1 , and p_2 and p_1 , are the amounts and the probabilities in the corresponding the RIGHT option and the LEFT option, respectively.

By using this design, they wanted to set $A_2 e^{-\rho t_2} = p_2 A_2$ and $A_1 e^{-\rho t_1} = p_1 A_1$, to ultimately have $\Delta NPV_1 \cong \Delta EV_1$. Since the amounts and probabilities were taken from predefined ranges, Luckman et al. (2018) had to round up some of them, which led to minor discrepancies. The differences in the net-present values (vertical axis) and the differences in the expected values (horizontal axis) are plotted and shown in **Figure 3**. The number of negative values is 75 and the number of positive values is 55, as you can see in **Figure A1** in **Appendix A**.

⁴ The Exponential Model is highly correlated with the Hyperbolic Model (Mazur, 1987), if one does not construct stimuli that will make them orthogonal (Doyle et al., 2011). Yet, the Hyperbolic Model produces rates steeper at short delays and flatter at long delays (Kirby & Marakovic, 1996).

Figure 3

The differences in the original net-present values and the differences in the original expected values



From **Figure 3**, we see that indeed $\Delta NPV_1 \cong \Delta EV_1$.

Because the options in the original study were hypothetical and involved timeframes that would

be not-implementable/non-credible with incentivised choices, we had to make changes and

adjustments (Figure 4). We faced several issues elaborated in more detail in Appendix D.

Figure 4

The differences in the non-adjusted net-present values and the differences in the non-adjusted expected values



Institute for Replication

Luckman et al. (2018) used the classical Exponential Model with continuous compounding (Doyle, 2013), as described in (1). In contrast, the "workhorse" model in the economics literature is Exponential Discounting with linear instantaneous utility, elaborated below:

$$DU_2(A_2, t_2) = \frac{A_2}{(1+\rho)^{t_2/52}}$$
(3)

and

$$DU_1(A_1, t_1) = \frac{A_1}{(1+\rho)^{t_1/52}}$$
(4)

gives

$$\Delta NPV_3 = DU_2(A_2, t_2) - DU_1(A_1, t_1) = \frac{A_2}{(1+\rho)^{t_2}/52} - \frac{A_1}{(1+\rho)^{t_1}/52}$$
(5)

While we prefer the specification in (5) over (1), it produces many differences in the construction of the Risky Task and Intertemporal Task. (Ultimately, they are equivalent.) Therefore, we use the classical Exponential Model (1) with continuous compounding, from the same family of discounting models, to be consistent with the set-up of Luckman et al. (2018).

There were several other problems. Specifically, we had to modify the Risky Task and the Intertemporal Task by implementing three main principles: i) $\Delta NPV_3 \cong \Delta EV_3$; ii) $\#(\Delta NPV_1) \cong$ $\#(\Delta NPV_3)$, $\#(-\Delta NPV_1) \cong \#(-\Delta NPV_3)$, $\#(\Delta EV_1) \cong \#(\Delta EV_3)$, and $\#(-\Delta EV_1) \cong \#(-\Delta EV_3)$; and iii) $p_1 > p_2$, $t_2 > t_1$, and $A_2 > A_1$. Following Luckman et al. (2018), only five choices [125 – 130 in the **Appendix F**] deviated from these principles in (iii) and they are used as attention checks, i.e., we did not analyse the data points from participants who "failed" on these choices. (A detailed elaboration on these principles is available in **Appendix D**.) **Figure 5a**, **Figure 5b**, **Figure 6a**, and **Figure 6b** illustrate that the implementation of these principles was successful and promised proper replication of Luckman et al. (2018) in line with their principles.

Figure 5a and Figure 5b

The differences in the adjusted net-present values and the differences in the adjusted expected values



Figure 6a and Figure 6b The differences $\Delta NPV_1 - \Delta EV_1$ and the differences $\Delta NPV_3 - \Delta EV_3$ (in dollars) respectively



Display

Participants in the Intertemporal Task, as well as in the Risk Task, faced two options, labelled "LEFT option" and "RIGHT option". To correspond to Luckman et al. (2018), choices were given in descriptive form i.e., in words and numbers. The task consisted of 131 choices, where the first choice was a practice round for participants to familiarise themselves with the task. In other words, we acknowledge the importance of gaining experience with the tasks (Charness & Chemaya, 2023; Charness et al., 2023), which

is why we included a practice round. However, we recognise that repeated and more extensive exposure is desirable and more meaningful, as it enhances consistency in decision-making (Charness & Chemaya, 2023; Charness et al., 2023). Still, we did not want the experimental procedures to deviate substantially from the experiment we set out to replicate.

An example of the Intertemporal Task is shown in **Figure 7a**, and an example of the Risky Task is shown in **Figure 7b**.

Figure 7a and Figure 7b

Intertemporal Task and Risky Task



Surveys

In the final part of the experiment, we went beyond replication and tried to identify the determinants of the decision-making in the Risky Task and the Intertemporal Task. We included three surveys: BNT, Big-5 survey of financial literacy, and the ES. The BNT and the Big-5 survey of financial literacy were meant to help us to peg the numeracy and literacy level of the participants in the experiment, whereas the ES helped us to collect information regarding socio-demographic and personal characteristics that are known to be relevant and potentially might affect risk-preference and time-preference estimates. The BNT, the Big-5 survey of financial literacy, and the ES are available in **Appendix G, Appendix H**, and **Appendix I**, accordingly. The questions in the ES are based on a survey

that we used in an experiment approved by HREAP: F under HC190793 [reduced]; it contains 30 questions, but it was necessary to slightly adjust it for the Prolific participants to answer their requirements. The BNT test contains 6 questions while the Big-5 survey of financial literacy contains 5 questions. We also collected information about participants' response time, as in Konstantinidis et al. (2020).

All information collected was treated as private and confidential. Identifying information was replaced with codes that were randomly generated. Specifically, before the experiment started, every participant was randomly assigned a code ["Participant number"] that s/he was asked to enter in the experiment. The list connecting participant number and participant information is kept safely (accessible only to the researchers) and separately.

Our replication experiment has both within-subject and between-subject design components, given that both of them have practical advantages and limitation (Charness et al., 2012), and it was conducted in an online environment to avoid potential COVID-19 disruptions and to address low laboratory response rates. The experiment was programmed and administered in z-Tree [Zurich Toolbox for Ready-made Economic Experiments] unleashed (Fischbacher, 2007; Duch et al., 2020) except for the "Withdrawal-of-Consent form" which was programmed in Qualtrics. The average completion time of the experiment was 54 minutes (median: 52 minutes, mode: 48 minutes), including the surveys (see more details in Table K3 in the Appendix K).

Regarding the earnings, as shown in in **Table K3** in the **Appendix K**, the participants paid under the flat incentive scheme received a fixed renumeration for their participation. On average, those participants recruited from the UNSW ORSEE pool, who were paid under the random incentive scheme, earned \$27.60 (median: \$28, mode: \$10) [RTF], \$24.07 (median: \$23.50, mode: \$10) [ITF], and \$32.13 (median: \$31.50, mode: \$10) [M]. For the participants recruited from the Prolific platform, the average earnings were \$26.40 (median: \$24, mode: \$10) [RTF], \$29 (median: \$29, mode: \$10) [ITF], and \$24.87

14R DP No. 205

(median: \$18.50, mode: \$10) [M]. In Fidanoski et al. (2024) are compared the choices of ORSEE and Prolific participants and established that there is no substantial difference in choice behaviour.

Estimation Strategy

Our estimation strategy follows Luckman et al. (2018) closely. We apply Bayesian Model Selection based on Bayes factors to discriminate between four competing semi-nested models that incorporate risky choices and intertemporal choices. Concretely, one model assumes a single value and choice function for both risky and intertemporal choices, another model assumes a separate value and choice function for each set of choices, and two intermediate models test whether one of these functions holds constant and the other varies. The models differed on whether the value function parameter, and the choice function parameter, were constrained to be the same for risky choices and intertemporal choices.

Models

We use Prospect Theory and the classical Hyperbolic Model in the estimation⁵. Luckman et al. (2018) relied on Prospect Theory because, if all outcomes occur immediately (pure risky choices), the

⁵ Ahlbrecht and Weber (1997) suggest that the Exponential Model is valid and more accurate for pairwise choices (as in our experiment), whereas the Hyperbolic Model is more appropriate for price lists. Going beyond a simple replication, in the separate follow-up project we also apply the classical Exponential Model with continuous compounding (1) and the Exponential Model with a linear instantaneous utility (3) as well as the estimation of the parameters through the Hierarchical Bayes framework. This is important because we based the construction of the tasks on the Exponential Model with continuous compounding (1).

Probability and Time Trade-off model (Baucells & Heukamp, 2010), the Multiplicative Hyperboloid Discounting model (Vanderveldt et al., 2015), and the Hyperbolic Discounting model of Yi et al. (2006) can be approximated by Prospect Theory (Kahneman & Tversky, 1979). In other words, considering that these models in their reduced form converge and can be explained by Prospect Theory, Hyperbolic Discounting, or both, Luckman et al. (2018) implemented the principles of Prospect Theory for the Risky Task, and the principles of the Hyperbolic Model for the Intertemporal Task.

Prospect Theory, Hyperbolic Model, Exponential Model with continuous compounding (1), and the Exponential Model with a linear instantaneous utility (3) are based on the evaluation of the respective utilities. Researchers in Prospect Theory are transforming the objective values into subjective values. Let us consider the amount for the LEFT option, A_1 . Exploiting Prospect Theory, we transform the objective amount A_1 via value function $v(A_1)$. The probabilities are also transformed into a decision weight by using a probability weighting function w.

A parsimonious utility model which embodies the value function, the probability weighting function, and the choice function is implemented here. Since we want to be consistent with the original study, we use a general representation, i.e., we denote the amount(s)/outcome(s) as x, probabilities as p, and the option as g. The corresponding utility of a particular option, in the general case, can then be expressed as:

$$U(g) = w(p)v(x) \tag{6}$$

A standard assumption is that the value function is concave in shape. Accordingly, we define the value function in the following format:

$$v(x) = x^{\alpha} \tag{7}$$

where α is a parameter of interest estimated from the observed data points. The parameter α ranges from 0 to 1. This follows Luckman et al. (2018) who pointed out that the value functions are

typically assumed to be concave, thus giving a rise to the range values of the parameter of interest that is attached to the value function.

