PROSPECT THEORY FOR INTERTEMPORAL CHOICE

Immanuel Lampe
University of St. Gallen, Switzerland

Matthias Weber
University of St. Gallen, Switzerland

Abstract
Prospect theory is well understood in contexts without a time dimension. In intertemporal contexts, however, it is unclear how prospect theory should be applied. In particular, it is unclear whether probabilities should be weighted within time periods or whether probabilities of discounted utilities (or present values) of outcome streams should be weighted. Furthermore, it is unclear what parametric specifications of probability weighting and utility functions should be used. We find in a pre-registered experiment on a representative sample that weighting probabilities of discounted utilities (or present values) predicts decisions best. The estimated probability weighting functions are inverse-S shaped, and the utility functions are almost linear. (JEL: D81, D90, G40, D15)

The editor in charge of this paper was Nicola Pavoni.

Acknowledgments: We thank Nicola Pavoni (the editor), three anonymous reviewers, Daron Acemoglu, Renée Adams, Francesco Audrino, Peter Bossaerts, Martin Brown, Patrick Chuard-Keller, Thomas Epper, Francesco Franzoni, Xiaoyi Guo, Özgür Gürerk, Frank Heinemann, Marit Hinnosaar, Lorenz Kueng, Boris van Leeuwen, Jona Linde, Paul Smeets, David Solomon, Jakub Steiner, Stefan Trautmann, Peter Wakker, Michael Weber, and audiences in various presentations for their comments and suggestions. IRB approval was obtained from the University of St. Gallen. We gratefully acknowledge GFF Project Funding from the University of St. Gallen. We are grateful to the Society for Experimental Finance, which awarded us the Vernon L. Smith Young Talent Award for this paper (titled "Intertemporal Prospect Theory" before). In this paper, we make use of data from the LISS panel (Longitudinal Internet Studies for the Social Sciences), which is administered and managed by the non-profit research institute Centerdata (Tilburg University, The Netherlands). Funding for the panel's ongoing operations comes from the Domain Plan SSH and ODISSEI since 2019. The initial set-up of the LISS panel in 2007 was funded through the MESS project by the Netherlands Organization for Scientific Research (NWO). In addition to his primary affiliation, Weber is also affiliated with the Swiss Finance Institute without being employed there.

E-mail: sitlampe@gmail.com (Lampe); matthias.weber@unisg.ch (Weber)

Journal of the European Economic Association 2026 24(2):701–735
https://doi.org/10.1093/jeea/jvaf037

Teaching Slides

A set of Teaching Slides to accompany this article is available online as Supplementary Data.

1. Introduction

Analyzing decisions under risk is an important topic in many fields of research, including economics, finance, business, health, and psychology. The most prominent contender of expected utility theory is (cumulative) prospect theory (Tversky and Kahneman 1992). Prospect theory entails two fundamental breakaways from the classical expected utility model. First, instead of defining preferences over total wealth, preferences are defined over changes with respect to a reference point, and it is assumed that negative changes (losses) are treated differently than positive changes (gains). Second, outcomes are not weighted by objective probabilities but rather by transformed probabilities (specifically, unlikely large gains and unlikely large losses are overweighted).

Considering financial decisions, prospect theory can, for example, explain the equity premium puzzle (Benartzi and Thaler 1995; Barberis and Huang 2006), the stock market participation puzzle (Dimmock and Kouwenberg 2010), and the disposition effect (e.g., Barberis and Xiong 2009; Meng and Weng 2017). In other fields of economics, prospect theory can explain the downward-sloping labor supply curve (Goette, Huffman, and Fehr 2004) or the avoidance of probabilistic insurance (Wakker, Thaler, and Tversky 1997) and index insurance (Lampe and Würtenberger 2020). DellaVigna (2009) and Barberis (2013) provide overviews of many more phenomena that are inconsistent with expected utility theory but can be explained with prospect theory.

Most applications of prospect theory assume an atemporal setting, that is, a setting where the gains and losses materialize at one point in time. In these settings, prospect theory has been thoroughly analyzed and is well understood. Many natural settings, however, involve gains and losses at different points in time. Recent studies integrate elements of prospect theory in such intertemporal settings, for instance, prominently in finance (Meng and Weng 2017; van Bilsen, Laeven, and Nijman 2020; among many others). However, in intertemporal applications, usually only some aspects of prospect theory are incorporated into a setting that is otherwise one of discounted expected utility maximization. Many applications include loss aversion in their models but ignore probability weighting (Easley and Yang 2015; Meng and Weng 2017; among many others).

To date, there is no generally accepted version of prospect theory in intertemporal settings. There are not only minor disagreements about the calibration of the parameters, but there is not even agreement on how the probability weighting should be applied. Considering that prospect theory is widely viewed as the most accurate available description of how people make risky decisions, this is surprising. Our paper provides evidence on how prospect theory should best be applied in intertemporal situations. Whereas the best way of applying prospect theory may depend on the context (as different application methods may capture different aspects

of intertemporal choice), we aim to analyze the question in a neutral setting, which does not favor one particular application method.

We refer to the first possible method that can be used as the time-separation method. With this method, outcomes are ranked within each time period and decision weights are calculated (with probability weighting functions similar to an atemporal setting). Given these decision weights, each time period's prospect theory value (PT value) can be calculated, and the overall PT value is then the time-discounted sum of these values. This first method has, for instance, been used in Andreoni and Sprenger (2012), De Giorgi and Legg (2012), Guo and He (2017), Krause, Ring, and Schmidt (2020), and van Bilsen and Laeven (2020).

The second method, which we refer to as the risk-separation method, first calculates the decision maker's discounted utility of each possible stream of outcomes. These utility values are then ranked, and their decision weights are calculated (as in atemporal prospect theory). The overall PT value is then calculated as the atemporal PT value of these discounted utilities. This method (or a version of it with rank-dependent utility, not distinguishing between gains and losses but including probability weighting) has been used in Chew and Epstein (1990), Bleichrodt and Eeckhoudt (2006), Halevy (2008), Drouhin (2015), Epper and Fehr-Duda (2015), and Blavatskyy (2016).[1]

Whereas the two methods yield the same results without probability weighting (that is, under classical discounted expected utility theory), they, in general, yield different results with probability weighting, which can lead to substantive differences (Section 6). The first step towards understanding how prospect theory should be applied in intertemporal situations is to understand which of these two methods describes human choices best in a neutral intertemporal setting.

A second open issue when moving from prospect theory for atemporal choice to prospect theory for intertemporal choice is the specification of the utility and probability weighting functions (for a given method of applying prospect theory). Many studies integrating elements of prospect theory into intertemporal models use functional forms and parameter calibrations obtained in atemporal settings. It is *a priori* not clear that these specifications of the utility and probability weighting functions are well-suited to analyze intertemporal choice. Abdellaoui et al. (2013a) and Abdellaoui, l'Haridon, and Paraschiv (2013b), for instance, suggest that loss aversion is smaller in intertemporal contexts than in atemporal contexts.

In this study, we elicit certainty equivalents of intertemporal lotteries in a representative sample to determine which of the two methods describes choices better. We also estimate the parameters for a variety of different utility, probability weighting, and time-discount functions jointly. Scholars and practitioners may benefit from our estimation results when applying prospect theory in intertemporal contexts, as they can then use parametrizations obtained in an intertemporal setting (estimating the parameters anew also assures that the results of the comparison of the methods are not driven by potentially unfitting atemporal calibrations).

To preview our results, the risk-separation method predicts observed choices better than the time-separation method for all considered combinations of functional forms.

1. Some authors were certainly aware that a choice needed to be made, which method to use. However, others were likely unaware that there were different options.

The prediction performance of the risk-separation method also improves considerably over the prediction performance of expected discounted utility (EDU) models: mean squared errors (MSEs) for EDU models are similar to mean squared errors for the time-separation method and about 25% higher than for the risk-separation method. The estimated probability weighting functions are inverse-S shaped and very similar to typical atemporal calibrations. The estimated utility functions are almost linear for gains and losses and have a loss aversion coefficient close to 1.[2]

While our main focus lies on the comparison of the time-separation method and the risk-separation method, we also compare these methods to a variant of the risk-separation method. This variant also separates risks, but it calculates present values of outcome streams (in monetary values) before entering them in a utility function (in contrast to the regular risk-separation method, which calculates discounted utilities, thus "present values" directly in utility terms). That variant performs as well as the regular risk-separation method.

Very closely related literature is scarce. Epper and Fehr-Duda (2015) nicely illustrate the co-existence of the two methods and show that the risk-separation method can explain a decision pattern that cannot be explained with the time-separation method. Rohde and Yu (2024) investigate intertemporal correlation aversion in a laboratory experiment. Intertemporal correlation aversion can arise if subjects use the risk-separation method to evaluate the lotteries but not if they use the time-separation method. Rohde and Yu (2024) find that the majority of subjects are averse to intertemporal correlation; their results thus lend support to the risk-separation method. Li, Rohde, and Wakker (2024) provide a theoretical framework for the optimization over two components (such as time and risk). They argue that risk is more easily separable than time. These recent studies, which are methodologically very different, are thus consistent with ours.

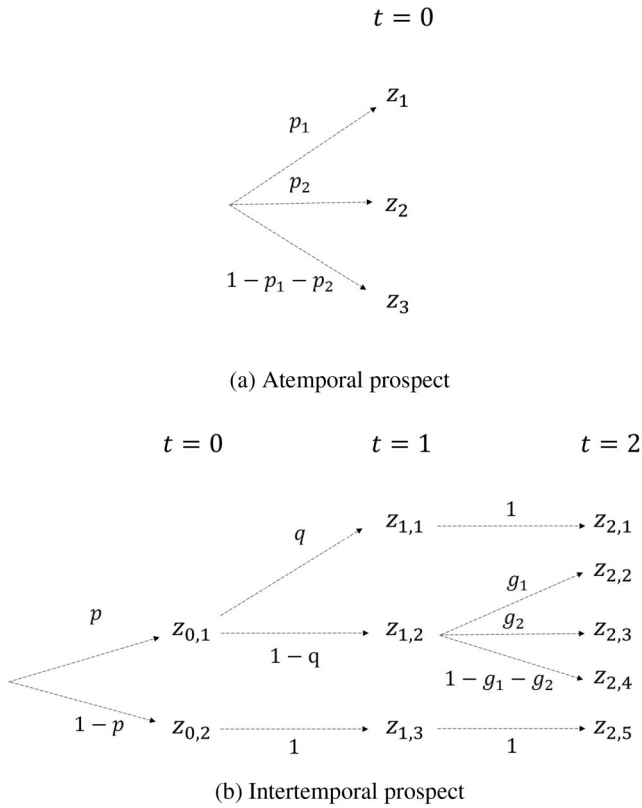
This paper is organized as follows. Section 2 introduces the two methods of applying prospect theory to intertemporal settings. Section 3 contains the experimental design and discusses the pre-registration and the experimental procedures. Section 4 discusses parametric specifications, the estimation method, and the measure of prediction performance. Section 5 shows the results. Section 6 sketches an application example, and Section 7 concludes.

2. Modeling Options for Intertemporal Prospects

2.1. Intertemporal Prospects

An atemporal prospect yields a payout in one time period. Figure 1(a) displays an example of such a prospect. The three potential outcomes (arriving at $t = 0$, without time discounting) are labeled z_1, z_2, and z_3. Outcome z_1, for instance, arrives with probability p_1.

2. In the experiment, we implement salient reference points of 0 in all time periods. In applications, reference points may differ across periods, of course.



(a) Atemporal prospect

(b) Intertemporal prospect

FIGURE 1. Examples of an atemporal and an intertemporal prospect.