Next, we need to stipulate a functional form for transforming the second component of the utility function in (6): probabilities. We assume that participants overweight small probabilities and underweight large probabilities which implies a *S*-shaped function. Specifically, we use Prelec's specification (1998) of Prospect Theory:

$$w(p) = e^{-(-lnp)^r} \tag{8}$$

where r is the parameter estimated from the observed data.

Luckman et al. (2018) were the first to include a choice or stochastic/noise function in the analysis of the risky intertemporal framework. The choices of participants do indeed exhibit typically stochastic patterns (see Harrison & Rutström, 2008, for more details). Luckman et al. (2018) analysed whether risky choices and intertemporal choices can be explained by a common choice function. Consequently, we also use a logistic choice function for the LEFT option (g_1) and the RIGHT option (g_2):

$$P(g_1, g_2) = \frac{1}{1 + e^{-s(U(g_1) - U(g_2))}}$$
(9)

where *s* is the free parameter from the choice function that we infer from the data.

Doyle (2013), in his exhaustive review, demonstrate that there are various models that explain time preferences. An important take-home message is that, by construction, most of the models share some common features: they consist of a value function, a discount function, and a choice function. We use the classical Hyperbolic Model to find an answer to our research questions.

The discounting function in the classical Hyperbolic Model (Mazur, 1987) is:

$$d(t) = \frac{1}{(1+ht)}$$
(10)

where t denotes the time delay and h is the parameter that is estimated from the data.

By the same token, we establish the value function as in (7), $v(x) = x^{\alpha}$, constraining α to be larger than 0, which differs from previous work but is important for our research question. As in the case of the Risky Task, we evaluate the utility in the Intertemporal Task as:

$$U(g) = d(t)v(x) \tag{11}$$

In the final step for the Intertemporal Task, we include the utilities in the logistic choice function, as we did in the Risky Task, that is same as the one in (9): $P(g_1,g_2) = \frac{1}{1+e^{-s(U(g_1)-U(g_2))}}$.

Analysis/Methods

We tested four semi-nested models on the data from the experiment. These models were identified by the two main parameters discussed in the previous sub-section: the parameter of the value function, α , and the parameter of the choice function, s. The difference is determined by the decision to either constrain these parameters to be the same for the risky and intertemporal choices or to let them be different. If the model is constrained to have a single-value parameter α for both the Risky Task and the Intertemporal Task we call it the Common value [Cv] model, otherwise we call it the Separate value [Sv] model with α_r and α_i for the Risky Task and the Intertemporal Task, respectively. Analogously, if the model is constrained to have a single choice parameter s for both the Risky Task and the Intertemporal Task and the Intertemporal Task is denoted as Common choice model [Cc]. The model that has separate choice parameter s_r for the Risky Task and separate choice parameter s_i for the Intertemporal Task is referred to as a Separate choice [Sc] model. The semi-nested models that we consider are thus Common-value Common-choice [CvCc], Common-value Separate-choice [CvSc], Separate-value Common-choice [SvCc], and Separate-value Separate-choice [SvSc] model.

Following closely the approach taken by Luckman et al. (2018), we select the most informative and best-fitting model via Bayes factors. Through our approach we identify the model that moderates the increased error variance from under- or over-identified models, accordingly. The model selection in

14R DP No. 205

our study is based on Bayes factors which penalise *extra* parameters. We calculate the marginal likelihood of each semi-nested model and then transform these marginal likelihoods into a model probability, assuming that each of the four candidate models has a prior probability of 0.25.

Model selection via Bayes factors requires setting up a prior distribution on the parameters of our models. Given that our study involves a replication of a previous experiment, we use the prior distribution for Luckman et al. (2018) which itself is informed by Luckman et al. (2015). Model selection via Bayes factors is appropriate in the current context because our models are not complex, i.e., they do not have many parameters that need to be estimated and compared. Hence our estimation procedure is not computationally expensive.

Note that in the Intertemporal Task we use the difference between the two value parameters α_I and $\alpha_R(\alpha_I - \alpha_R)$. Further, the marginal likelihood of each of the models described in the previous subsection is fitted to the data for every participant in the experiment by using the brute-force method (Vandekerckhove et al., 2015), as in Luckman et al. (2018). Concretely, we use *N* samples from the prior distributions of each parameter p(a) and p(s), and for the weighted set of parameters θ_i we compute the probability of the data $p(D|\theta_i)$ and we average over *N* samples. One limitation of this model is that *N* should be large enough because the marginal likelihood is only true if *N* converges to infinity. Otherwise, it would be inefficient and would not approximate the true state⁶ (Farrell & Lewandowsky, 2018). Therefore, Luckman et al. (2018) proposed to set N = 600.000. We adjust the estimate of the marginal likelihood to the posterior model probability, assuming 0.25 for each model. Strictly speaking, this transformation is calculated for each of the four *j* models described above by applying the following formula:

⁶ In that case, it is recommended to implement the importance sampling model (Farrell & Lewandowsky, 2018).

$$p(M_j|D) = \frac{m(D|M_j)}{\sum_k m(D|M_k)}$$
(12)

Given that we assign an equal prior probability to each of the four models, a model probability higher than 0.25 indicates that the data identifies that particular model as more parsimonious. We do this calculation for every individual that participated in our experiment, excluding those who chose dominated choices, in order to test the degree of support for each model described above.

Luckman et al. (2018) observed that the behaviour and choices of 20 out of 66 participants cannot be explained by the CvCc model (but for 46 participants it could be explained). The information we collected through the standardised version of the BNT test, and the Big-5 survey of financial literacy, was meant to help dig deeper into the reasons and unpack why participants in the experiment make choices that cannot be explained by the CvCc model or *vice versa*.

Results

Our analysis of the data points is conducted in RStudio (R Development Core Team, 2021) and Python. We exclude from the main analysis those participants in the experiment who fail on more than one of the attention-check questions.

Group Choice

This part of the analysis corresponds to the section *"Group choice"* from the original study (Luckman et al., 2018). (The analyses we conducted go beyond the original one. Specifically, we discuss the safer decisions in the Risky Task and sooner decisions in the Intertemporal Task in **Appendix K**.)

In line with the original study, we calculated the proportion of the participants who chose the safe option (LEFT option) for every choice in the Risky Task and the proportion of participants who chose the sooner option (LEFT option) for every choice in the Intertemporal Task. This replicates the group-level analysis in Luckman et al. (2018, p. 789) which is presented in **Figure 1** of their study.

and

Further, for the signifiers FLAT or RANDOM, the equations are the same, but instead of 360 participants, we have 180 participants. Yet again, for each of these 180 participants, we have 130 choices and thus 130 data points. For instance, for FLAT or RANDOM, we apply the following equations:

and

Lastly, for the signifiers ITF, RTF or M, the equations are the same, but we have 120 participants

and 130 choices and thus have 130 data points. Formally, for the RTF treatment (and likewise for the

treatments labelled ITF and M) we used the following equations:

and

We now move to the analysis of choices that were made by the participants in the Risky Task and in the Intertemporal Task, at the group level. **Figure 8** contains the boxplots of the Risky Task and the Intertemporal Task, respectively. These boxplots represent the distribution of the proportion of participants who have chosen the safer option (LEFT option) in the Risky Task for each of the 130 choices and the distribution of the proportion of participants who have chosen the sooner option (LEFT option) in the Intertemporal Task for each of the 130 choices. The long black line indicates the median value for each group, whereas the red circle indicates the mean values for each group. Recall that *prop.ALL* is based on the decisions made by 360 participants in 130 risky choices and 130 intertemporal choices.

Figure 8





Similarly, INCENTIVE SYSTEM contrasts the outcomes of the FLAT and RANDOM (performancebased) payment schemes whereas TREATMENT captures the results of RTF, ITF, and M treatments, i.e., the order in which we administered Risky and Intertemporal Task. (They are calculated in accordance with 14*a*, 14*b*, and 15*a* and 15*b*, respectively.) Boxplots in the **Online Materials** contain information about the proportions of participants that have chosen the safer option in the different conditions

14R DP No. 205

(FLAT, RANDOM, RTF, ITF, and M) and the proportion of participants that have chosen the sooner option in the different conditions (FLAT, RANDOM, RTF, ITF, and M), respectively.

We infer from **Figure 8** that the distributions of risky and intertemporal choices differ. We observe that the span of the most frequent concentration as well as the median line for the Intertemporal Task are substantially below the span of the most frequent concentration and the median line for the Risky Task. Further, we observe that both median and mean values differ considerably across both the Risky and the Intertemporal Task. We conclude that the distribution of the proportions of participants that have chosen the safer option differs from the proportion of participants that have chosen the various boxplots in the **Online Materials** illustrate that these differences are robust across the different conditions.

In the last section, we provide a wider theoretical discussion about the parameter of the utility curvature for the Risky Task and Intertemporal Task, which implies that the two tasks differ. As a consequence, our analysis is not as robust as would be desirable, but it is informative when the stimuli (and the corresponding decisions) are analysed, and when one recall that identification between the stimuli in the Risky Task and the Intertemporal Task is almost perfect.

We have also created scatterplots that mimic those in the original study (**Figure 1** from Luckman et al., 2018), **Figure 9a**, for the Risky Task, and **Figure 9b** for the Intertemporal Task.

The visual inspection of the group choices suggests that the intertemporal choices (almost all of which lie below 0.8), are noisier, and their distribution less concave than the risky choices. This indicates that risky choices are not the same as intertemporal choices and unlikely to be explainable by a single model on the aggregate level. Note that, both visually and from a qualitative perspective, they might look similar to the reader but there are distinct quantitative differences. Additional scatterplots are available in the **Online Materials.** They suggest that the normality assumption is violated.

The first two tables in the **Online Materials** contain information about the proportion of the participants that have chosen the safer option in the different conditions, and the proportion of the participants that have chosen the sooner option in the different conditions, respectively. These tables informed the other parts of our analysis.

Figure 9a and Figure 9b

Scatterplot of the Risky Task (proportion of safe choices) and of the Intertemporal Task (proportion of the sooner choices) at the group level



We now move to a formal statistical analysis of the data points from the experiment.

Firstly, we ran a Wilcoxon test (paired [Wilcoxon signed-rank test] and unpaired [Wilcoxon rank sum test or Mann-Whitney test]) as a non-parametric alternative to the t-test (t-test statistics are available in the **Supplemental Materials**) when the data are not normally distributed, as in our case. As a bonus, the Wilcoxon test does not lose relevance when extreme values (outliers) are present. The null hypothesis of the Wilcoxon test is that the median of the considered variable is the same for the two compared groups.

Feltovich (2003) tested a modification of the Wilcoxon-Mann-Whitney test, an allegedly more robust alternative proposed by Fligner-Policello, and concluded that it should be used instead of the Wilcoxon-Mann-Whitney test if the population dispersions are different, when the exact critical values for the robust rank-order test are available, and when the sample sizes are large (40+). The null hypothesis is the same as for the Wilcoxon-Mann-Whitney test. While the Fligner-Policello test does not assume that the shape of the distribution is similar for the two groups, it does assume that the

distributions of the variables or groups of interests are symmetric.

In **Table 1** we present the results that we obtained from the Wilcoxon (paired and unpaired) and

Fligner-Policello tests. (Nota bene, we show the results in the first part for the Fligner-Policello test in

order to be consistent in our presentation. This test is not specifically designed to compare central

tendences of paired data, but it is an alternative to the Wilcoxon-Mann-Whitney test i.e., it is adequate

for unpaired [independent] data. Therefore, these results should be treated with caution.)

Table 1

Wilcoxon, and Fligner-Pollicello, tests of risky choices and the intertemporal choices

	WILCOXON TEST	FLIGNER-POLICELLO TEST
ALL	8152***	-10.19888***
FLAT	8243.5***	-9.468542***
RANDOM	8280***	-10.77485***
RTF	8180***	-11.56998***
ITF	7504***	-6.51648***
Μ	8360.5***	-12.64584***
	RI	SKY TASK
INCENTIVES	8845.5	-0.6570591
RTF – ITF	8685	-0.391574
RTF – M	8891.5	-0.7408341
ITF – M	8648.5	-0.3311087
	INTERT	EMPORAL TASK
INCENTIVES	9622.5**	-1.941453**
RTF – ITF	6091***	3.923049***
RTF – M	9151	-1.149128
ITF – M	11165***	-4.521599***

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively.

Comparing **Table 1** and **Table K6** (from the **Appendix K**), we observe that the statistically significant findings in the first part of the tables, across the various treatments, and the incentive systems, are robust and identical. Moreover, there are no statistically significant differences between the results from the Wilcoxon test and those from the Fligner-Policello test.

We highlight the statistically significant differences in the median values from the proportion of the participants that have chosen the safer option (Risky Task) and the median values from the

14R DP No. 205

proportion of the participants that have chosen the sooner option (Intertemporal Task). This difference is robust across incentive systems and treatments for both tests.

In terms of the Risky Task (second part of **Table 1**), we see no statistically significant difference, which implies that the median values from the proportion of the participants that have chosen the safer option in the Risky Task is robust across the incentive systems and treatments. However, this is not the case for the Intertemporal Task (third part of **Table 1**) where statistically significant differences in the median values from the proportion of the participants that have chosen the sooner option in the incentive systems, ITF treatment and RTF treatment, and ITF treatment and M treatment are evident.

We also ran a Kolmogorov-Smirnov test which is used to test whether the compared groups come from the same distribution i.e., the Kolmogorov-Smirnov helps us to evaluate the equivalence in distribution between pairs of data. The null hypothesis of the Kolmogorov-Smirnov test states that the two groups of interest come from the same distribution. The results that we have obtained from the Kolmogorov-Smirnov test for different incentive systems and treatments are available in Table 2.

Table 2

Kolmogorov-Smirnov test o	f the risky choices an	nd the intertemporal choices
---------------------------	------------------------	------------------------------

	1251 51411511C5				
	Risky Task and Intertemporal Task	Risky Task	Intertemporal Task		
ALL	0.50769***	-	-		
FLAT	0.5***	-	-		
RANDOM	0.51538***	-	-		
RTF	0.55385***	-	-		
ITF	0.36923***	-	-		
Μ	0.58462***	-	-		
INCENTIVES	-	0.069231	0.14615		
RTF- ITF	-	0.069231	0.34615***		
RTF – M	-	0.092308	0.12308		
ITF – M	-	0.069231	0.4***		

TECT CTATICTICC

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively.

Table 2 reveals the substantially different distributions of the proportion of safe choices (Risky Task) and the proportion of sooner choices (Intertemporal Task). We find thus additional empirical evidence that the distribution of the risky choices is different from the distribution of the intertemporal
choices. Note that this result is robust across different incentive systems, and treatments. This finding is not a strong indication/proof because the Kolmogorov-Smirnov test is (typically) designed for treating independent samples, hence the results should be taken with caution when the paired data points are at stake. Regarding the tasks, the results for the Risky Task (second column in **Table 2**) indicate that there is no statistically significant difference between the distribution of the proportion of safe choices across the different incentive systems, and treatments. Contrary to this, for the Intertemporal Task (third column in **Table 2**), we find that the distribution of the proportion of sooner choices is different when ITF treatment and RTF treatment, and ITF treatment and M treatment, are compared.

We also did standard correlation, distance covariance, and distance correlation analyses. The

results are presented in Table 3.

Table 3

Correlation analysis at the aroun level					
correlation	Pearson	Kendall	Spearman	Distance covariance	Distance correlation
ΔΠ	0 7542031***	0 6301602***	0 8187256***	0 1272244	0 7588383
		0.0301002	0.0107230	0.1290777	0.7560505
FLAI	0.7503575	0.02/9/18	0.8101088	0.1280777	0.7551156
RANDOM	0.7465259***	0.6255179***	0.814139***	0.1256091	0.7559091
RTF	0.7148516***	0.5880296***	0.7827983***	0.1195607	0.7244985
ITF	0.7359897***	0.6197614***	0.8052184***	0.1366098	0.7270352
M	0.7240589***	0.6452266***	0.8246188***	0.1242993	0.7905518
	RISKY TASK				
INCENTIVES	0.98713***	0.9071996***	0.9856662***	0.1605975	0.9831464
RTF – ITF	0.9797498***	0.8733356***	0.9726318***	0.1630992	0.9764242
RTF – M	0.9184689***	0.817559***	0.9376408***	0.1550841	0.9311492
ITF – M	0.9117114***	0.8152373***	0.9297886***	0.1518415	0.9285239
	INTERTEMPORAL TASK				
INCENTIVES	0.9866689***	0.9080653***	0.9860413***	0.1700116	0.9853431
RTF – ITF	0.9650496***	0.8961066***	0.9810819***	0.1805017	0.9723378
RTF – M	0.952854***	0.8078461***	0.9469225***	0.1458684	0.9363104
ITF – M	0.9094369***	0.8068737***	0.9481342***	0.1654296	0.9156839

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively.

From **Table 3**, we see that there is a moderate to strong positive correlation between the

proportion of the safe choices for every choice (Risky Task) and the proportion of the sooner choices for

every choice (Intertemporal Task). This result is robust across the different correlation coefficients as

well as across the different treatments, and incentive systems. We also found a positive, almost perfect,

correlation between the proportion of the safe choices (Risky Task) in the pairwise comparison of the incentive systems, and treatments. A strong positive correlation is also found in the Intertemporal Task across the incentive systems, and treatments. Nevertheless, given the possibility of non-linear affine transformations of risky and intertemporal choices we must be cautious when drawing general conclusions from the correlation analysis.

Individual Choice

After the descriptive analysis and the analysis of the data at the group level, we move to the analysis at the individual level. See section **Estimation Strategy** for motivation and theory.

We exclude from this analysis the participants who failed to choose the dominant option on more than one of the five check questions in either choice sets. (See the **Supplemental Materials**).

In this last part of our analysis, we identify the best semi-nested model at the individual level by using Bayesian statistics. To define, "the Bayesian inference is reallocation of credibility over possibilities. In model comparison, the focal possibilities are the models, and Bayesian model comparison reallocates credibility across the models, given the data" (Kruschke, 2015, pp. 265-266). Bayesian model comparison/selection helps us to determine, based on the data that we have, which model performs better relative to others.

A key to Bayesian model comparison/selection is the Bayes factor. The popularity of the Bayes factor is perhaps due to the fact that it does not require parameter estimation. Rather it provides relative evidence in favour of the predictive performance of one model over another i.e., it tells us whether the observed data are (relatively) more probable under one over another model. It does not depend on a single value parameter but rather the entire prior distribution of the parameter and the actual data are considered in the unique framework simultaneously.

Formally, we write Bayes factors as: $BF_{12} = \frac{P(D|M_1)}{P(D|M_2)}$ or $BF_{12} = \frac{posterior_{12}}{prior_{12}}$. Note that Bayes

factors⁷ are related to the Akaike Information Criterion approach to approximately calculate the posterior probability. More precisely, "the Bayes factor is the factor by which a rational agent changes her prior odds in the light of observed data to arrive at the posterior odds. More intuitively, the Bayes factor quantifies the strength of evidence given by the data about the models of interest. It expresses this evidence in terms of the models' relative prior predictive accuracy" (Franke, 2021).

Given there is no substantial computational burden involved in estimating Bayes factors via brute-force (or grid search) in our case, and because it is the methodology that was applied in Luckman et al. (2018), we proceeded with this approach which allows us to systematically evaluate the semi-

⁷ Concretely, "the Bayes factors (and posterior model probabilities) tell us how much evidence the data (and priors) provide in favour of one model or another. That is, they allow us to perform inferences on the model space, i.e., to learn how much each hypothesis is consistent with the data" (Nicenboim et al., 2022). Analytically, we define the Bayes factor as the ratio of margi nal likelihoods — which is the likelihood of the observed data given the model specification — derived from Bayesian updating or as the degree of shift in prior beliefs about the relative credibility of the two models (Makowski et al., 2019). In contrast, frequentist statistics evaluates a model by computing the maximum likelihood, that is, the likelihood of the data given the b estfitting model parameter relative to the data points. Yet, it does not take into account the uncertainty associated with the estimates, and it could lead to a situation of Lindley's paradox. Bayesian inference attempts to quantify this uncertainty i. e., it helps to evaluate the marginal likelihood that entails the likelihood given all plausible values and thus the uncertainty for the model parameters. The marginal likelihood is not a function which means that it does not depend on the model parameters as they are *marginalised*. Concretely, the likelihood is evaluated for every possible parameter value, weighted by the prior plausibility and summed along. Hence, the prior distribution is inevitably vital for the procedure. Nevertheless, note that the marginal likelihood is based on the data observations (Nicenboim et al., 2022).