An intertemporal prospect yields payouts at different time periods. Figure 1(b) provides an example. The potential outcomes of period t are labeled by $z_{t,1}, z_{t,2}, \ldots$. A period's outcomes are not necessarily distinct (that is, $z_{1,2} = z_{1,3}$, for instance, is possible) and the labels do not necessarily reflect the outcomes' rank-order. In the first period ($t = 0$), either outcome $z_{0,1}$ arrives (with probability p) or outcome $z_{0,2}$ (with probability $1 - p$). Outcomes at $t = 1$ and $t = 2$ depend in general on the outcomes of previous periods. The probability that one specific outcome arrives is, thus, equal to the product of its path probabilities. The probability that the outcome at $t = 1$ equals $z_{1,1}$ is, for instance, given by $p \cdot q$.

This paper analyzes the evaluation of such intertemporal prospects at one point in time (at the beginning). We thus do not study dynamic decision making. Time consistency, for instance, a dynamic decision principle, plays no role in our study.

2.2. Time-Separation Method

To evaluate a prospect with outcomes in T periods, the time-separation method first calculates each time period's PT value (note that we call any evaluation of a lottery in

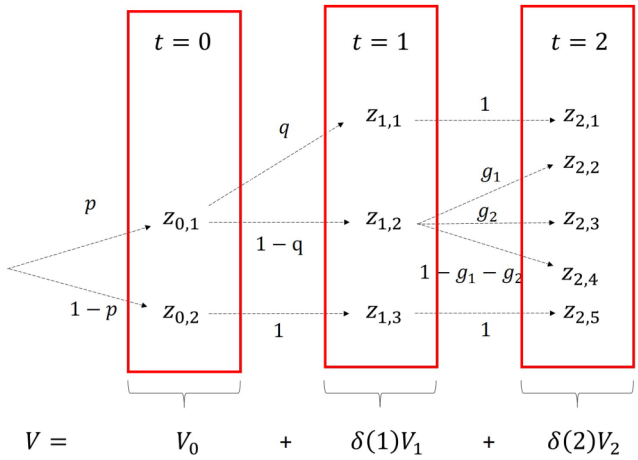


FIGURE 2. Illustration of the time-separation method.

utility terms with a version of prospect theory a PT value; this holds for calculations of such values within time periods in the time-separation method but also for the overall evaluation of an intertemporal lottery with the time-separation or the risk-separation method). These values are then discounted with a given time-discount function δ (generally with $\delta(0) = 1$) and added up. We denote the PT value of period t by V_t, $t = 0, \ldots, T-1$. Applying the time-separation method to the prospect in Figure 1(b) implies that its overall PT value is the sum of its three time-discounted per-period PT values, as illustrated in Figure 2.

To calculate each time period's PT value, prospect theory is applied as in an atemporal setting. Periods' outcomes are ranked and their objective probabilities are transformed into subjective decision weights. Denoted by $x_{t,1}, \ldots, x_{t,n_t}$ the ordered and distinct outcomes of period t and by $p_{t,1}, \ldots, p_{t,n_t}$ their objective probabilities. Each time period's outcome set may consist of gains and losses, so that $x_{t,1} > x_{t,2} > \cdots > x_{t,k_t} > 0 > x_{t,k_t+1} > \cdots > x_{t,n_t}$. Note that the objective probabilities are compound probabilities (for example, if $z_{1,1} > z_{1,2}$ and $z_{1,1} > z_{1,3}$ in Figures 1B and 2, then $x_{1,1} = z_{1,1}$ and $p_{1,1} = p \cdot q$). The PT value of period t can then be calculated using the formula:

$$V_t = \sum_{i=1}^{k_t} \pi_{t,i}^{+} v(x_{t,i}) + \sum_{l=k_t+1}^{n_t} \pi_{t,l}^{-} v(x_{t,l}),$$

with v denoting the decision maker's utility function (also called value function in prospect theory), which is strictly increasing. We assume that the reference point is at 0 and $v(0) = 0$. Further, the decision weights $\pi_{t,i}^{+}$ and $\pi_{t,i}^{-}$ are defined by

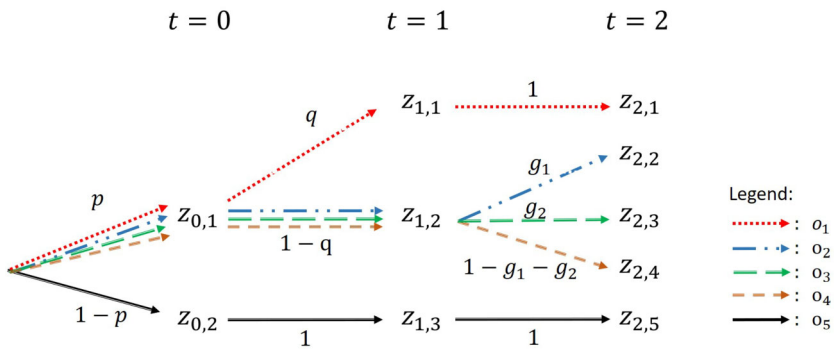


FIGURE 3. Illustration of the risk-separation method.

$$\pi^+_{t,1} = w^+(p_{t,1}),\ \pi^-_{t,n_t} = w^-(p_{t,n_t}),$$
$$\pi^+_{t,i} = w^+(p_{t,1} + \cdots + p_{t,i}) - w^+(p_{t,1} + \cdots + p_{t,i-1}),\ \text{for } 1 < i \le k_t,$$
$$\pi^-_{t,l} = w^-(p_{t,l} + \cdots + p_{t,n_t}) - w^-(p_{t,l+1} + \cdots + p_{t,n_t}),\ \text{for } k_t < l < n_t.$$

Here, w^+ is the probability weighting function for gains and w^- that for losses, satisfying $w^+(0) = w^-(0) = 0$, $w^+(1) = w^-(1) = 1$. Both weighting functions are nondecreasing.

The overall value of the prospect, which we label V, is then the sum of all time-discounted per-period PT values, thus $V = \sum_{t=0}^{T-1} \delta(t)V_t$. We call this method time-separation method, because the different time periods are first evaluated separately, before the aggregation over time takes place.

2.3. *Risk-Separation Method*

The risk-separation method evaluates an intertemporal prospect by calculating the discounted utilities of the possible streams of outcomes and then weighting the probabilities of these discounted utilities (one could also call the discounted utilities present values in utils to avoid that people think of EDU theory here; present values in monetary terms will play a role later, in Section 5.4). An outcome stream is a sequence of outcomes that arrives with positive probability. We denote the different possible outcome streams by $o_1, o_2, \ldots$. An outcome stream o_j thus consists of T outcomes, one in each time period, $o_j = (o_{j,0}, \ldots, o_{j,T-1})$, with $o_{j,t}$ denoting the outcome in period t. Figure 3 illustrates the outcome streams of the example given in Figure 1(b) (for o_2, for instance, $o_{2,0} = z_{0,1}$, $o_{2,1} = z_{1,2}$, and $o_{2,2} = z_{2,2}$).

The discounted utility (or present value in utils) of outcome stream o_j is given by $PV(o_j) = \sum_{t=0}^{T-1} \delta(t)v(o_{j,t})$. The discounted utility of o_1 in Figure 3 is, for instance, given by $v(z_{0,1}) + \delta(1)v(z_{1,1}) + \delta(2)v(z_{2,1})$, with $\delta(0) = 1$. We denote the number of distinct discounted utilities by m and rank-order them, so that $PV_1 > \cdots > PV_k > 0 > PV_{k+1} > \cdots > PV_m$. The objective probability of receiving the outcome stream

with discounted utility PV_j is denoted by q_j. For example, the objective probability of receiving (the discounted utility of) outcome stream o_1 in Figure 3 is $p \cdot q \cdot 1$.

The overall PT value of the intertemporal prospect, W, is the sum of all discounted utilities multiplied by their decision weights as in atemporal prospect theory,

$$W = \sum_{j=1}^{k} \pi_j^+ PV_j + \sum_{h=k+1}^{m} \pi_h^- PV_h,$$

with the decision weights for gains and losses defined by

$$\begin{aligned}
&\pi_1^+ = w^+(q_1), \pi_m^- = w^-(q_m),\\
&\pi_j^+ = w^+(q_1 + \cdots + q_j) - w^+(q_1 + \cdots + q_{j-1}), \text{ for } 1 < j \leq k,\\
&\pi_h^- = w^-(q_h + \cdots + q_m) - w^-(q_{h+1} + \cdots + q_m), \text{ for } k < h < m.
\end{aligned}$$

We call this method risk-separation method (in an earlier version of this paper, the method was called present-value method).[3]

2.4. Differences between the Methods

In general, with non-linear weighting of probabilities, both methods yield different evaluations of prospects (see Li, Rohde, and Wakker 2024; for a brief application example, see Section 6). However, for a subset of lotteries, both methods yield the same PT values (thus, exactly the same evaluations, assuming the same parametrization for both methods). In particular, the methods give the same PT value for a lottery if any outcome stream of the lottery contains only non-negative or non-positive outcomes and if, in addition, the lottery possesses a property that we call rank-order stability. Intuitively, rank-order stability means that if one outcome is larger than another outcome within one period, then, for any other period, the outcomes that may arrive in addition to the larger outcome (i.e., outcomes that are in one of the outcome streams containing this larger outcome) must be at least as large as the outcomes that may arrive in addition to the smaller outcome. Rank-order stability, therefore, implies that the best outcome stream bundles the best outcomes of each period. In Online Appendix A, we provide an example, a formal definition, and a proof that lotteries with these properties are indeed evaluated equally.[4]

3. Note that both of the methods involve compound probabilities. It is also possible to model intertemporal choice without compound probabilities (e.g., Kreps and Porteus 1978; Epstein and Zin 1989). However, we focus on analyzing these particular two methods and on a variant of the risk-separation method (Section 5.4); analyzing further very different models would be beyond the scope of this paper.

4. For most economic or financial applications, the evaluations between the methods will differ. For instance, if a good outcome in one period can be followed by a bad outcome in another period, rank-order stability does not hold. In addition, even if rank-order stability holds, the evaluations can differ if some outcome streams contain both positive and negative outcomes.

3. Experimental Design and Procedures

3.1. Decision Tasks

Each subject makes 48 decisions. For each decision, a subject is shown a lottery (framed neutrally as a risky option) with a time horizon of three periods ($t = 0, 1, 2$). The first period ($t = 0$) refers to the time of the experiment, the second period ($t = 1$) to 3 months after the experiment, and the third period ($t = 2$) to 6 months after the experiment. Sixteen of the lotteries only contain positive outcomes (gain lotteries), 16 only contain negative outcomes (loss lotteries), and 16 contain positive and negative outcomes (mixed lotteries).

For each lottery, we elicit a subject's switching point from the lottery to a safe option that yields three certain and identical payouts (with the same timing as for the lotteries; that is, at the time of the experiment, 3 months later, and 6 months later). We elicit the switching point by asking subjects to position a slider that indicates for which payout amounts of the safe option they prefer the lottery over the safe option (and vice versa). The safe option contains three payouts with the same timing as the lottery payouts, so that the key difference between a lottery and a certainty equivalent lies in the riskiness of the options, as for certainty equivalents in atemporal choice (we thus do not introduce an additional difference in the timing of the payouts).

Figure 4 shows an example of such a decision task. At the top, subjects see the intertemporal lottery of this task (with non-trivial probabilities illustrated using a wheel of fortune). We designed the lottery illustration with the aim not to favor one method over the other (as our aim is to compare the application methods in an abstract situation that is as neutral as possible). Arrows are shown, which are related to outcome streams, but therefore gray boxes highlighting the time dimension are also shown.