14R DP No. 205

nested models and to check which of them outperforms other models. An advantage of this approach, beyond its simplicity, is the possibility to visually represent the results⁸.

In order to identify the best model at the individual level, we computed the Bayes factor of each semi-nested model for each individual. In our case, with 4 semi-nested models and (planned) 360 participants, that requires a total of 1340 Bayes factors at the individual level (we have calculated the Bayes factors for 10 participants that failed on more than one checker too but excluded them from the subsequent analysis).

In our estimation we used the prior distribution from Luckman et al. (2018) which helps us to overcome issues that arise when uninformative priors would be attached. Moreover, we also applied alternative prior specifications that help us to avoid these issues, and that relate to the other comparisons in the **Supplementary Materials** of Luckman et al. (2018). We have provided an additional comparison, based on "CUSTOM" priors. Concretely, the prior distributions for the parameters of the models which were used in Luckman et al. (2018) in their **Comparison 1** (used in the main manuscript) included gamma distributions for the value parameters with relatively big scale and shape parameters. These gamma distributions produced samples of value parameters with a mean around 12. This is an enormous value for these parameters since they are used as exponents in the subjective value function which results in choice function probabilities of either 0 or 1 due to underflow or overflow of the exponent function in the logistic function. As explained by Luckman et al. (2018), the subjective value function is concave for the values in the parameter range [0, 1]. The sampling of the value parameter

⁸ Yet, Kangasrääsiö et al. (2019) noted that the method is less efficient and very expensive when the dimensionality of the space is high and when we need to infer multiple parameters simultaneously. Fortunately, we do not face those challenges and obstacles in our estimation.

using gamma distributions with mean bigger than 12 produces almost no samples in the considered range i.e., it results in a "typical" misleading bias. This is why we additionally employed the **"CUSTOM"** set of prior distributions which contains the same priors for all parameters except for the value, for which we have chosen a uniform distribution in the range of [0,1]. Finally, considering that "large sample studies, and/or replication studies can lead to reliable inferences from empirical data in the cognitive sciences" (Nicenboim, et al., 2022), we believe that the Bayesian inference in our case provides credible evidence.

Following the estimation strategy of Luckman et al. (2018), we have not estimated the best fitting parameters for each of the semi-nested models at the aggregate and at the individual level. Instead, at the individual level, we used the brute-force method. This methodology enables us to calculate the marginal likelihood for each model for each participant separately. This analysis relies heavily based on the information regarding the prior distributions available in the **Supplementary Materials** in Luckman et al. (2018) and the **"CUSTOM"** priors as explained above. The last analysis is done at the level of individual participants.

Following Luckman et al. (2018), the basic procedure/algorithm that we implement is: 1) choose a participant and a particular semi-nested model; 2) draw a set of parameter values for that model from its priors reported in the **Supplementary Materials** in Luckman et al. (2018)⁹; 3) calculate the likelihood

⁹ The prior distributions are obtained from the dataset with responses from a separate set of participants to a similar set of choices (Luckman et al., 2015). Luckman et al. (2015) fitted the semi-nested models to the dataset, and by using the Maximum Likelihood Estimation they estimated best parameters for each of the semi-nested models, for each participant in their experiment separately. Then, Luckman et al. (2015) looked at the distribution of values across participants for each parameter r. Ultimately, they removed the top and bottom most extreme two percent of the of values, and fitted a distribution to these (uniform, lognormal, normal etc.), depending upon restrictions on the parameters and the distribution.

of the observed data for that participant and for the selected semi-nested model using the sampled set of parameter values; 4) repeat 600.000 times, i.e., randomly draw 600.000 sets of parameters, and calculate the likelihood of that participant's data for each set, to calculate the marginal likelihood; 4) repeat this procedure for each of the other semi-nested models for the participants; and 5) repeat this procedure for the other participants from the experiment. See more details in **Appendix K**.

We replicate **Figure 2** from Luckman et al. (2018) [**Comparison 1**]¹⁰. Concretely, we produce a violin plot of the posterior model probabilities of the four semi-nested models for each individual in the experiment, excluding those who failed on more than one dominated choice. The results are available in **Figures 10a-10g**, where **Figure 10a** corresponds to **Figure 2** from Luckman et al. (2018). Note that we present the results in dollar amounts, as we did not find substantial differences between the dollar amounts and talers.

An important note, the white dot in **Figures 10a-10g** represents the median, whereas the bold line shows the span of the most frequent distribution of the model probabilities.

¹⁰ We replicated other figures reported in the **Supplementary Materials** of Luckman et al. (2018) in our **Online Materials**, as **Comparison 2**, **Comparison 3**, **Comparison 4**, as well as more added and created figures for the **Comparison CUSTOM**. We found that the results in **Comparison 2** (152 participants, 272 participants, 162 participants and 159 participants), **Comparison 3** (341 participants, 52 participants, 229 participants, and 64 participants), **Comparison 4** (344 participants, 101 participants, 289 participants, and 0 participants) and **Comparison CUSTOM** (350 participants, 7 participants, 209 participants, and 0 participants) differed in the number of participants where the model probability is greater than 0.25 for the four semi-nested models, accordingly: CvCc, CvSc, SvCc, and SvSc.

Figure 10a

Model probabilities for ALL (350 participants)











From **Figure 10a**, which corresponds to **Figure 2** reported in Luckman et al. (2018), we learn that the CvSc model has the largest model probability for the majority of the participants in the experiment. Our result differs from Luckman et al. (2018) who found evidence in favour of the CvCc model for 46 out of 66 participants, i.e., for about 7 out of 10 participants. Instead, we found evidence in support of the CvSc model for 317 out of 350 participants, i.e., for about 9 out of 10 participants. Thus, we identify CvSc as the best model to accommodate risky and intertemporal choices.

An important difference is that the model probabilities in our study span from 0.20 to 0.30, which is different from Luckman et al. (2018) where they span from 0 to almost 1. This differs across the different treatments, as one can see it in the **Online Materials**. While the span of the model probability is slightly different for our various comparisons, overall, it does not differ much. It just moves around substantially the number of participants for which the model probability for the different semi-nested model is larger than 0.25.

We do not have a good and reasonable explanation for the (substantially) different ranges of model probabilities in our study and that of Luckman et al. (2018). An obvious guess is — priors. Namely, the parameters of the same underlying construct for which improper priors are imposed are (usually) identical across models. The original priors (Luckman et al., 2018) are obtained from a 2015 experiment where the participants were "paid" with credit points. It could be that these priors are not the most adequate ones for our conceptual replication (but, since no better candidate was suggesting itself, and because we did a conceptual replication, we relied on them).

Our result seems to be robust across the incentive systems (FLAT and RANDOM) and order of presentation of the tasks (RTF, ITF, and M). That said, from the visual inspection of **Figures 10a-10g**, we observe that for the other semi-nested models, there are some differences across incentive systems and orders of presentation. Therefore, we provide descriptive statistics in **Tables 4a-4c** that facilitate understanding of the model probabilities of the semi-nested models.

Descriptive statistics for model probabilities for ALL (350 participants)						
MODELS	MODELS MODEL PROBABILITY					
(LARGER THAN 0.25)						
CvCc	92 (26.29%)	46 (13.14%)				
CvSc	317 (90.57%)	286 (81.71%)				
SvCc	138 (39.43%)	17 (4.86%)				
SvSc	167 (47.71%)	1 (0.29%)				

Table 4a

Table 4b

Descriptive statistics for model probabilities for FLAT (left) and RANDOM (right) treatments

MODELS	FLAT		RANDOM		
[MODEL PROBABILITY	LARGEST MODEL	MODEL PROBABILITY	LARGEST MODEL	
	(LARGER THAN 0.25)	PROBABILITY	(LARGER THAN 0.25)	PROBABILITY	
CvCc	38 (21.71%)	19 (10.86%)	54 (30.86%)	27 (15.43%)	
CvSc	161 (92%)	148 (84.57%)	156 (89.14%)	138 (78.86%)	
SvCc	60 (34.29%)	8 (4.57%)	78 (44.57%)	9 (5.14%)	
SvSc	93 (53.14%)	0	74 (42.29%)	1 (0.57%)	

Table 4c

Descriptive statistics for model probabilities for RTF (left), ITF (center), and M (right) treatments

MODELS	RTF		ITF		М	
	MODEL	LARGEST MODEL	MODEL	LARGEST	MODEL	LARGEST MODEL
	PROBABILITY	PROBABILITY	PROBABILITY	MODEL	PROBABILITY	PROBABILITY
	(LARGER THAN		(LARGER	PROBABILITY	(LARGER THAN	
	0.25)		THAN 0.25)		0.25)	
CvCc	35 (29.91%)	19 (16.24%)	15 (12.50%)	7 (5.83%)	42 (37.17)	20 (17.70%)
CvSc	103 (88.03%)	92 (78.63%)	113 (94.17%)	109 (90.83%)	101 (89.38%)	85 (75.22%)
SvCc	52 (44.44%)	6 (5.13%)	28 (23.33%)	3 (2.5%)	58 (51.33%)	8 (7.08%)
SvSc	53 (45.30%)	0	77 (64.17%)	1 (0.83%	37 (32.74%)	0

Table 4a confirms our conclusion that the CvSc model accommodates the decisions of the

overwhelming majority of the participants in the experiment (317 out of 350 participants). The secondmost preferred semi-nested model, the SvSc model, is found to have a probability greater than 0.25 for 167 out of 350 participants. The SvCc semi-nested model has a model probability greater than 0.25 for 138 out of 350 participants, and the CvCc model for 92 out of 350 participants. We conclude that the CvCc model - contrary to the findings in Luckman et al. (2018) -- is not the best model to account for risky and intertemporal choices in a unified framework. Indeed, when we rank the model probabilities for each of the 350 participants, we find that the model probability of the CvSc model is largest for 286

out of 350 participants, with 46 participants for the CvCc model, 17 participants for the SvCc model, and 1 out of 350 participants for the SvSc model.

As regards incentive systems (**Table 4b**), the performance of the CvSc model is consistent across them. We do find differences in proportion for the other semi-nested models. The number of participants from the FLAT treatment where the model probability is greater than 0.25 is different from the number of participants from the RANDOM treatment where the model probability is greater than 0.25. This difference is larger than 10 participants for the CvCc model (16 participants), the SvCc model (18 participants), and the SvCc model (19 participants). Regarding the number of participants and seminested models with the largest model probability, we observe a difference of 10 participants for the CvSc model (148 participants in the FLAT treatment and 138 participants in the RANDOM treatment.) We do not observe a difference larger than 10 participants for the other semi-nested models.

In **Table 4c**, we have excluded 10 participants who violated the dominance principle, which leaves us with 117 participants in the RTF treatment, 120 participants in the ITF treatment, and 113 participants in the M treatment. Yet again we find that the CvSc model has the dominant model probability of greater than 0.25 for the RTF (88.03%), ITF (94.17%), and M (89.38%) treatments. This is more or less consistent across different orders of presentation. Yet, that is not the case for the other semi-nested models. For instance, the CvCc model is found to have a probability greater than 0.25 for RTF (29.91%), ITF (12.50%), and M (37.17%) treatments. Further, the SvCc model is found to have a probability greater than 0.25 for RTF (44.44%), ITF (23.33%), and M (51.33%) treatments. The SvSc model is found to have a probability greater than 0.25 for RTF (45.30%), ITF (64.17%), and M (32.74%) treatments. Finally, regarding the largest model probability, for the four respective seminested models in **Table 4d**, we find that the percentage of the participants for the RTF treatment are 16.24%, 78.63%, 5.13%, and 0%, percentage for the ITF treatment are 5.83%, 90.83%, 2.5%, and

14R DP No. 205

0.83%, and percentage for the M treatment are 17.70%, 75.22%, 7.08%, and 0%. The bottom line from this simple (descriptive) analysis is that order effects matter.

We now move to the formal analysis of the individual choices in the experiment, which goes beyond Luckman et al. (2018). We ran three logit regressions where the binary outcome variable is whether the CvCc model has a model probability greater than 0.25 (the outcome variable is 1) or otherwise (the outcome variable is 0). The first logit regression ("DETERMINANTS") contains the predictors form the surveys in the experiment. The second logit regression ("EXPERIMENT") contains the predictors relevant for the experiment, and the third logit regression "TASK ATTRIBUTES" contains the predictors from the task features and attributes. **Table 5** shows the results. We note that we controlled multicollinearity by applying the Variance Inflation Factor (VIF) and we found that there are no predictors that potentially undermine the results from our logit regressions. As an important caveat, recall that large standard errors affect statistical significance, but that is not the case for all of the coefficients in the subsequent analyses.

From the first column in **Table 5**, which corresponds to the first logit regression, we found that only the BNT score (risk numeracy) has a statistically significant effect on the binary outcome variable (at the 10% level of statistical significance). In other words, for every one-unit change in BNT score (risk numeracy), the log odds that the CvCc has a model probability larger than 0.25 increases by 0.176339. Other predictors in the first logit regression are not statistically significant.

Regarding the second logit regression, which corresponds to the predictors related to the experiment (second column in **Table 5**), we find that when the incentive system is RANDOM, instead of FLAT, the log odds that the CvCc has a model probability larger than 0.25 increases by 0.4683. This result is statistically significant at the 10% level. For the other predictors we fail to find a statistically significant relationship with the binary outcome variable.

Table 5

Logit model of the CvCc semi-nested model

CVCC (BINARY)	DETERMINANTS	EXPERIMENT	TASK ATTRIBUTES
AGE	0.039160		
	(0.110599)		
GENDER	-0.009463		
	(0.245911)		
EDUCATION	-0.038118		
	(0.187686)		
INCOME	-0.013847		
	(0.090691)		
JOB	0.020882		
	(0.105954)		
BNT SCORE	0.176339*		
	(0.095291)		
HILDA SCORE	0.256938		
	(0.168186)		
POOL		0.1120	
		(0.2450)	
INCENTIVES		0.4683*	
		(0.2464)	
ORDER		0.1784	
		(0.1513)	
COMPREHENSION			0.367549**
(INTERTEMPORAL TASK)			(0.174703)
COMPREHENSION			0.508299***
(RISKY TASK)			(0.181268)
RESPONSE TIME			-0.019504**
			(0.009685)
CONFIDENCE LEVEL			0.221741*
(INTERTEMPORAL TASK)			(0.134606)
CONFIDENCE LEVEL			0.038405
(RISKY TASK)			(0.141165)
EFFORT LEVEL			-0.573859***
(INTERTEMPORAL TASK)			(0.118825)
EFFORT LEVEL			0.380083***
(RISKY TASK)			(0.128887)
			0.082968
(INTERTEMPORAL TASK)			(0.116824)
			-0.026875
(RISKY TASK)			(0.116156)

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively. Standard errors are reported in brackets.

Lastly, the third column in **Table 5** corresponds to the logit regression where the predictors are specific task attributes and features. We find that the task understanding, which is particularly important for the choice function (n.b., relative performance of the model is not fully dependent on the task scores, but understanding the task, which is one potential attribute, could potentially direct the drive wheel), indicated by the comprehension score, has a statistically significant effect on the binary outcome variable for both tasks at the statistical significance level of 1%. In particular, the log odds that the CvCc has a model probability larger than 0.25 increases by 0.367549 and 0.508299 for every one unit increase in the comprehension score for the Intertemporal Task and the Risky Task, respectively.

Further, from the third column in **Table 5**, we found that the log odds that the CvCc has a model probability larger than 0.25 decreases by 0.019504 for each one instantaneous increase in the response time (statistically significant at 5%). The reported confidence level is found to be statistically significant only for the Intertemporal Task (statistically significant at 10%). In other words, the log odds that the CvCc has a model probability larger than 0.25 increases by 0.221741 for every one unit increase in the confidence level for the Intertemporal Task. Interestingly, we found opposite effects for the reported effort levels in the Intertemporal Task and Risky Task. More precisely, the log odds that the CvCc has a model probability larger than 0.25 decreases by 0.573859 for every one unit increase in the stated effort level for the Intertemporal Task. Alas, log odds that the CvCc has a model probability larger than 0.25 decreases in the stated effort level for the Intertemporal Task. Alas, log odds that the CvCc has a model probability larger than 0.25 decreases in the stated effort level for the Risky Task. The results are statistically significant at 1%.

Next, we move to the analysis of the marginal effects (**Table 6**.). The marginal effects show the direction and magnitude of the change in probability when the predictor increases by one unit. The results from the three respective logit regression models that we discussed are available in **Table 6**. (In **Appendix M** are reported available relative risk [odds] ratios from this analysis.)

Table 6

Marginal effects of the CvCc semi-nested model

CVCC (BINARY)	DETERMINANTS	EXPERIMENT	TASK ATTRIBUTES
AGE	1.039937		
	(0.115016)		
GENDER	0.990581		
55.101.710.1	(0.243594)		
EDUCATION	0.962599		
	(0.180666)		
INCOME	0.986248		
	(0.089443)		
JOB	1.021101		
	(0.10819)		
BNT SCORE	1.192842*		
	(0.113667)		
HILDA SCORE	1.292965		
POOL	(0.217458)	1 11055	
POOL		1.11855	
		(0.27402)	
INCENTIVES		1.59750*	
OPDER		1 10525	
ORDER		1.19555	
		(0.18091)	1 ///100/**
			(0.2523043)
COMPREHENSION			1 667462***
(RISKY TASK)			(0 3013518)
RESPONSE TIME			0.9806847**
			(0.0094978)
CONFIDENCE LEVEL			1.2482479*
(INTERTEMPORAL TASK)			(0.1680211)
CONFIDENCE LEVEL			1.0391517
(RISKY TASK)			(0.1466914)
EFFORT LEVEL			0.5633473***
(INTERTEMPORAL TASK)			(0.06694)
EFFORT LEVEL			1.4624056***
(RISKY TASK)			(0.1884849)
COMPLEXITY LEVEL			1.0865066
(INTERTEMPORAL TASK)			(0.12693)
COMPLEXITY LEVEL			0.9734825
(RISKY TASK)			(0.1130761)

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively. Standard errors are reported in brackets.

From the first logit regression from **Table 6**, again, the only statistically significant predictor is the BNT score (statistically significance at 10%). Yet, the interpretation is different now. A one unit increase in the BNT score is associated with a 1.192842 increase in the probability of a CvCc model larger than 0.25, *ceteris paribus*. The second logit regression from **Table 6** indicates that when the incentive system is RANDOM, it is associated with a 1.59730 increase in the probability of a CvCc model larger than 0.25 (statistical significance of 10%), *ceteris paribus*.

In the last part of this analysis at the individual choice level, we implement a standard ordered logit model (**Table 7**). An important note, instead of a variable with binary outcome, in the standard logit model we have a dependent variable that is categorical. More precisely, when the model probability for the CvCc outperforms other semi-nested models — it is the largest model probability — we have that dependent categorical variable is 1. By the same token, the dependent categorical variable is 2 when the model probability of the CvSc model is largest, 3 when the model probability of the SvCc model is largest, and 4 when the model probability of the SvSc model is largest.

Interestingly, from the first standard ordered logit model, reported in first column in **Table 7**, we did not find any statistically significant result. The same holds for the second column in **Table 7**. Alas, we found that four predictors in the third standard ordered logit model (third column in **Table 7**) are statistically significant: comprehension score for the Intertemporal Task (5%), comprehension score for the Risky Task (5%), confidence level for the Risky Task (10%), and effort level for the Intertemporal Task (5%). So, for the comprehension score from the Intertemporal Task, we would say that for a one unit increase in the comprehension score, we expect a 0.4175823 decrease in the logarithm odds that CvSc, SvCc or SvSc models have the largest model probability, *ceteris paribus*. The result is identical for the comprehension score from the Risky Task. Yet, the magnitude is to some extent smaller in the confidence level from the Risky Task, but it is opposite in the reported effort level for the Intertemporal Task.