Below the lottery, subjects see an illustration of the safe option (without the monetary values filled in because these depend on a subject's choice in the task). Furthermore, the decision screen shows a short description of the meaning of the position of the slider and the slider that they can position. The slider is initially always placed at the leftmost position (as depicted in Figure 4). When subjects move the slider to the right, they increase the amount that they need to be offered in the safe option (three times) in order to prefer the safe option over the lottery. The bold numbers in the row above the slider change when the slider is moved. The difference between these two bold numbers is always 1 euros (in the analysis, we take the midpoint between these two numbers as the certainty equivalent). The limits of the slider (i.e., the leftmost and rightmost positions) are the lottery's minimal and maximal possible sums of payouts across the three time periods (without time discounting) divided by three, as the safe option payout amount is received three times.[5]

5. Admittedly, the decision screen remains complex. We believe that this reflects the fact that the studied decisions are complex, so that it cannot be avoided. The stimuli are also arguably not more complex than the stimuli in real-world applications.

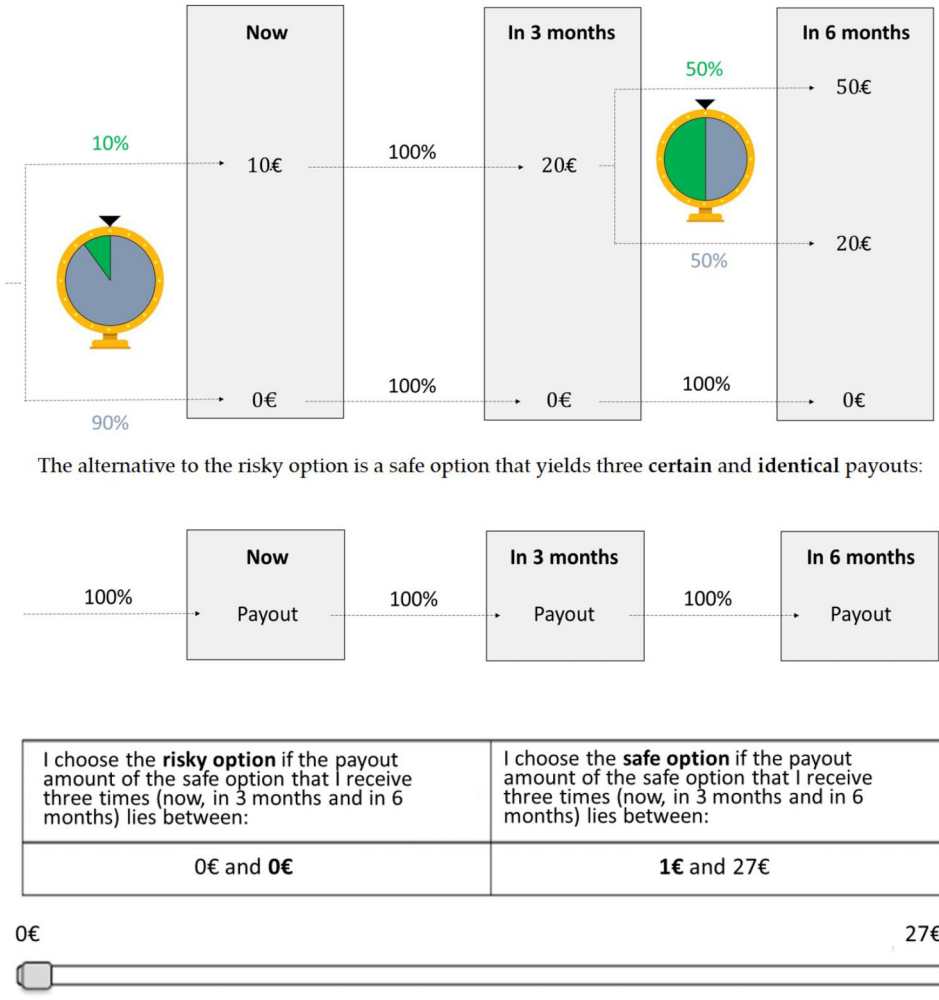


FIGURE 4. Example of a decision task. On top is the decision task's intertemporal lottery, followed by an illustration of the safe option (without payout values). The bottom part shows a description of the slider and the slider to elicit the certainty equivalent of the intertemporal lottery.

3.2. Lotteries

All lotteries used in the experiment, with a description of why they have been chosen, can be found in Online Appendix B. Here, we give an overview of the structure behind the chosen lotteries.

The following design features hold for all lotteries and aim at keeping the experiment comprehensible for subjects: (1) each branching in the trees describing the intertemporal lotteries leads to at most two distinct outcomes in the following period; (2) in the last period, the number of distinct outcomes is limited to four (this is achieved

TABLE 1. Overview lottery sets.

	Gains	Losses	Gains & losses
Low stakes	Set 1	Set 2	Set 3
High stakes	Set 4	Set 5	Set 6

by relying on outcomes that arrive with probability one in one time period); and (3) outcomes appear only with probabilities 1, 9/10, 2/3, 1/2, 1/3, or 1/10.

The 48 lotteries of the experiment can be categorized into six sets, each containing eight lotteries. The first three sets consist of low-stake lotteries: low-stake gain lotteries (Set 1), low-stake loss lotteries (Set 2), and low-stake mixed lotteries (Set 3). The other three sets consist of high-stake gain lotteries (Set 4), high-stake loss lotteries (Set 5), and high-stake mixed lotteries (Set 6). The sets are summarized in Table 1. The different sets appear in random order. Lotteries within each set similarly appear in random order.

In general, the loss lotteries (Sets 2 and 5) are obtained by multiplying all outcomes of the gain lotteries (Sets 1 and 4) by −1. The high-stake lotteries (Sets 4–6) are obtained by multiplying the low-stake lotteries (Sets 1–3) by 20. The 8 lotteries of each set consist of 6 calibration and 2 test lotteries, implying a total of 36 calibration and twelve test lotteries.

Calibration Lotteries. We use the certainty equivalents reported for the calibration lotteries to estimate the parameters of utility function, weighting functions, and time-discount function jointly. The calibration lotteries ensure that all functions can be estimated on a relatively dense grid, for the utility functions, including small, intermediate, and large positive and negative values.

The outcome streams of the calibration lotteries contain only non-negative or non-positive outcomes (also in the mixed-lottery sets), and the calibration lotteries are rank-order-stable. As explained in Section 2.4, this implies that for any given combination of utility function, probability weighting functions, and time-discount function, the evaluations under the time-separation method and the risk-separation method are identical. The parameter estimates obtained on the calibration lotteries are, thus, the same for both methods.

Figure 5 shows two explanatory calibration lotteries from Set 1 (low-stake gain lotteries). All outcomes are in euros. Lottery 1 contains several unlikely good outcomes. Lottery 3, in contrast, contains several likely good outcomes. The variation in outcomes contributes to the utility function being estimated on several small positive values. Similarly, the variation in good outcomes' arrival probabilities contributes to the probability weighting function for gains being estimated on several values in the interval [0,1].

Test Lotteries. The test lotteries are designed such that the evaluations under the two methods give (sizably) different results. The test lotteries can therefore be used

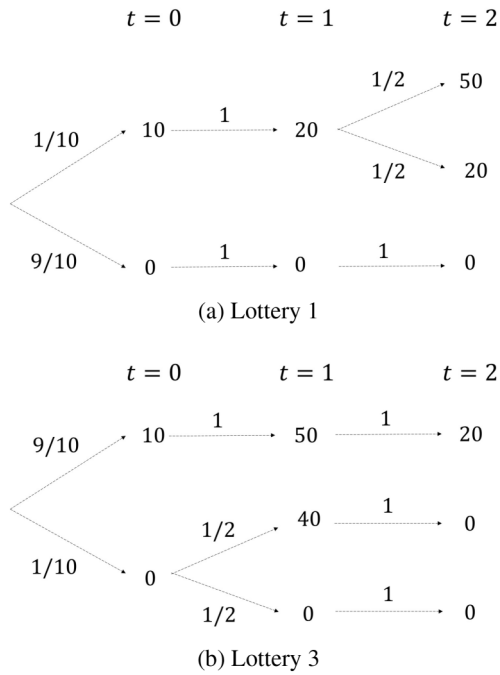
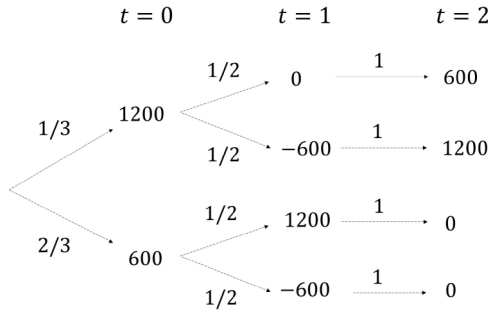


FIGURE 5. Calibration lottery examples, Lottery Set 1 (low-stake gain lotteries).

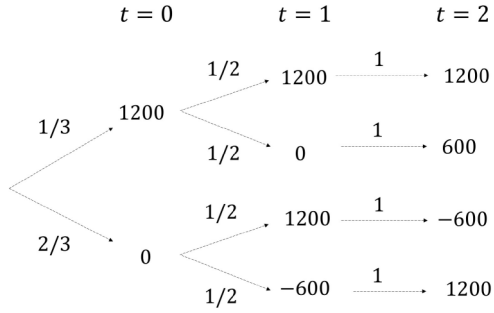
to compare the out-of-sample explanatory power of the two methods. In each of the six sets, one of the two test lotteries is chosen such that the time-separation method leads to a higher (multi-period) certainty equivalent than linear probability weighting if parameters are similar to the ones usually found for atemporal prospect theory, while the risk-separation method leads to a lower certainty equivalent. The opposite holds for the other test lottery. In order to assure a fair comparison between the two methods, we choose the test lotteries such that the degree to which one method over and the other method undervalues the lottery is similar (assuming typical atemporal calibrations).

Figure 6 shows the two test lotteries from Set 6 (high-stake mixed lotteries). In Lottery 47, the time-separation method would simultaneously underweight the most negative outcome at $t = 1$ (-600 arrives with probability $1/2$) and overweight the most positive outcome at $t = 2$ (1200 arrives with probability $1/6$). The time-separation method would then overvalue the lottery. For the risk-separation method, the lottery contains three outcome streams with positive discounted utility: (1200, -600, 1200), (1200, 0, 600), and (600, 1200, 0). The discounted utility of these streams should be close to each other. The probability that one of these three streams arrives is $2/3$. The risk-separation method would therefore underweight the probability of such a positive discounted utility and undervalue the lottery.

In Lottery 48, at $t = 1$ and at $t = 2$, the outcome 1200 arrives with probability $1/2$. The time-separation method would underweight the positive outcome 1200 in these



(a) Lottery 47



(b) Lottery 48

FIGURE 6. Test lottery examples, Lottery Set 6 (high-stake mixed lotteries).

TABLE 2. Test lottery examples, certainty equivalent (CE) calculation.

	No weighting	Time-separation		Risk-separation	
	CE (1)	CE (2)	Rel. deviation (3)	CE (4)	Rel. deviation (5)
Lottery 47	220.3	265.4	+20%	171.8	−22%
Lottery 48	195.9	141.1	−28%	248.8	+27%

Notes: To calculate the certainty equivalents, we use the parametric specifications suggested by Tversky and Kahneman (1992), that is, $v(x) = \mathbb{1}(x \geq 0)x^{0.88} - \mathbb{1}(x < 0)2.25(-x)^{0.88}$ and $w^+(p) = w^-(p) = \frac{p^{0.69}}{\left(p^{0.69}+(1-p)^{0.69}\right)^{1/0.69}}$ (without time discounting).

two periods and undervalue the lottery. For the risk-separation method, note that the outcome stream (1200, 1200, 1200) arrives with probability 1/6. The risk-separation method therefore overweights this stream and overvalues the lottery.

In order to quantify the over- and undervaluation, we calculate the certainty equivalent of each lottery using typical atemporal specifications of the weighting and utility functions (omitting time discounting). In line with our experimental design, this certainty equivalent would be received three times. Table 2 displays the results for two lotteries as an illustration, showing that the degrees of over- and

undervaluation are similar. Column (1) shows the certainty equivalent that is calculated with linear probability weighting. Column (2) displays the certainty equivalent that is calculated with the time-separation method, and column (3) its relative deviation from the evaluation with linear probability weighting. Column (4) shows the certainty equivalent implied by an evaluation with the risk-separation method and column (5) its relative deviation from the evaluation with linear probability weighting. As shown in Online Appendix B, these characteristics are shared by all twelve test lotteries, enabling an *a priori* fair comparison of the two methods.

3.3. Hypothetical and Incentivized Lotteries

Treatment comparisons are not the main focus of the paper. We implement two treatments, in which the remuneration of subjects varies. For subjects in treatment T1, all decisions are hypothetical and subjects receive a fixed payment only (15 euros). For subjects in treatment T2 (who also receive the fixed payment of 15 euros), the gain lotteries are incentivized (mixed and loss lotteries are hypothetical as in T1). Each subject in the experiment has a 25% chance to be assigned to the incentivized treatment T2. Subjects know whether a part of their choices have monetary implications (and which choices these are) or whether all choices are hypothetical. These two treatments serve as robustness check of whether choices are consistent with and without monetary consequences. In the incentivized treatment, the payouts from the three different time periods are actually paid at three different times. This is clearly communicated to subjects in the experimental instructions.[6]

For subjects in T2, one gain lottery is randomly selected for payment at the end of the experiment. There is a 99% chance that a low-stake lottery is selected and a 1% chance that a high-stake lottery is selected (for evidence that incentives implemented with random problem-selection procedures work well in individual decision experiments without strategic interaction, see Umer 2023). Then a standard payment mechanism is employed. A random number is drawn from a uniform distribution, defined over the interval from $LP_{min}/3$ to $LP_{max}/3$, where LP_{min} is the minimal and LP_{max} the maximal non-discounted sum of payouts of the selected lottery (these are the limits of the sliders in the decision tasks, as described in Section 3.1). The terms LP_{min} and LP_{max} are divided by three because the certainty equivalent is paid thee times. If the drawn random number is greater than or equal to the subject's stated switch point, the subject receives the random number three times. If the random number

6. The reason for using these treatments is as follows. Conducting the experiment with monetary incentives for all lotteries is not possible in a straightforward manner because of the loss lotteries (implementing real losses is not possible with this subject pool, and it would be ethically questionable; resorting to a large endowment from which losses are deducted would be problematic because it is unclear whether subjects would consider such deductions as actual losses). Conducting the experiment only with hypothetical lotteries might lead to the criticism that hypothetical choices do not necessarily correspond to choices with monetary incentives. If hypothetical and incentivized choices are similar for the gain lotteries, this gives good reason to believe that the hypothetical choices in our setting would generally be similar to incentivized choices.

is lower, the subject keeps the lottery and the reward is determined by a realization of the lottery.[7]

3.4. Pre-Analysis Plan

The study was pre-registered at the Open Science Foundation (https://osf.io/7zyp4) with a detailed pre-analysis plan (https://osf.io/fv8a2). We follow the pre-analysis closely for our main analysis. Minor existing deviations from the pre-analysis plan are explicitly mentioned.

3.5. Procedures

The data was collected on the LISS (Longitudinal Internet Studies for the Social Sciences) panel, a subject pool that is representative of the Dutch population, consisting of about 9000 Dutch individuals (at the time of the data collection).[8] Data collection and data analysis were entirely separated. The experiment was conducted in Dutch. Instructions and other materials were translated from English to Dutch by the LISS panel team. We made sure in several iterations that the translations were accurate (one of us speaks Dutch, and we checked with an experimental economist who is a native speaker). Before starting the decision tasks, subjects had to correctly answer a set of comprehension test questions. The English version of the instructions and comprehension test questions can be found in Online Appendix C. All subjects who completed the experiment received 15 euros for participating, subjects in the incentivized treatment received, on average, 84 euros in addition to this. Payments were made by bank transfers (in the incentivized treatment, the payments from the first time period were transferred jointly with the payment for participation; the other payments were transferred 3 and 6 months later). Subjects are used to receiving such bank transfers, as this is the standard payment mechanism used for the LISS panel. Our experiment constituted an own "survey," it was not integrated into another survey. Before running the experiment, two pilots with about 25 subjects each were conducted

7. This random price mechanism is relatively difficult for subjects to understand. An alternative would have been using choice lists (which would then have needed to be implemented in both treatments instead of the slider). We opted for the slider (which implied using the random price mechanism in the incentivized treatment), because choice lists would either have become confusingly large for some lotteries, or we would have needed to use a much larger step size for the choices, giving us less precise estimates, which we wanted to avoid. Ex post, the fact that choices are very similar in both treatments suggests that using the random price mechanism was not a problematic design choice.

8. The description of the panel by Centerdata is as follows: "The LISS panel is a representative sample of Dutch individuals who participate in monthly internet surveys. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Self-registration is not possible, and households that would otherwise be unable to participate are provided with a computer and internet connection. The longitudinal LISS Core Study, consisting of separate topical surveys, has been conducted in the panel every year since 2007, covering a large variety of domains including health, work, education, income, housing, leisure and time use, political views, values and personality." More information can be found at www.lissdata.nl/ and in Scherpenzeel and Das (2010).

online with the student subject pool of the Behavioral Lab of the University of St. Gallen. These pilots served to make the elicitation procedure and the instructions as comprehensible as possible.

In a post-experimental questionnaire, we asked three questions about the comprehension of the tasks and the attention paid in the experiment:

- During the decision tasks, we asked you to position a slider. Was it clear to you what the position of the slider meant?
- During the decision tasks, we asked you to evaluate risky options with uncertain payouts at three different points in time. How complicated were these decision tasks?
- Please honestly report your attention during the questionnaire.

Each of these questions only had three answer possibilities. As outlined in our pre-analysis plan, we exclude all subjects from the analysis that answer at least one of these questions with the "worst" answer possibility (indicating that it was unclear what the slider meant, that the decision tasks were too complicated, or that a subject only paid attention in a few parts of the experiment or not at all). In addition (and as pre-registered), we exclude all subjects that completed the tasks with a median decision time of less than 15 seconds. Such strict exclusion criteria serve to make sure that only data are used that stem from subjects who understood the decision tasks and took them seriously. It is thus unlikely that the results are driven by subjects who misunderstood the choice problem or by subjects who clicked through the experiment fast to collect the participation fee in a short period of time.

A total of 391 subjects started the decision tasks, 13 of them did not finish the experiment.[9] Of the remaining 378 subjects, we exclude 192 subjects based on the answers to the post-experimental questionnaire and an additional 86 subjects that finished the decision tasks with a median time per task of less than 15 seconds. Following our pre-registered data exclusion rules, we therefore drop about 74% of our sample. Importantly, using the full sample of 378 subjects does not affect our main conclusion of the superiority of the risk-separation method over the time-separation method. One could also choose different and less strict exclusion criteria than the ones that we pre-registered. The main results for a set of less strict exclusion criteria and for the full sample are provided in Online Appendix D. Note that exclusion rates for studies estimating preference parameters on representative samples are often only slightly below ours, despite the fact that they usually have less complex tasks, without time and risk dimensions simultaneously (e.g., Booij, Van Praag, and Van De Kuilen

9. Our pre-registration specifies that the target number of subjects was 240. If 240 had been reached by the end of September 2020, the experiment should have stopped, otherwise continued in October. At the end of September, 205 subjects had completed the experiment. We did not receive the data at this point, only summary information on payments and on the responses from the post-experimental questionnaire. Based on this, we asked the LISS panel team to continue the data collection in October to obtain in total between 350 and 400 subjects. Data collection then stopped at the end of October, and we obtained the data in December.

TABLE 3. Function specifications.

Specification	Parameters
Utility functions	
Power utility: $v(x) = \mathbb{1}_{x\geq 0}x^{\alpha} - \mathbb{1}_{x<0}\lambda(-x)^{\alpha}$	α, λ
Exponential utility: $v(x) = \mathbb{1}_{x\geq 0}\frac{1-\exp(-\alpha x)}{\alpha} - \mathbb{1}_{x<0}\lambda\frac{1-\exp(\beta x)}{\beta}$	α, β, λ
Probability weighting functions (gains and losses)	
Tversky and Kahneman (1992): $w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma}}$	γ^{+}, γ^{-}
Prelec et al. (1998): $w(p) = \exp(-\nu(-ln(p))^{\gamma})$	$\gamma^{+}, \gamma^{-}, \nu^{+}, \nu^{-}$
Goldstein Einhorn (1987): $w(p) = \frac{\nu p^{\gamma}}{\nu p^{\gamma}+(1-p)^{\gamma}}$	$\gamma^{+}, \gamma^{-}, \nu^{+}, \nu^{-}$
Time-discount functions	
Exponential discounting: $\delta(t) = \exp(-rt)$	r
Quasi-Hyperbolic discounting: $\delta(t) = \mathbb{1}_{t=0} + \mathbb{1}_{t>0}k\exp(-rt)$	k, r

Notes: This table shows the formulas for the considered utility functions, probability weighting functions (for gains and losses), and time-discount functions. For each function, the table lists the parameters that have to be estimated.

2010, who use the CentERPanel in The Netherlands, and Guiso and Paiella 2008, who use a representative sample in Italy, with exclusion rates of 61% and 57%).

Out of the exactly 100 subjects that remain in our main analysis, 27 are in the incentivized treatment and 73 are in the hypothetical treatment. Demographic variables of these 100 subjects are similar to the demographic variables of all subjects starting the experiment and also of the LISS panel subject pool. Details on responses in the post-game survey, the distribution of median decision times, and demographic characteristics of subjects and the LISS panel subject pool can be found in Online Appendix D.

4. Estimation Procedure

Here, we describe the parametric specification, the maximum-likelihood procedure to estimate the model parameters on the calibration set, the measurement of prediction performance on the test set, the calculation of tests and standard errors, and the final re-estimation of the parameters. The described approach is exactly as in the pre-analysis plan.

4.1. Parametric Specification

We consider two families of utility functions, three families of weighting functions, and two families of time-discount functions. These are briefly discussed below and shown in Table 3.

TABLE 4. Overview function combinations.

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Value	Power	x	x	x	x	x	x						
	Exponential							x	x	x	x	x	x
Weighting	T+K (1992)	x	x					x	x				
	Prelec et al. (1998)			x	x					x	x		
	G+E (1987)					x	x					x	x
Discount	Exponential	x		x		x		x		x		x	
	Quasi hyp.		x		x		x		x		x		x

Notes: This table shows the twelve combinations of the two types of utility functions, three types of probability weighting functions, and two types of time-discount functions.

As common in the literature, we assume that the utility function is composed of a loss aversion coefficient λ and basic utility functions for gains and losses. We focus on the two most common specifications, power utility (forcing the power coefficients of gains and losses to be identical, as is common to obtain an interpretable loss aversion coefficient; e.g., Köbberling and Wakker 2005; Tanaka, Camerer, and Nguyen 2010) and exponential utility. We use three specifications of the probability weighting functions. The same type of function is always used for gains and losses, but we allow the parameters to differ between gains and losses. The first option we consider is the function used by Tversky and Kahneman (1992), with only one parameter. The second and third weighting functions are the two-parameter specifications proposed by Prelec et al. (1998) and Goldstein and Einhorn (1987). For time discounting, we use exponential and quasi-hyperbolic discounting. As displayed in Table 4, we use each possible combination of these functions (there are twelve such combinations).