Table 7

Standard ordered logit model for the semi-nested models

MODEL (CATEGORICAL)	DETERMINANTS	EXPERIMENT	TASK ATTRIBUTES
AGE	0.012654		
	(0.124911)		
GENDER	0.15461		
	(0.279567)		
EDUCATION	-0.295182		
	(0.211701)		
INCOME	0.017595		
	(0.101131)		
JOB	-0.094551		
	(0.119144)		
BNT SCORE	-0.165319		
	(0.108285)		
HILDA SCORE	0.199661		
	(0.161571)		
POOL		-0.336210	
		(0.278429)	
INCENTIVES		-0.213279	
		(0.276700)	
ORDER		0.010873	
		(0.171515)	
COMPREHENSION			-0.4175823**
(INTERTEMPORAL TASK)			(0.1777004)
COMPREHENSION			-0.4462661**
(RISKY TASK)			(0.1894606)
RESPONSE TIME			0.0070909
			(0.009548)
CONFIDENCE LEVEL			-0.1512953
(INTERTEMPORAL TASK)			(0.1394044)
CONFIDENCE LEVEL			-0.2743219*
(RISKY TASK)			(0.149921)
EFFORT LEVEL			0.2721057**
(INTERTEMPORAL TASK)			(0.1206402)
EFFORT LEVEL			-0.1884949
(RISKY TASK)			(0.1292943)
			-0.0135722
(INTERTEMPORAL TASK)			(0.1246107)
COMPLEXITY LEVEL			-0.0459158
(RISKY TASK)			(0.122188)

Note. ***, **, * indicates a statistical significance at 1%, 5%, and 10%, respectively. Standard errors are reported in brackets.

It is not straightforward to interpret the coefficients from the standard ordered logit model. Therefore, we have calculated the marginal effects for these models. The results are reported in the **Online Materials**.

Discussion and Concluding Remarks

We present a brief review of related studies. We also summarise the key results of the experiment.

A Brief Review of Related Literature

We discussed some relevant literature in the first two sections when motivating our experiment. Here we complete this review.

The need for replication studies has increasingly been acknowledged due to the part they play in restoring confidence in the scientific enterprise (Anderson & Maxwell 2016; see also Ioannidis et al., 2017; Ortmann 2021; Brodeur et al., 2024). The focus of our replication study is on risk [risky choices] and time preferences [intertemporal choices] and their interaction. Anderhub and Güth (2001) reported that risk-averse participants discount the future more heavily than participants classified as risk-neutral and risk-seeking. Andersen et al. (2008) found that the joint elicitation of risk and time preferences results in significantly lower discount rates than separate elicitation. Somasundaram and Eli (2022) found similar results. Cheung (2020) suggests that the utility function elicited directly from choices over time is concave, but far closer to linear than the utility function that is elicited under risk, whereas Abdellaoui et al. (2013) found that utility functions tend to be concave for risky choices and linear for intertemporal choices. Andreoni and Sprenger (2012) identified a significant difference between risk and time preferences within the monetary and environmental domains.

In the context of risky intertemporal choices, Ahlbrecht and Weber (1997) implemented both matching task and choice task and tested the short-term/long-term asymmetry for certain and risky future outcomes for both gain and losses, the certainty/risk asymmetry for short-term and long-term decisions and for both gain and losses, the gain/loss asymmetry for certain/risky future outcomes, both for short-term and long-term decisions. When the matching task was implemented, they found evidence in favour of the short-term asymmetry which states that individual discount rates are higher for the short-term than the long term both for the case of certainty and the case of risk. They did not find empirical evidence that supports the gain/loss asymmetry, but they found that risky choices are discounted less than equally attractive certain choices in the matching task. Nevertheless, asymmetries were not observed once the choice procedure was established. Although these results are of interest for the literature on risky intertemporal choices, they are based on participants' response to hypothetical outcomes.

Similarly, Sun and Li (2010) presented experimental evidence that rejects the "reductionist assumption" underlying the Luckman et al. (2018) study when risk is induced in intertemporal choices through the standard choice titration procedure. Specifically, in two experiments they found that risk and time delay had opposite effects on the decision-making in an intertemporal context, i.e., the level of risk increased the degree of discounting and resulted in a lower discount factor, while the opposite effect was true for time delay. Yet, their study also involves hypothetical outcomes, and the authors stressed that it is important to test the effects of real incentives in this setting. Using real payoffs, Anderson and Stafford (2009) posited a relationship between risky choices and intertemporal choices which suggests that risk-taking behaviour is conditional on time. While only a small number of participants were found to exhibit consistent risky decision making, imposing risk on intertemporal choices seems to result in decreased patience levels regardless of the delay length. Implementing multiple price lists, Holt and Laury (2002) and Noussair and Wu (2006) found that more than 38% of

their participants tend to be risk-loving for choices that offer amounts that are realised in the distant future. In other words, they found that the level of risk aversion falls as delays of the options increase.

Önçüler and Onay (2009), conducting three experiments with which they tested path dependency in decision processes for risky intertemporal settings, found that the Discounted Expected Utility model is not the most appropriate specification because it assumes independent time and risk preferences. However, they suggested that, in the setting of the risky intertemporal choices, there is a two-stage decision process where the participants first discount the future by calculating the netpresent value and then they evaluate the certainty equivalent. Moreover, Ferecatu and Önçüler (2016) reported a negative correlation between risk and time preferences.

Cohen et al. (2011) conducted three experiments that involved students, participants from the general population, and portfolio managers and tested the relationship between risk, time/patience, and ambiguity/imprecision. Based on the results of their experiments, they stress that it is important to consider at least three different parameters for risk preferences, time preferences, and ambiguity preferences since they find no correlation between them in their experiment. They suggest that decision-making in these three areas differs and thus cannot be evaluated through a unifying/reductionist concept.

Reeck et al. (2017) elaborated that there is a comparative or integrative search in the decision processes that involve intertemporal choices. Therefore, the examination of the search processes is an important research avenue. Besides the fact that there are related studies in the literature of risky intertemporal choices (Wall et al., 2018)¹¹ and risky choices (Hey & Orme, 1994) in which also many choices are presented to the participants, it would be interesting to cross-validate the results by using

¹¹ Our access to these stimuli on the OSF was approved by Daniel G. Wall on 04-Sep-2021.

14R DP No. 205

the choices/stimuli based on price lists, convex budgets, or Bayesian adaptive design. The context for elicitation of risk preferences and time preferences in a unique framework is of importance as suggested by the findings in Ahlbrecht and Weber (1997).

Luckman and his colleagues wrote a series of related papers in order to address the limitations of Luckman et al. (2018). Specifically, drawing heavily on their design and implementation choices in in Luckman et al. (2018), they evaluated different models of risky intertemporal choices (Luckman et al., 2020) and in another project, conducted around the same time as Luckman et al. (2018), they included a titration procedure, testing of the magnitude and peanut effects, and "Risk vs Delay" choices (Luckman et al., 2017). Additionally, Konstantinidis et al. (2020) as well as Fitch and Kvam (2024) analysed cognitive process models and heuristics in decision making in risky choices and the intertemporal choices. The research of decision heuristics is important in the context of risky intertemporal choices because experimental participants make decisions in experiments that test the reductionist assumption through a large number of choices [e.g., 130 risky choices and 130 intertemporal choices]. Our caveats regarding design and implementation choices and the proper powering up of studies pertain with might to Luckman et al. (2020).

In the existing setting of pure risky choices and pure intertemporal choices, we removed the delays in Risky Choices and added temporal trade-offs in Intertemporal Choices. Hence, it would be interesting to make contextual changes and conduct an analysis when these effects are separated and observed in isolation. Relatedly, Chung et al. (2019) analysed these effects in the context or risk preferences. In other words, they observed stark differences between the riskless utility that is estimated for choices over bundles, and the risky utility.

An important issue in the relation between risky and intertemporal choices is the utility curvature. The identification of the curvature of utility functions in time-preference elicitation has gained only recently an attention in the literature relative to the long-standing interest for identification

I4R DP No. 205

of the curvature of utility functions for risk-preference elicitation. The curvature of the utility function is a concern because in the literature of risky intertemporal choices it is unclear how the utility function is identified given that the effect is not isolated. When the utility function is concave in the Intertemporal Task, it behaviourally indicates that there is a preference to smooth payoffs over time, and in the Risky Task it means that there is a preference for equal rather than unequal payoffs across different states of the nature and implies risk aversion. When doing these kinds of experiments, it is hence necessary to compare the utility (curvature) elicited under risk to the utility elicited over time, without interacting with the two of them. When there is an interaction, we have to deal with correlation aversion, an effect that is particularly relevant for the mixture setting of "Risk vs Delay" choices (Cheung, 2016). Luckman et al. (2018) found, using Maximum Likelihood Estimation, that the utility function for intertemporal choices was closer to linear than the utility function for risky choices, yet concave.

From a theoretical and empirical perspective, we conclude that the evidence provided by the existing theoretical and experimental studies regarding the relation of the risk preferences and time preferences is contradictory. Theoretically, it can be argued that the risk and time preferences have a one-to-one relationship within the discounted-utility model, given that the intertemporal elasticity of substitution can be used (though imperfectly) to define the inverse curvature of the utility function. Another view holds that the risk and time preferences are ingredients of a multi-attribute utility function, which gives rise for the careful analysis of the correlation aversion.

There are limitations of our study as well as of the original study (Luckman et al., 2018). First, the number of choice decisions (which is, as you will recall, large) might induce heuristics. Second, using the power value function, and constraining the parameter of interest, imposes limitations on the results. Other functional forms, other decision theories, and other choice models are bound to lead to different results and are worthy of investigation. As pertaining to our conceptual replication, we tried to stay as close as possible to the original. For instance, the attribute-wise model of Dai and Busemeyer (2014) or

14R DP No. 205

some other models of time preferences (see Frederick et al., 2002) and risk preferences (see Wu et al., 2004) which are also discussed by Luckman et al. (2018) could potentially better, and in a more adequate way, explain some important attributes and phenomena such as magnitude effects and peanut effects. In other words, alternative characterisation of value function and choice function, as well as applying other decision theories, will strengthen the analysis. Although we examined the evolution of the decision-making process across the choices, it still remains an open question to determine the decision patterns and cues. Namely, there are differences between the first and second half of the choices both within the same task and when both tasks are compared. Importantly, we did not make a correction for multiple testing, because the tests in the main analysis are powered enough according to our calculation attached to the replication of the main findings from Luckman et al. (2018). Yet, we strongly encourage those who will use the additional analysis from the **Supplemental Materials** and **Online Materials**, to correct for multiple testing. Lastly, a direct estimation of the parameters can explain (potentially) why the choice parameter is different in our study for the risky choices and intertemporal choices. Yet, instead of using the traditional Maximum Likelihood Estimation approach, we believe that Hierarchical Model Estimation is more appropriate in this setting. Furthermore, extracting different Bayesian prior distribution that could potentially explain the sensitivity of our results vis-à-vis results from Luckman et al. (2018) is another avenue of interest. Nevertheless, one of the main reasons why we did not extend our analysis along the lines just discussed is that the purpose of our study was replication of the main findings presented in Luckman et al. (2018). Therefore, we followed the approach and structure of that paper. Doing things different from the models or estimation approach that Luckman et al. (2018) implemented would move us even further from the (conceptual) replication that we set out to do.