4.2. Maximum-Likelihood Estimation

Each subject reports a (multi-period) certainty equivalent for a total of 48 lotteries, consisting of 36 calibration and 12 test lotteries. We use the certainty equivalents reported for the calibration lotteries to estimate the model parameters. Consider model Combination C1 with parameters α, λ, γ^+, γ^- and r, where the parameters have the meanings as above (the other model specifications are handled accordingly).

Denote by $V(L_j|\alpha, \lambda, \gamma^+, \gamma^-, r)$ the PT value of the j-th lottery (L_j), with a given parametric specification. The (multi-period) certainty equivalent of the lottery, ce_j, is a certain payout that arises three times (at $t = 0$, at $t = 1$, and at $t = 2$), so that the three payments jointly have a PT value equal to the PT value of the lottery. The PT value of the certainty equivalent is $v(ce_j) + \exp(-r)v(ce_j) + \exp(-2r)v(ce_j)$, independently of which application method of prospect theory is used (the payment is certain, there is thus no probability weighting). Indifference between the certainty equivalent and the

lottery implies that the certainty equivalent of a lottery can be calculated as

$$ce_j(\alpha, \lambda, \gamma^+, \gamma^-, r) = v^{-1}\left(\frac{1}{1+\exp(-r)+\exp(-2r)} V(L_j|\alpha, \lambda, \gamma^+, \gamma^-, r)\right),$$

with v^{-1} denoting the inverse of the utility function.

Following existing literature (e.g., Hey, Morone, and Schmidt 2009; Bruhin, Fehr-Duda, and Epper 2010), we write the observed certainty equivalent of subject i, $CE_{i,j}$, as $CE_{i,j} = ce_j + \varepsilon_{i,j}$ (note that we use capital letters for observed data here). We assume that $\varepsilon_{i,j}$ is normally distributed with mean 0 and standard deviation $\sigma_{i,j}$. We allow for two sources of heteroskedasticity. First, as the employed set of lotteries has different payout ranges, we assume that the error variance is proportional to the payout range of the lottery. The payout range of lottery j, w_j, is calculated as $w_j = |LP_{j,max} - LP_{j,min}|$, where $LP_{j,max}$ denotes the maximal and $LP_{j,min}$ denotes the minimal non-discounted sum of payouts of lottery j. Second, as subjects are heterogeneous, we allow the error variance to differ by subject. This yields the form $\sigma_{i,j} = \varepsilon_i w_j$ for the standard deviation of the error term, where ε_i denotes a subject-specific parameter.

Given our assumptions on the distribution of the error term, the contribution of subject i to the likelihood function L can be expressed as

$$f(\theta, \varepsilon_i|CE_i) = \prod_{j=1}^{36} \frac{1}{\sigma_{i,j}} \varphi\left(\frac{CE_{i,j} - ce_j(\theta)}{\sigma_{i,j}}\right),$$

where the vector $CE_i = (CE_{i,1}, \ldots, CE_{i,36})$ contains the certainty equivalents subject i reports for the 36 calibration lotteries, and the vector $\theta = (\alpha, \lambda, \gamma^+, \gamma^-, r)$ contains the parameters of the model. Furthermore, φ denotes the density of the standard normal distribution. Using data from all n subjects, the log-likelihood function is given by

$$\begin{aligned}\ell(\theta, \varepsilon|CE) &= \log\ L(\theta, \varepsilon|CE) = \sum_{i=1}^{n} \log\ f(\theta, \varepsilon_i|CE_i) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{36} \log\left[\frac{1}{\sigma_{i,j}} \varphi\left(\frac{CE_{i,j} - ce_j(\theta)}{\sigma_{i,j}}\right)\right],\end{aligned}$$

where the vector

$$CE = (CE_1, \ldots, CE_n) = (CE_{1,1}, \ldots, CE_{1,36}, \ldots, CE_{n,1}, \ldots, CE_{n,36})$$

contains the certainty equivalents reported for the calibration lotteries by all n subjects, and the vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ contains the subject-specific error variance parameters.

For the estimation of the parameters, we maximize the log-likelihood function using standard iterative algorithms. With this estimation strategy, we simultaneously estimate the model parameters $(\alpha, \lambda, \gamma^+, \gamma^-, r)$ and the subject-specific error variance parameters $\varepsilon_1, \ldots, \varepsilon_n$.

4.3. Measurement of Prediction Performance on the Test Set

We use the certainty equivalents reported for the twelve test lotteries to measure the (out-of-sample) performance of the two methods. For each of the two methods, we use the model parameters that were estimated on the calibration set (e.g., $\widehat{\alpha}$, $\widehat{\lambda}$, $\widehat{\gamma^+}$, $\widehat{\gamma^-}$, and $\widehat{r}$) to predict choices of the certainty equivalents in the test set,

$$\widehat{ce_j} = v^{-1}\left(\frac{1}{1+\exp(-\widehat{r})+\exp(-2\widehat{r})}V(L_j|\widehat{\alpha},\widehat{\lambda},\widehat{\gamma^+},\widehat{\gamma^-},\widehat{r})\right).$$

As measure of prediction performance, we use the (weighted) mean squared errors. The (weighted) mean squared error of subject i is calculated as

$$MSE_i = \frac{1}{12}\sum_{j=1}^{12}\left(\frac{1}{w_j}\left(CE_{i,j}-\widehat{ce_j}\right)\right)^2,$$

where $w_j = |LP_{j,max} - LP_{j,min}|$, as before. The weighting ensures that the error is not disproportionately driven by lottery choices involving large (absolute) payout amounts. Using data from all n subjects, we define our main outcome variable as

$$MSE = \frac{1}{n}\sum_{i=1}^{n} MSE_i.$$

Out of all 24 combinations of application method and functional specification, we consider the combination best that leads to the lowest MSE.

The predicted certainty equivalents $\widehat{ce_j}$ and thus MSE_i and MSE depend on the choice of the application method. We denote by MSE_{TS} and MSE_{RS} the MSE resulting from the evaluation under the time-separation method and under the risk-separation method, respectively (given a combination of utility, probability weighting, and time-discount functions). We denote the difference between these two values by

$$\triangle MSE := MSE_{TS} - MSE_{RS}.$$

For each of the twelve functional specifications, $\triangle MSE$ indicates whether one application method describes decisions better than the other.

4.4. Statistical Tests, Standard Errors, and Re-Estimation

To test whether the two application methods differ with respect to their prediction performance, we investigate the twelve functional specifications in isolation. For each specification, we test whether the difference between the two (weighted) mean squared errors, that is, $\triangle MSE$, is statistically significantly different from 0. The null hypothesis H_0 is $\triangle MSE = 0$ and the (two-sided) alternative hypothesis H_a is $\triangle MSE \neq 0$. We test this with (non-parametric) paired bootstrap tests. We draw from our sample with replacement to obtain bootstrap samples of the same size as our original sample (resampling is done at the level of subjects). For each bootstrap sample we follow the

procedure described above: We first estimate the parameters on the calibration lotteries and then calculate the two (weighted) mean squared errors on the test lotteries (one for each application method). We denote by $MSE_{b,TS}$ and $MSE_{b,RS}$, the two MSEs that are estimated on bootstrap sample b and by $\triangle MSE_b$ their difference, $\triangle MSE_b = MSE_{b,TS} - MSE_{b,RS}$. We reject the null hypothesis if $\triangle MSE_b \geq 0$ in less than 2.5% of the bootstrap samples or if $\triangle MSE_b \leq 0$ in less than 2.5% of the bootstrap samples. We also use these bootstrapped samples to provide standard errors of the parameter estimates.

In a final step, we pool the data on the calibration and test lotteries (thus making use of the full 48 certainty equivalents per subject) to re-estimate the parameters.

5. Results

We first focus on the main result, which is the comparison of the prediction performance of the two application methods (Section 5.1). Thereafter, we re-estimate the parameters on the pooled training and test set (Section 5.2). Analyses that are pre-registered as additional analyses can be found thereafter in Sections 5.3 and 5.4. In Section 5.3, we analyze which components of prospect theory are essential for good prediction performance and compare the prediction performance to that of EDU. In Section 5.4, we analyze a variant of the risk-separation method. We analyze the data from both treatments jointly, as they are sufficiently similar (according to the criteria defined in the pre-analysis plan; details can be found in Online Appendix E.10).[10] The number of observations for the analyses is 100.

5.1. Time-Separation Method versus Risk-Separation Method

We first estimate the parameters for each of the twelve combinations separately on the calibration lotteries. We then use these parameters to compare the prediction performance of the methods on the test lotteries. The parameter estimates on the calibration set are not very different from those on the whole sample (shown in Section 5.2), therefore we do not discuss them (they can be found in Online Appendix E).

Figure 7 shows the main result. The figure shows the mean squared errors on the test set for both methods, for all twelve combinations of function specifications. It can be seen that the risk-separation method always performs better than the time-separation method. This difference is sizable and highly statistically significant: Figure 8 shows

10. That there is no considerable difference between incentivized and hypothetical choices is in line with the literature (e.g., Abdellaoui, l'Haridon, and Paraschiv 2011; Noussair, Trautmann, and Van de Kuilen 2014; Brañas-Garza et al. 2021; Hackethal et al. 2023).

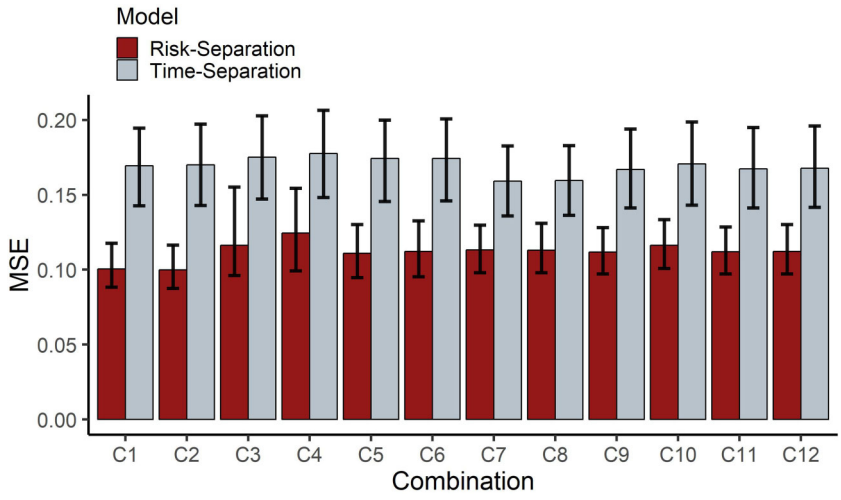


FIGURE 7. MSE for both application methods (all function combinations). This figure shows mean squared errors of the risk-separation and the time-separation method on the test set for all twelve function combinations (with bootstrapped 95% confidence intervals).

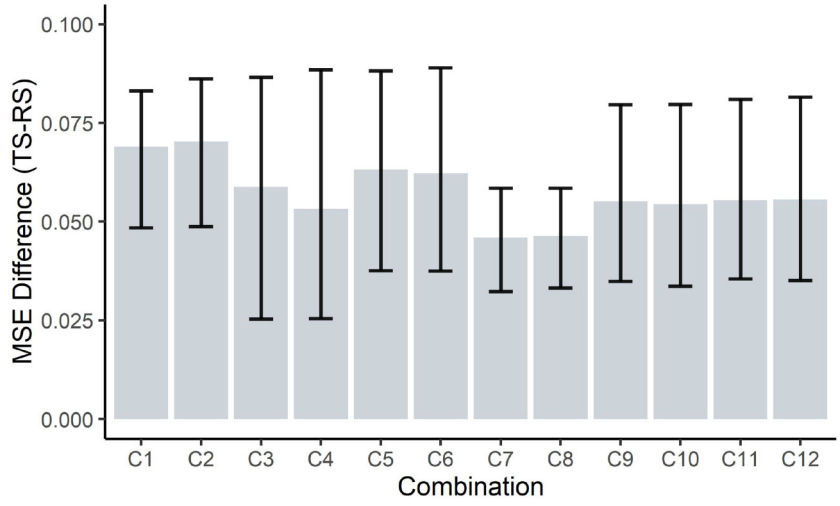


FIGURE 8. Difference in MSE between the application methods (all function combinations). This figure shows the average differences in mean squared errors (MSE of time-separation method minus MSE of risk-separation method) with bootstrapped 95% confidence intervals.