Summary of the Results and Outlook

We have studied decision-making in related Risky and Intertemporal Tasks and tested the reductionist assumption which suggests that risky choices and intertemporal choices are special cases of risky intertemporal choices, as reported by Luckman et al. (2018) under the conditions of their experiment.

In the **Introduction** we identified our three research questions: 1) can we replicate the results in Luckman et al. (2018) in a non-incentivised laboratory setting; 2) do monetary incentives shift risky and intertemporal choices; and 3) can a single model account for both risky and intertemporal decisions once choices are properly incentivised?

Overall, we believe to be on safe grounds saying that the answers to 1) and 3) are negative, and that there is considerable evidence that the answer to 2) is broadly positive, although some such assessment warrants differentiated argumentation.

At the group level, we do not find evidence that pure risky choices are the same as pure intertemporal choices. While we have not observed differences in decision-making in the Risky Task between participants paid under the flat incentive scheme and those paid under the random incentive scheme, we observe some such differences in the Intertemporal Task. The proportion of the participants that have chosen the sooner option (for every choice) is larger in the FLAT condition than the RANDOM condition. Therefore, regarding our second research question, monetary incentives are important for the Intertemporal Task (but not for the Risky Task). This contradicts our priors, and we conjecture that stakes have to be considerably higher for their importance to show in the Risky Task (e.g., Holt & Laury 2002; Harrison et al., 2005a). Regarding the first research question, while we were able to replicate the results from the Risky Task, we failed to do so for the Intertemporal Task.

We have also performed a Kolmogorov-Smirnov test of the choices at the group level. From this, we found substantially different distributions of the proportion of the safe choices for every choice

(Risky Task) and the proportion of the sooner choices made for every choice (Intertemporal Task). In other words, we found additional empirical evidence — albeit not produced by a strong test — that the distribution of risky choices differs from the distribution of intertemporal choices.

At the group level, we have detected statistically significant differences in the median values of the proportion of the participants that have chosen the safer option for every choice in the Risky Task and the median values from the proportion of the participants that have chosen the sooner option for every choice in the Intertemporal Task. This difference is found to be robust across the incentive systems and treatments.

At the individual level, we failed to find evidence in favour of the CvCc model – the overwhelming model of choice by about two thirds of the participants in Luckman et al. (2018). Instead, we found that the decision-making for the majority of the participants in our experiment can be accommodated by the CvSc model. Based on the existing literature, we expected to find evidence that the most adequate model for the majority of the participants would be the SvSc semi-nested model (a pre-registered hypothesis). Our result is robust across the different treatments. In other words, we find that a common value function accommodates both risky and intertemporal choices, but the choice function does not.

Our result is sensitive to the prior distribution (see the **Online Materials**). Further, we found some evidence that monetary incentives (and order effects) matter for the other semi-nested models, particularly for the CvCc model. A case in point: when the participants are remunerated under the random incentive system, it is 0.09 times more likely that the CvCc model to has the model probability larger than 0.25 (as reported in **Appendix M**).

In addition to the analysis that is informed by Luckman et al. (2018), we also analysed in more detail the safer options (LEFT option) for the Risky Task, and the sooner options (LEFT option) for the

Intertemporal Task¹² as well as the overlap of the preferred options in both tasks. Specifically, we calculated — at the individual/participant level — the proportion of safer decisions (LEFT option) for the Risky Task out of 130 choices, and the proportion of sooner decisions (LEFT option) for the Intertemporal Task out of 130 choices. The main takeaway from our analysis of left choices is in line with the results reported in the body of the text: decision-making in the Risky Task does not correspond to decision-making in the Intertemporal Task, under the conditions of our experiment. In addition to this, we found statistical evidence that order effects matter for the Intertemporal Task. These results strengthen our previous conclusion that the Risky Task and Intertemporal Task cannot be accommodated by the same model, and that order effects have an important effect for decision-making with delayed option. Details of this analysis may be found in **Appendix K**.

Finally, to stress again, as with other research avenues, research of the same or different models and the same tasks in different domains of decision-making such as environment, transportation, health and so on seems desirable. (Again, context modification of the tasks, i.e., using a graphical presentation, allowing losses (Shelley, 1994), or using some other standardised instruments like price lists, adaptive methods, or convex budgets, might be interesting.) Conducting a re-tests session in a multi-period setting can be used as an indicator of the temporal stability of the conclusion and can further confirm the reliability of the results. Also, given that the elicitation mechanisms from the laboratory do not accurately explain field behaviour (Charness et al., 2020), this is an important limitation of our study that should be addressed in future projects. Going beyond the estimation strategy in Luckman et al (2018), in order to select among the models at the group levels, it is also relevant to apply two metrics, the group Bayes factor and the pooled Bayes factor (Regenwetter et al., 2018). Other characterisations of the

¹² We thank Sandro Ambuehl for suggesting this analysis.

value function and choice function as well as applying other decision theories, particularly the exponential model, will be of particular relevance for related studies. An important open issue that we address in a follow-up project is the estimation of the risk preference parameter and time preference parameter, as well as the respective choice parameters, through Maximum Likelihood Estimation and Hierarchical model. Testing of these tasks in the artefactual experimental setting could additionally strengthen the conclusions drawn from the experiment because decision-making in the time preference could be different when the general population is tested (Sun & Li, 2010; Cohen et al., 2011; Charness et al., 2020).

References

- Abdellaoui, M., Bleichrodt, H., l'Haridon, O., & Paraschiv, C. (2013). Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time. *Management Science*, *59*(9), 2153–2169. <u>https://doi.org/10.1287/mnsc.1120.1690</u>
- Ahlbrecht, M., & Weber, M. (1997). An empirical study on intertemporal decision making under risk. *Management Science*, *43*(6), 813–826. https://doi.org/10.1287/mnsc.43.6.813
- Anderhub, V., & Güth, W. (2001). On the interaction of risk and time preferences: An experimental study. *German Economic Review*, 2(3), 239–253. <u>https://doi.org/10.1111/1468-0475.00036</u>
- Andersen, S., Di Girolamo, A., Harrison, G. W., & Lau, M. I. (2014). Risk and time preferences of entrepreneurs: Evidence from a Danish field experiment. *Theory and Decision*, *77*(October), 341–357. <u>https://doi.org/10.1007/s11238-014-9446-z</u>
- Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3), 583–618. <u>https://doi.org/10.1111/j.1468-0262.2008.00848.x</u>
- Anderson, L. R., & Stafford, S. L. (2009). Individual decision-making experiments with risk and intertemporal choice. *Journal of Risk and Uncertainty*, *38*(February), 51–72.

https://doi.org/10.1007/s11166-008-9059-4

Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, *21*(1), 1–12.

https://doi.org/10.1037/met0000051

- Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *American Economic Review*, *102*(7), 3357–3376. <u>https://doi.org/10.1257/aer.102.7.3357</u>
- Baucells, M., & Heukamp, F. H. (2010). Common ratio using delay. *Theory and Decision, 68*(February), 149–158. https://doi.org/10.1007/s11238-008-9130-2

- Brañas-Garza, P., Jorrat, D., Espín, A. M., & Sánchez, A. (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *26*(April), 412–434. https://doi.org/10.1007/s10683-022-09776-5
- Brodeur, A., Mikola, D., & Cook, N. (2024). *Mass reproducibility and replicability: A new hope*. IZA Discussion Paper No. 16912, IZA Institute of Labor Economics, Bonn, Germany.
- Charness, G., & Chemaya, N. (2023). Repeated experience and consistent risk preferences. *Economics Letters*, 233(December), 111375. <u>https://doi.org/10.1016/j.econlet.2023.111375</u>
- Charness, G., & Viceisza, A. (2016). Three risk-elicitation methods in the field: Evidence from rural Senegal. *Review of Behavioral Economics*, *3*(2), 145–171.

http://dx.doi.org/10.1561/105.0000046

- Charness, G., Chemaya, N., & Trujano-Ochoa, D. (2023). Learning your own risk preferences. *Journal of Risk and Uncertainty*, 67(August), 1–19. <u>https://doi.org/10.1007/s11166-023-09413-3</u>
- Charness, G., Garcia, T., Offerman, T., & Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory. *Journal of Risk and Uncertainty, 60*(April), 99–123. <u>https://doi.org/10.1007/s11166-020-09325-6</u>
- Charness, G., Gneezy, U., & Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131(Part A, November), 141–150.

https://doi.org/10.1016/j.jebo.2016.08.010

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87(March), 43–51.

https://doi.org/10.1016/j.jebo.2012.12.023

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and withinsubject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009 Cheung, S. L. (2016). Recent developments in the experimental elicitation of time preference. *Journal of Behavioral and Experimental Finance*, *11*(September), 1–8.

https://doi.org/10.1016/j.jbef.2016.04.001

- Cheung, S. L. (2020). Eliciting utility curvature in time preference. *Experimental Economics, 23*(June), 493–525. <u>https://doi.org/10.1007/s10683-019-09621-2</u>
- Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and reproducible social science research*. University of California Press.
- Chung, H.-K., Glimcher, P., & Tymula, A. (2019). An experimental comparison of risky and riskless choice
 limitations of Prospect Theory and Expected Utility Theory. *American Economic Journal: Microeconomics*, 11(3), 34–67. <u>https://doi.org/10.1257/mic.20170112</u>
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153. <u>https://doi.org/10.1037/h0045186</u>
- Cohen, M., Tallon, J.-M., & Vergnaud, J.-C. (2011). An experimental investigation of imprecision attitude and its relation with risk attitude and impatience. *Theory and Decision*, *71*(July), 81–109.

https://doi.org/10.1007/s11238-010-9205-8

- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47. https://doi.org/10.1017/s1930297500001819
- Coller, M., & Williams, M. (1999). Eliciting individual discount rates. *Experimental Economics*, 2(December), 107–127. <u>https://doi.org/10.1007/bf01673482</u>