TABLE 5. Mean absolute prediction error (by lottery).

	Low-stake lotteries						High-stake lotteries					
	L7	L8	L15	L16	L23	L24	L31	L32	L39	L40	L47	L48
Payout range	23	20	23	20	30	50	467	400	467	400	600	1000
Mean error time-separation method	6	8.1	7.3	5.6	7.9	16.6	144.7	148.6	146.4	124.6	180.3	319.5
Mean error risk-separation method	5.2	5.6	6.1	5.6	7.2	14.6	108.7	118.1	114.4	115.1	150.7	276.7

Notes: The mean absolute prediction error of Lottery j is calculated as mean($|CE_{i,j} - \widehat{ce}_j|$), with $CE_{i,j}$ denoting the certainty equivalent subject j reported for lottery j and $\widehat{ce}_j$ denoting the predicted certainty equivalent resulting from the parameters estimated on the calibration set.

the difference between the mean squared errors jointly with bootstrapped 95% confidence intervals.[11]

The difference in prediction performance is not driven by a particular subset of the test lotteries. For each of the twelve combinations, the risk-separation method has a lower MSE than the time-separation method for each single test lottery (Table E.2 in the online appendix contains the MSE per lottery for each function combination).

To illustrate the fit of the predictions jointly with the difference in prediction performance, Table 5 shows mean absolute prediction errors (in euros) on a lottery-by-lottery basis for all test lotteries (for the combination of functional forms C1; Table E.1 in the online appendix contains the absolute error per lottery for all twelve combinations of functional forms). Considering the absolute error in relation to the payout range of a lottery, the absolute error for the risk-separation method is in the range of 20%–30% of the payout range for all lotteries.

One can also illustrate the difference by classifying subjects as time-separation or risk-separation types. If the MSE of one subject for one method is greater than the MSE for the other with a difference of at least one standard error (of the MSE difference between the two methods), a subject is classified as a time-separation or a risk-separation type (otherwise, the subject remains unclassified). Table 6 displays the results. For all twelve combinations, the vast majority of subjects are classified as risk-separation type. The share of risk-separation types ranges from 81% to 89%, with 8% to 18% unclassified and 1% to 10% time-separation types. Figure E.1 in the online appendix displays the distribution of the difference between the two MSEs, showing that these are sizable for a majority of the subjects.

Interpreting the results, we believe that there are two main reasons why the risk-separation method predicts choices better. First, separating different outcome streams may be more natural for people than separating different time periods (as

11. The pre-registration states that we also provide a test in which we compare the best performing of the twelve combinations for the risk-separation method with the best performing combination for the time-separation method. This difference is naturally also significant because the performance of the function combinations C1–C12 within one application method is very similar (which can be seen from Figures 7 and 8).
In a few cases the predictions of the time-separation method are so extreme that they fall outside of the slider limits. In these cases we replace the predictions by the slider limits (not adjusting the prediction would lead to an even larger difference in MSE between the two methods).

TABLE 6. Subject types.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Time-separation types	10	5	1	1	1	1	5	5	2	1	2	2
Risk-separation types	81	85	84	81	89	88	87	87	86	89	88	88
Unclassified	9	10	15	18	10	11	8	8	12	10	10	10

Notes: Subject i is classified as time-separation type if $MSE_i^{RS} - SE(\triangle MSE) > MSE_i^{TS}$ or as risk-separation type if $MSE_i^{TS} - SE(\triangle MSE) > MSE_i^{RS}$. $SE(\triangle MSE)$ denotes the standard error of the difference in MSE between the two methods.

argued in Li, Rohde, and Wakker 2024). Separating outcome streams seems relatively straightforward, as the different outcome streams are disjoint probabilistic events (if one outcome stream is realized, no other outcome stream can be realized). Separating time periods seems difficult because these are not independent. To illustrate the complexity of the time-separation method, imagine that a probability of an outcome in the first period of a lottery changes. This change does not only affect probabilities for the first period, but also in general the (compound) probabilities of outcomes in all later periods. Second, while both methods contain compound probabilities, the time-separation method involves more different probabilities than the risk-separation method (often, the compound probabilities involved in the risk-separation method are a subset of the probabilities involved in the time-separation method—those used for the PT value of the last time period only). Both of these reasons have in common that the time-separation method may be considered more complex (and thus less intuitive) than the risk-separation method (for evidence that complexity plays a role in intertemporal choice in a different setting, see Enke, Graeber, and Oprea 2023).[12]

5.2. Parameter Estimates

Here, we present and discuss the parameter estimates from the re-estimation on the full set of lotteries for the risk-separation method (those for the time-separation method can be found in Online Appendix E). Which utility function and which of the two-parameter weighting functions are used have only minor implications for the shapes of the calibrated functions. We therefore only show the parametrizations of the four combinations that employ a power utility and either one-parameter weighting functions, as suggested by Tversky and Kahneman (1992), C1 and C2, or two-parameter weighting functions suggested by Goldstein and Einhorn (1987), C5 and C6.

12. Note that we do not consider it essential that subjects actually compute the compound probabilities. The weighting of probabilities is an irrational and inexact process anyway (it may be considered a misperception or misrepresentation of probabilities). Thus, it will be sufficient for the models to be useful if subjects follow reasoning based on heuristics leading to imprecise representations of the involved probabilities. For instance, in the risk-separation method, subjects may consider outcome streams and assign (consciously or unconsciously) rough estimates of the objective compound probabilities to these outcome streams. In atemporal prospect theory, an imprecise non-mathematical process (probability weighting) is described successfully by a mathematical formula–in intertemporal situations, a slightly more imprecise process (the forming of rough estimates of the probabilities plus their weighting) may similarly be successfully described by a mathematical formula.

TABLE 7. Overview parameter estimates on full sample (48 lotteries).

	Value			Weighting				Discounting	
Combination	α	β	λ	γ^+	γ^-	ν^+	ν^-	r	k
C1	1.195		1.069	0.535	0.552			0.118	
	(0.068)		(0.096)	(0.045)	(0.059)			(0.036)	
C2	1.172		1.065	0.548	0.565			0.011	0.863
	(0.060)		(0.113)	(0.042)	(0.057)			(0.025)	(0.042)
C5	1.091		1.037	0.426	0.448	0.766	0.785	0.070	
	(0.035)		(0.124)	(0.066)	(0.073)	(0.072)	(0.070)	(0.028)	
C6	1.085		1.101	0.424	0.451	0.786	0.796	0.001	0.885
	(0.029)		(0.163)	(0.069)	(0.073)	(0.073)	(0.068)	(0.010)	(0.042)

Notes: This table shows the parameter estimates on the full set of lotteries (calibration plus test set) for the risk-separation method, for four selected function combinations. Standard errors are shown in parentheses. Estimates for all twelve function combinations and for the time-separation method can be found in Online Appendix E.

C1 and C5 use exponential time-discounting, C2 and C6 quasi-hyperbolic discounting. The estimates are shown in Table 7. The estimates for all twelve model combinations can be found in Online Appendix E.

Utility Function. The utility functions consist of basic utility functions for gains and losses and a loss aversion parameter (λ). The basic utility functions are in general close to linear, with utility in the gain domain mostly slightly convex and in the loss domain slightly concave (for the four combinations C9 to C12 that employ an exponential utility and two-parameter probability weighting functions, the estimated utility is slightly concave for gains and convex for losses, see the full Table E.4 in the online appendix). In the atemporal literature, the utility for gains is generally found to be concave, and that for losses to be convex or linear. However, similar shapes have already been observed in the atemporal literature.[13] The second component of the utility function is the loss aversion parameter λ. The loss-aversion parameters are on average only slightly above 1.[14] This is lower than what is usually observed in the atemporal literature.[15]

13. Bruhin, Fehr-Duda, and Epper (2010), for instance, calibrate a combination of Goldstein–Einhorn weighting functions and power utility, as our C5 and C6, with very similar estimates of weighting and utility functions, including basic utility that is convex for gains and concave for losses.

14. A potential explanation for such a low loss-aversion coefficient could be that subjects may aggregate losses and gains rather than evaluating these in isolation. Suppose, for instance, that an outcome stream yields 60 at $t = 1$ and -40 at $t = 2$. If subjects aggregate gains and losses, the evaluation of the stream should (neglecting time discounting) be close to the evaluation of a payout of 20 at $t = 1$ and no payout at $t = 2$. The risk-separation method (as also the time-separation method), however, transforms each outcome to a utility term before aggregating. A version of the risk-separation method (pre-registered as additional analysis) calculating present values (in monetary terms) before entering them in a utility function is treated below in Section 5.4. Estimates of the utility function with that version are similar to those obtained with the regular risk-separation method (loss aversion coefficients are only slightly higher), so that this explanation does not seem to be the main driver of such low loss-aversion coefficient.

15. According to a recent meta study (Brown et al. 2024), the median loss-aversion parameter among 286 aggregate level studies is about 1.5. However, the few studies that estimate loss aversion in an intertemporal

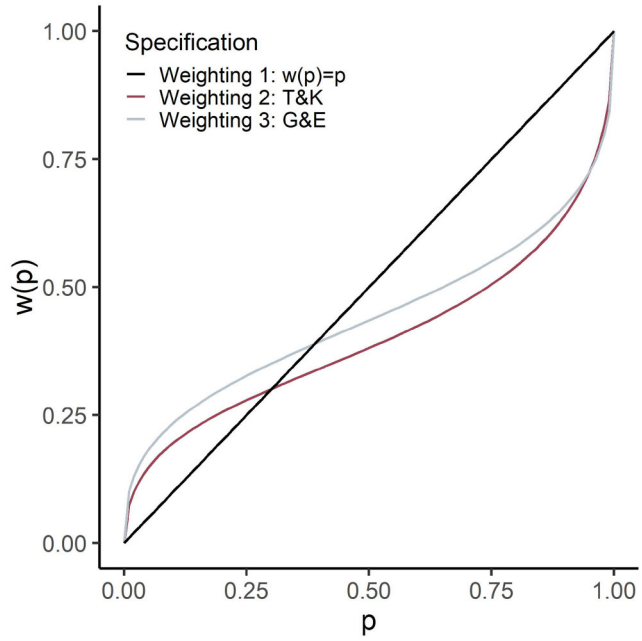


FIGURE 9. Estimated probability weighting functions. This figure shows the estimated weighting functions for gains for function combinations C1 and C5 (for the risk-separation method). C1 and C5 both use a power utility function and exponential discounting.

Probability Weighting Functions. The calibrations of the probability weighting functions turn out to be very similar for gains and for losses for all function combinations, suggesting that one can use the same functions in the gain and in the loss domain (reducing the number of model parameters). For all specifications, an inverse S-shape is found, as common in the atemporal literature. The shapes of the weighting functions are also similar across the different specifications, with the two-parameter weighting functions ascending faster for low values than the one-parametric Tversky–Kahneman specification. Figure 9 displays the estimated weighting functions for gains for model combinations C1 and C5 as illustrations. In general, our estimates of probability weighting functions are very similar to those estimated in atemporal contexts (e.g., Booij, Van Praag, and Van De Kuilen 2010; Bruhin, Fehr-Duda, and Epper 2010).

Time-Discount Function. Exponential discounting and quasi-hyperbolic discounting lead to very similar prediction performances in our setting. The estimates for exponential discounting reach from zero to about 12% per quarter. Estimates for

context also find loss-aversion coefficients closer to 1 (Abdellaoui et al. 2013a;Abdellaoui, l'Haridon, and Paraschiv 2013b). Chapman et al. (2025) find that there is less loss aversion among the general population than among people with high cognitive ability (such as the typical student subject pools).

quasi-hyperbolic discounting reach from no discounting at all to discounting all future payments by about 15% (with very small additional exponential discounting). These discounting patterns are similar to those found in riskless intertemporal decision tasks, as reported in the recent meta study by Matousek, Havranek, and Irsova (2022).

Parameter Calibrations for Applications. In applications in which one is only interested in predictive power (possibly favoring a combination with few parameters), one could use the following specification of a power utility and Tversky–Kahneman probability weighting functions for gains and losses, combined with an exponential discounting of 12% per quarter:

$$v(x) = \mathbb{1}(x \geq 0)x^{1.20} - \mathbb{1}(x < 0)1.07(-x)^{1.20},$$
$$\text{and} \quad w^+(p) = w^-(p) = \frac{p^{0.54}}{\left(p^{0.54} + (1-p)^{0.54}\right)^{1/0.54}}$$

(labeled combination C1 in Table 4).[16] As alternative, if one prefers to use a two-parametric probability weighting function to allow for more flexibility (for theoretical arguments in favor of a two-parametric function, see Fehr-Duda and Epper 2012), one may want to use the following specifications of a power utility and Goldstein–Einhorn weighting functions, combined with exponential discounting of 7% per quarter:

$$v(x) = \mathbb{1}(x \geq 0)x^{1.09} - \mathbb{1}(x < 0)1.04(-x)^{1.09},$$
$$\text{and} \quad w^+(p) = w^-(p) = \frac{0.78p^{0.44}}{0.78p^{0.44} + (1-p)^{0.44}}$$

(C5). This specification has a prediction performance that is almost as good as the specification mentioned earlier, and the parameter values seem intuitively more natural and are more closely aligned with the atemporal literature (they are, for instance, very similar to the estimates of Bruhin, Fehr-Duda, and Epper 2010).

5.3. Relative Importance of the Components of Prospect Theory

Here, we analyze how the prediction performance deteriorates when the probability weighting functions are assumed to be linear, when there is no loss aversion, or when the utility functions are linear (with all free parameters estimated anew). We also compare the prediction performance to that of EDU maximization. We only show the

16. This is the combination with the best out-of-sample explanatory power. It is also the combination with the lowest number of parameters. The calibration stems from the whole set of lotteries—in the exact calibration (Table 7), the parameters for the weighting functions for gains and losses are 0.535 and 0.552, respectively. As these values are so close, we believe that it is reasonable to use the same parameter in the gain and loss domains (the argument for using identical specifications of weighting functions for gains and losses also applies to other specifications).

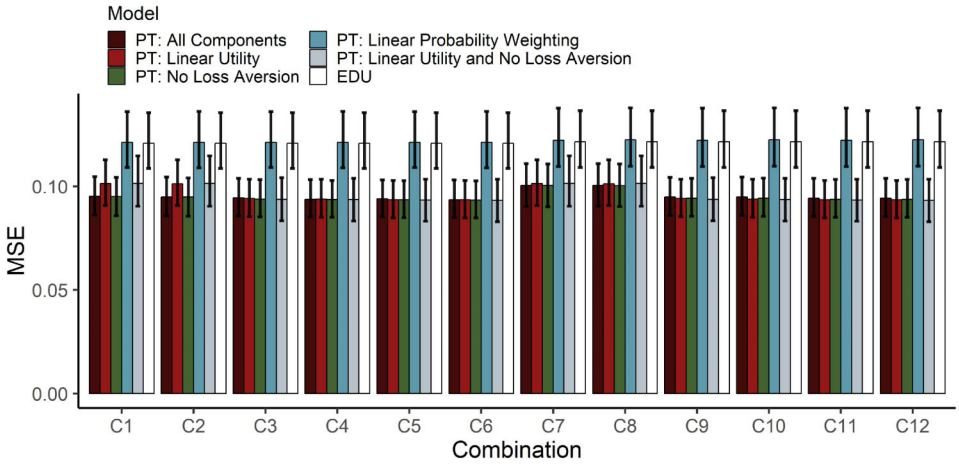


FIGURE 10. MSE risk-separation method, PT components (all function combinations). This figure compares the prediction performance of the risk-separation method to versions of it with some components absent and to expected utility theory (parameters are estimated anew for each method). The versions are (from left to right for each of the twelve function combinations) the risk-separation method with all components ("PT: All Components"), with linear utility for gains and losses still including a loss-aversion parameter ("PT: Linear Utility"), with a loss-aversion parameter of 1 ("PT: No Loss Aversion"), with linear-probability weighting ("PT: Linear Probability Weighting"), with linear utility for gains and losses and a loss-aversion parameter of 1 ("PT: Linear Utility and No Loss Aversion"); on the right, expected discounted utility is shown ("EDU").

results for the risk-separation method, an investigation for the time-separation method can be found in Online Appendix E.7.[17]

The results are shown in Figure 10. The prediction performances of models with linear utility for gains and losses ("PT: Linear Utility"; this model allows for a loss aversion coefficient different from one) or no loss aversion ("PT: No Loss Aversion") are almost identical to the performance of the full PT application ("PT: All Components"). The same holds if utility is not only required to be linear for gains

17. The analyses in Sections 5.3 and 5.4 are described as additional analyses in the pre-analysis plan (with little detail; the analysis is kept as similar to the main analysis as possible). The core design feature of our calibration set is that the calibrations are the same for time-separation and risk-separation method. In addition, the test lotteries are such that the two methods over- or undervalue the lotteries to a similar extent with typical atemporal calibrations. These features ensure that our design does not favor one of the two methods. Both features, however, do not hold when also considering other methods (including EDU). Therefore, the comparison here differs as follows.
We randomly select six of the eight lotteries from each set, implying a total of 36 lotteries, as the new calibration lotteries. We repeat this random selection process ten times. For each of these ten repetitions the 36 calibration lotteries are used to calibrate a models' parameters. The other twelve lotteries (two per set) are used as test lotteries to compare the predictive performance, measured by the (weighted) MSE. The numbers of calibration and test lotteries are therefore the same as in the main analysis. We denote by MSE_{S_i} the (weighted) MSE of the random selection S_i, $i \leq 10$. The main outcome variable is the mean of these ten values, $MSE = \frac{1}{10}\sum_{i=1}^{10} MSE_{S_i}$. Bootstrapped confidence intervals for the MSE are obtained as in the main analyses.

and losses, but on the whole domain, including the absence of loss aversion ("PT: Linear Utility and No Loss Aversion"). This implies that utility curvature and loss aversion can be omitted if a more parsimonious model is desired. However, forcing linear-probability weighting ("PT: Linear Probability Weighting") predicts decisions significantly worse.[18] These findings are in line with the parameter estimates discussed in Section 5.2, which exhibit almost linear utility with a loss aversion parameter close to 1 and highly non-linear probability weighting functions.

We also compare the prediction performance of the risk-separation method to that of standard EDU models. Figure 10 shows that EDU predicts decisions significantly and considerably worse than the risk-separation version of prospect theory, with MSEs that are about 25% higher (for all combinations of functional forms, except for the combinations C7 and C8, where the MSE is about 20% higher for EDU). The prediction performance of the time-separation method turns out to be almost identical to that of EDU (for details, see Figure E.2 in the online appendix).

5.4. Risk-Separation Present-Value Method

In addition to the two main methods, we also analyze a variant of the risk-separation method (as described in the pre-analysis plan), which we call risk-separation present-value method. Whereas the regular risk-separation method calculates the discounted utility of an outcome stream (before probability weighting plays a role) as $PV(o_j) = \sum_{t=0}^{T-1} \delta(t) v(o_{j,t})$, the risk-separation present-value method calculates the corresponding term as $PV(o_j) = v\left(\sum_{t=0}^{T-1} \delta(t) o_{j,t}\right)$. That is, instead of calculating the term by discounting the utility obtained each period, it first calculates a present value in monetary terms and then transforms this into utility (this happens again before probability weighting plays a role).[19]

To illustrate how these methods can differ with a square root utility function (for gains and ignoring time discounting), consider the following outcome streams in monetary terms: first, a stream with 9 monetary units in each of the three time periods; second, a stream with 81 monetary units in the first period and 0 in the other two periods. The regular risk-separation method assigns a utility value of 9 to both streams (calculated as $\sqrt{9} + \sqrt{9} + \sqrt{9}$ for one stream and as $\sqrt{81} + \sqrt{0} + \sqrt{0}$ for the other stream). The risk-separation present-value method assigns a utility value of $\sqrt{27}$ to the first stream (calculated as $\sqrt{9 + 9 + 9}$) and a utility value of $\sqrt{81} = 9$ to the second stream (calculated as $\sqrt{81 + 0 + 0}$).

18. Several of the MSE bars in Figure 10 are identical across the function combinations. For instance, when utility is assumed to be linear, there is no more distinction between power utility (C1–C6) and exponential utility (C7–C12). For EDU, there are even only two specifications: one with power utility (C1–C6) and one with exponential utility (C7–C12), as there is neither probability weighting nor hyperbolic discounting.

19. The "time-first" specification in Rohde and Yu (2024) can be restricted to become the regular risk-separation method or the risk-separation present-value method. If there is no time discounting, the risk-separation present-value method corresponds to the prospect theory applications in Barberis (2012), Ebert and Strack (2015), and Heimer et al. (2021).

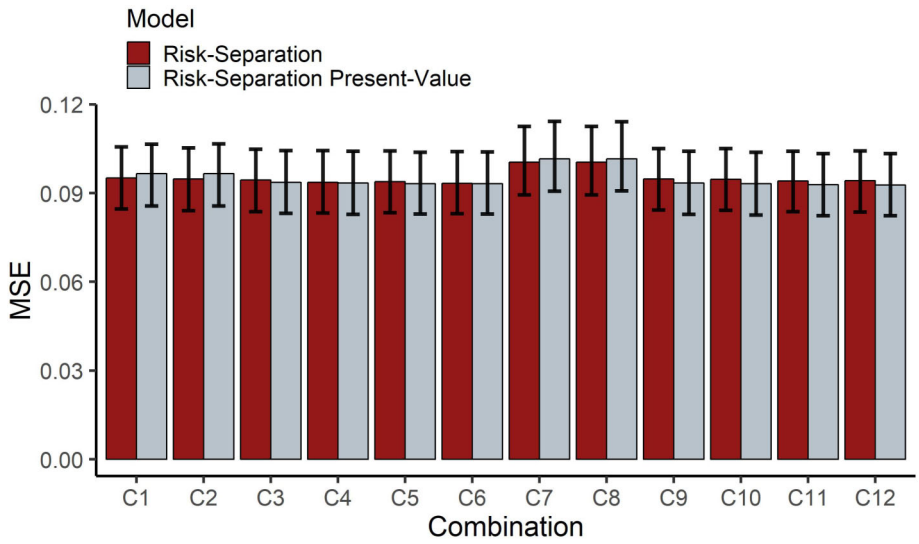


FIGURE 11. MSE for the risk-separation method and the risk-separation present-value method. This figure shows mean squared errors of the risk-separation present-value and the (regular) risk-separation method on the test set for all twelve function combinations (with bootstrapped 95% confidence intervals).

The risk-separation present-value method seems natural because the outcomes of the lotteries are of a monetary nature, and money can be transferred from one period to another. The time discounting allows for the value of monetary payments to be different from period to period, but there is no motive for the decision maker to smooth payouts across periods (as would be the case with concave per-period utility functions in the regular risk-separation method). Therefore, we find the risk-separation present-value method intuitive for settings with financial outcomes; in contrast, we find the regular risk-separation method more intuitive if the outcomes are not of a monetary nature but final consumption, which cannot be transferred from one period to another.