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, 143(4), 1489–1514. <u>https://doi.org/10.1037/a0035976</u>

- Doyle, J. R. (2013). Survey of time preference, delay discounting models. *Judgment and Decision Making*, *8*(2), 116–135. <u>https://doi.org/10.1017/s1930297500005052</u>
- Doyle, J. R., Catherine Huirong Chen, & Savani, K. (2011). New designs for research in delay discounting. *Judgment and Decision Making*, 6(8), 759–770. <u>https://doi.org/10.1017/s1930297500004198</u>
- Duch, M. L., Grossmann, M. R. P., & Lauer, T. (2020). z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet. *Journal of Behavioral and Experimental Finance*, *28*(December), 100400. <u>https://doi.org/10.1016/j.jbef.2020.100400</u>
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.
- Feltovich, N. (2003). Nonparametric tests of differences in medians: Comparison of the Wilcoxon– Mann–Whitney and robust rank-order tests. *Experimental Economics*, 6(November), 273–297. <u>https://doi.org/10.1023/a:1026273319211</u>
- Ferecatu, A., & Önçüler, A. (2016). Heterogeneous risk and time preferences. *Journal of Risk and Uncertainty*, *53*(August), 1–28. <u>https://doi.org/10.1007/s11166-016-9243-x</u>
- Fidanoski, F., Dixit, V., & Ortmann, A. (2024). *Zooming risky-intertemporal models across participant pools*. Mimeo.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(June), 171–178. <u>https://doi.org/10.1007/s10683-006-9159-4</u>
- Fitch, A. K., & Kvam, P. D. (2024). Cognitive mechanisms of subjective value in multi-attribute pricing. *PsyArXiv*, y5v79. <u>https://doi.org/10.31234/osf.io/y5v79</u>
- Franke, M. (2021). An introduction to data analysis. In *michael-franke.github.io*. Retrieved on May 15, 2022, from <u>https://michael-franke.github.io/intro-data-analysis/</u>
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2), 351–401. <u>https://doi.org/10.1257/jel.40.2.351</u>

- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125. <u>https://doi.org/10.1007/s40881-015-0004-4</u>
- Hackethal, A., Kirchler, M., Laudenbach, C., Razen, M., & Weber, A. (2023). On the role of monetary incentives in risk preference elicitation experiments. *Journal of Risk and Uncertainty, 66*(April), 189–213. <u>https://doi.org/10.1007/s11166-022-09377-w</u>
- Harrison, G. W. (2011). Experimental methods and the welfare evaluation of policy lotteries. *European Review of Agricultural Economics*, *38*(3), 335–360. <u>https://doi.org/10.1093/erae/jbr029</u>
- Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments* (pp. 41–196). Emerald Group Publishing Limited.

https://doi.org/10.1016/s0193-2306(08)00003-3

Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005a). Risk aversion and incentive effects: Comment. *American Economic Review*, *95*(3), 897–901.

https://doi.org/10.1257/0002828054201378

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*, *92*(5), 1606–1617.

https://doi.org/10.1257/000282802762024674

Harrison, G. W., Lau, M. I., Rutström, E. E., & Sullivan, M. B. (2005b). Eliciting risk and time preferences using field experiments: Some methodological issues. In G. W. Harrison, J. Carpenter & J. A. List. (Eds.), *Field experiments in economics* (pp. 125–218). Emerald Group Publishing Limited.

https://doi.org/10.1016/S0193-2306(04)10005-7

- Hey, J. D., & Orme, C. (1994). Investigating generalizations of Expected Utility Theory using experimental data. *Econometrica*, 62(6), 1291–1326. <u>https://doi.org/10.2307/2951750</u>
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655. <u>https://doi.org/10.1257/000282802762024700</u>

- Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *Economic Journal*, 127(605), F236–F265. <u>https://doi.org/10.1111/ecoj.12461</u>
- Ioannou, C. A., & Sadeh, J. (2016). Time preferences and risk aversion: Tests on domain differences. Journal of Risk and Uncertainty, 53(August), 29–54. <u>https://doi.org/10.1007/s11166-016-9245-8</u>
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292. https://doi.org/10.2307/1914185
- Kangasrääsiö, A., Jokinen, J. P. P., Oulasvirta, A., Howes, A., & Kaski, S. (2019). Parameter inference for computational cognitive models with approximate Bayesian computation. *Cognitive Science*, 43(6), e12738. https://doi.org/10.1111/cogs.12738
- Kirby, K. N., & Maraković, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review*, *3*(March), 100–104.

https://doi.org/10.3758/bf03210748

Konstantinidis, E., van Ravenzwaaij, D., Güney, Ş., & Newell, B. R. (2020). Now for sure or later with a risk? Modeling risky intertemporal choice as accumulated preference. *Decision*, *7*(2), 91–120.

https://doi.org/10.1037/dec0000103

- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, Elsevier Inc.
- Luckman, A. (2016). *Choosing to gamble and/or wait: How do people make choices which involve both probabilistic and delayed outcomes?* [Doctoral dissertation, University of New South Wales]. UNSW Library. https://doi.org/10.26190/unsworks/19237

Luckman, A., Donkin, C., & Newell, B. R. (2015). Exploring the concept of utility: Are separate value functions required for risky and inter-temporal choice? In D. C. Noelle, R. Dale, A. S.
Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1440–1446). Cognitive Science Society.

- Luckman, A., Donkin, C., & Newell, B. R. (2017). People wait longer when the alternative is risky: The relation between preferences in risky and inter-temporal choice. *Journal of Behavioral Decision Making*, *30*(5), 1078–1092. <u>https://doi.org/10.1002/bdm.2025</u>
- Luckman, A., Donkin, C., & Newell, B. R. (2018). Can a single model account for both risky choices and inter-temporal choices? Testing the assumptions underlying models of risky inter-temporal choice. *Psychonomic Bulletin & Review*, 25(April), 785–792. <u>https://doi.org/10.3758/s13423-</u> 017-1330-8
- Luckman, A., Donkin, C., & Newell, B. R. (2020). An evaluation and comparison of models of risky intertemporal choice. *Psychological Review*, *127*(6), 1097–1138.

https://doi.org/10.1037/rev0000223

- Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics & Finance*, *10*(4), 497–508. <u>https://doi.org/10.1017/s1474747211000448</u>
- Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. <u>https://doi.org/10.21105/joss.01541</u>
- Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: The effect of delay and of intervening events on reinforcement value* (Vol. 5, pp. 55–73). Lawrence Erlbaum Associates, Inc.
- Nicenboim, B., Schad, D. J., & Vasishth, S. (2022). An introduction to Bayesian data analysis for cognitive science. In *vasishth.github.io*. Retrieved on May 12, 2022, from

https://vasishth.github.io/bayescogsci/book/

Noussair, C., & Wu, P. (2006). Risk tolerance in the present and the future: An experimental study. *Managerial and Decision Economics*, 27(6), 401–412. <u>https://doi.org/10.1002/mde.1278</u>

- Önçüler, A., & Onay, S. (2009). How do we evaluate future gambles? Experimental evidence on path dependency in risky intertemporal choice. *Journal of Behavioral Decision Making*, 22(3), 280– 300. https://doi.org/10.1002/bdm.626
- Ortmann, A. (2021). On the foundations of behavioural and experimental economics. In H. Kincaid & D. Ross (Eds.), *A modern guide to philosophy of economics* (pp. 157-181). Edward Elgar Publishing. https://doi.org/10.4337/9781788974462.00013

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3), 497–527.

https://doi.org/10.2307/2998573

- R Development Core Team. (2021). R: A language and environment for statistical computing. *In R Foundation for Statistical Computing*. Retrieved on January 7, 2022, from <u>https://www.R-</u> project.org/
- Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*, *114*(45), 11890–11895.

https://doi.org/10.1073/pnas.1707040114

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision*, *5*(2), 63–94.

https://doi.org/10.1037/dec0000069

- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1446–1465. https://doi.org/10.1037/a0013646
- Samuelson, P. A. (1937). A note on measurement of utility. *Review of Economic Studies*, 4(2), 155–161. https://doi.org/10.2307/2967612
- Shelley, M. K. (1994). Gain/loss asymmetry in risky intertemporal choice. *Organizational Behavior and Human Decision Processes*, *59*(1), 124–159. <u>https://doi.org/10.1006/obhd.1994.1053</u>

- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341
- Somasundaram, J., & Eli, V. (2022). Risk and time preferences interaction: An experimental measurement. *Journal of Risk and Uncertainty*, 65(October), 215–238.

https://doi.org/10.1007/s11166-022-09394-9

- Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics*, 21(June), 383–433. <u>https://doi.org/10.1007/s10683-017-9528-1</u>
- Sun, Y., & Li, S. (2010). The effect of risk on intertemporal choice. *Journal of Risk Research*, *13*(6), 805–820. https://doi.org/10.1080/13669871003606224
- Vandekerckhove, J., Matzke, D., & Wagenmakers, E. J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, J. Townsend, Z. J. Wang & A. Aidels (Eds.), Oxford handbook of computational and mathematical psychology (pp. 300–319). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.14
- Vanderveldt, A., Green, L., & Myerson, J. (2015). Discounting of monetary rewards that are both delayed and probabilistic: Delay and probability combine multiplicatively, not additively. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 148–162.

https://doi.org/10.1037/xlm0000029

- Wall, D., Chapman, G., & Hemmer, P. (2018). Risky intertemporal choice with multiple outcomes and individual differences. In T. T. Rogers, M. Rau, X. Zhu & C. W. Kalish (Eds.), *Proceedings of the* 40th Annual Conference of the Cognitive Science Society Meeting (pp. 2645–2650). Cognitive Science Society.
- Wilkins, R., & Lass, I. (2018). The Household, Income and Labor Dynamics in Australia survey: Selected findings from waves 1 to 16. Melbourne Institute of Applied Economic & Social Research, University of Melbourne.

- Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In D. J. Koehler & N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making (pp. 399–423). Blackwell Publishing. <u>https://doi.org/10.1002/9780470752937.ch20</u>
- Yi, R., de la Piedad, X., & Bickel, W. K. (2006). The combined effects of delay and probability in discounting. *Behavioural Processes*, *73*(2), 149–155.

https://doi.org/10.1016/j.beproc.2006.05.001