Figure 11 shows the predictive power of the risk-separation present-value method and the regular risk-separation method.[20] It can be seen that the predictive power of the two risk-separation methods is virtually identical.

Estimated Functions. We also estimate the functions for the risk-separation present-value method on the full set of lotteries. As the choice of the utility and two-parameter probability weighting functions has again only minor implications, we only show combinations C1, C2, C5, and C6 in Table 8, as before (for estimates of all combinations, see Online Appendix E).

20. The comparison is again based on ten random draws of calibration and test lotteries, as in Section 5.3 (the degrees of freedom are always identical for a given combination of functional forms).

TABLE 8. Calibration risk-separation present-value method.

	Value			Weighting				Discounting	
Combination	α	β	λ	γ^+	γ^-	ν^+	ν^-	r	k
C1	1.179		1.134	0.579	0.592			0	
	(0.272)		(0.326)	(0.085)	(0.094)			(0)	
C2	1.193		1.622	0.577	0.577			0	1
	(0.273)		(0.327)	(0.085)	(0.094)			(0)	(0.002)
C5	0.916		1.034	0.404	0.406	1.037	1.05	0.042	
	(0.069)		(0.228)	(0.096)	(0.105)	(0.109)	(0.118)	(0.026)	
C6	0.915		1.087	0.407	0.422	1.03	1.049	0.002	0.939
	(0.068)		(0.216)	(0.098)	(0.106)	(0.106)	(0.119)	(0.008)	(0.037)

Notes: This table shows the parameter estimates on the full set of lotteries (calibration plus test set) for the risk-separation present-value method, for four selected function combinations. Estimates for all twelve function combinations can be found in Online Appendix E.

The estimated parameters are overall similar to those of the regular risk-separation method (Table 7). There are some differences, however. The loss-aversion coefficient is, on average, slightly higher for the risk-separation present-value method. This makes intuitively sense if one considers loss aversion to be present at the "overall" level rather than in each period: the loss-aversion coefficient in the risk-separation present-value method captures aversion to losses when the payouts over the periods are summed up (after discounting), while the loss-aversion coefficient in the regular risk-separation method captures aversion to losses within each period. While the utility is still mildly convex for gains and mildly concave for losses for the combinations that employ the one-parameter weighting function, the utility is now mildly concave for gains and mildly convex for losses for combinations that employ a two-parameter weighting function. Estimated utility functions are still close to linear, as for the regular risk-separation method. This explains the similarity of the prediction performance of the two risk-separation methods: for linear utility (with a loss-aversion coefficient of 1), both risk-separation methods give identical results.

For all combinations, the curvature parameters of the probability weighting functions (γ^+ and γ^-) are almost identical to those estimated for the regular risk-separation method. All functions, thus, again have an inverse-S shape.

Estimates of time discounting are a bit lower for the risk-separation present-value method than for the regular one. For exponential discounting, estimated discounting per quarter reaches from 0% to about 7%. For quasi-hyperbolic discounting, all future outcomes are estimated to be discounted by 0% to about 12% (with very little additional exponential discounting).

Parameter Calibration for Applications. For applications of this method, one could use the following specification of a power utility and Goldstein–Einhorn weighting functions, combined with exponential discounting of 4% per quarter:

$$v(x) = \mathbb{1}(x \geq 0)x^{0.92} - \mathbb{1}(x < 0)1.03(-x)^{0.92},$$
$$\text{and} \quad w^+(p) = w^-(p) = \frac{1.04p^{0.40}}{1.04p^{0.40} + (1-p)^{0.40}}$$

(C5). This specification has very high explanatory power and intuitive parameter estimates.

6. Sketch of an Application Example

Which method is used in applications cannot only lead to minor differences in quantities, but the two methods may even lead to opposite conclusions. Assume, for example, an insurable unfavorable event (a loss) that may occur in any of several time periods. The event occurs with relatively small probability in any single time period, but the probability that the event occurs at one point in time is large. The time-separation method would then predict overinsurance because the small probability of an event in any single time period would be overweighted. The risk-separation method would predict underinsurance because the joint probability of outcome streams with an event at some point in time is large and would be underweighted.

An example of such a situation is, for instance, (private) long-term care insurance. Long-term care is one of the greatest financial risks that people (especially the elderly) face in many countries, including the U.S. (e.g., Brown and Finkelstein 2011). Between a third and half of the elderly enter a nursing home at one point in their life (while the probability for this to happen is relatively small in any single year). Given the size of the costs (staying in a private room in a nursing home for the average duration of about two years in the U.S. costs about $200,000), there is a case for risk pooling. However, only between 10% and 20% of the elderly are covered by private long-term care insurance. There is thus underinsurance in this market (see Ameriks et al. 2016; Brown and Finkelstein 2011). This underinsurance is in line with people making decisions with the risk-separation method, but it stands in contrast to people making decisions with the time-separation method.

7. Concluding Remarks

Our pre-registered experiment on a representative sample of the Dutch population tests the performance of applications of prospect theory in intertemporal settings. We find that the risk-separation method (aggregating first over time) performs much better than the time-separation method (aggregating first over risk). The risk-separation method also performs much better than EDU, which has a prediction performance that is almost identical to that of the time-separation method (with mean squared errors about 25% higher than for the risk-separation method).

We also estimate utility, probability weighting, and time-discount functions. We find very low levels of loss aversion and almost linear utility functions. Probability weighting functions in our study are as in the typical atemporal settings. While the choice of risk-separation method or time-separation method has a sizable impact on the explanatory power, the choice of the functional forms of utility, probability weighting, and time-discounting functions has only a negligible effect on the explanatory power. We provide parameter estimates for twelve combinations of functional forms.

We also compare these two application methods to a variant of the risk-separation method, the risk-separation present-value method. This variant calculates present values in monetary terms before assigning them a utility value, instead of aggregating utility terms from different periods. We consider this a natural way of applying prospect theory when the payouts are of a financial nature, as money can be transferred from one period to another. This variant performs as well as the regular risk-separation method.

References

Abdellaoui, Mohammed, Han Bleichrodt, Olivier l'Haridon, and Corina Paraschiv (2013a). "Is There One Unifying Concept of Utility? An Experimental Comparison of Utility under Risk and Utility over Time." *Management Science*, 59, 2153–2169.

Abdellaoui, Mohammed, Olivier l'Haridon, and Corina Paraschiv (2011). "Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications?" *Management Science*, 57, 1879–1895.

Abdellaoui, Mohammed, Olivier l'Haridon, and Corinar Paraschiv (2013b). "Sign Dependence in Intertemporal Choice." *Journal of Risk and Uncertainty*, 47, 225–253.

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, Christopher Tonetti et al. (2016). "Late-in-Life Risks and the under-Insurance Puzzle." NBER Working Paper No. 22726.

Andreoni, James and Charles Sprenger (2012). "Risk Preferences Are Not Time Preferences." *American Economic Review*, 102(7), 3357–76.

Barberis, Nicholas (2012). "A Model of Casino Gambling." *Management Science*, 58, 35–51.

Barberis, Nicholas and Ming Huang (2006). "The Loss Aversion/Narrow Framing Approach to the Equity Premium Puzzle." NBER Working Paper No. 12378.

Barberis, Nicholas and Wei Xiong (2009). "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation." *The Journal of Finance*, 64, 751–784.

Barberis, Nicholas C. (2013). "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives*, 27, 173–96.

Benartzi, Shlomo and Richard H. Thaler (1995). "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics*, 110, 73–92.

Blavatskyy, Pavlo R. (2016). "A Monotone Model of Intertemporal Choice." *Economic Theory*, 62, 785–812.

Bleichrodt, Han and Louis Eeckhoudt (2006). "Survival Risks, Intertemporal Consumption, and Insurance: The Case of Distorted Probabilities." *Insurance: Mathematics and Economics*, 38, 335–346.

Booij, Adam S., Bernard M. S. Van Praag, and Gijs Van De Kuilen (2010). "A Parametric Analysis of Prospect Theory's Functionals for the General Population." *Theory and Decision*, 68, 115–148.

Brañas-Garza, Pablo, Lorenzo Estepa-Mohedano, Diego Jorrat, Victor Orozco, and Ericka Rascón-Ramírez (2021). "To Pay or Not to Pay: Measuring Risk Preferences in Lab and Field." *Judgment and Decision Making*, 16, 1290–1313.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer (2024). "Meta-Analysis of Empirical Estimates of Loss Aversion." *Journal of Economic Literature*, 62, 485–516.

Brown, Jeffrey R. and Amy Finkelstein (2011). "Insuring Long-Term Care in the United States." *Journal of Economic Perspectives*, 25, 119–42.

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica*, 78, 1375–1412.

Chapman, Jonathan, Erik Snowberg, Stephanie W. Wang, and Colin Camerer (2025). "Looming Large or Seeming Small? Attitudes towards Losses in a Representative Sample." *The Review of Economic Studies*, 92, 2893–2922.

Chew, Soo Hong and Larry G. Epstein (1990). "Nonexpected Utility Preferences in a Temporal Framework with an Application to Consumption-Savings Behaviour." *Journal of Economic Theory*, 50, 54–81.
De Giorgi, Enrico G. and Shane Legg (2012). "Dynamic Portfolio Choice and Asset Pricing with Narrow Framing and Probability Weighting." *Journal of Economic Dynamics and Control*, 36, 951–972.
DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47, 315–72.
Dimmock, Stephen G. and Roy Kouwenberg (2010). "Loss-Aversion and Household Portfolio Choice." *Journal of Empirical Finance*, 17, 441–459.
Drouhin, Nicolas (2015). "A Rank-Dependent Utility Model of Uncertain Lifetime." *Journal of Economic Dynamics and Control*, 53, 208–224.
Easley, David and Liyan Yang (2015). "Loss Aversion, Survival and Asset Prices." *Journal of Economic Theory*, 160, 494–516.
Ebert, Sebastian and Philipp Strack (2015). "Until the Bitter End: On Prospect Theory in a Dynamic Context." *American Economic Review*, 105(4), 1618–1633.
Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2023). "Complexity and Time." CESifo Working Paper No. 10327.
Epper, Thomas and Helga Fehr-Duda (2015). "Risk Preferences are not Time Preferences: Balancing on a Budget Line: Comment." *American Economic Review*, 105(7), 2261–71.
Epstein, Larry G. and Stanley E. Zin (1989). "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica*, 57, 937–969.
Fehr-Duda, Helga and Thomas Epper (2012). "Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences." *Annual Review of Economics*, 4, 567–593.
Goette, Lorenz, David Huffman, and Ernst Fehr (2004). "Loss Aversion and Labor Supply." *Journal of the European Economic Association*, 2, 216–228.
Goldstein, William M. and Hillel J. Einhorn (1987). "Expression Theory and the Preference Reversal Phenomena." *Psychological Review*, 94, 236.
Guiso, Luigi and Monica Paiella (2008). "Risk Aversion, Wealth, and Background Risk." *Journal of the European Economic Association*, 6, 1109–1150.
Guo, Jing and Xue Dong He (2017). "Equilibrium Asset Pricing with Epstein-Zin and Loss-Averse Investors." *Journal of Economic Dynamics and Control*, 76, 86–108.
Hackethal, Andreas, Michael Kirchler, Christine Laudenbach, Michael Razen, and Annika Weber (2023). "On the Role of Monetary Incentives in Risk Preference Elicitation Experiments." *Journal of Risk and Uncertainty*, 66, 189–213.
Halevy, Yoram (2008). "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect." *American Economic Review*, 98(3), 1145–62.
Heimer, Rawley, Zwetelina Iliewa, Alex Imas, and Martin Weber (2021). "Dynamic Inconsistency in Risky Choice: Evidence from the Lab and Field." ECONtribute Discussion Paper No. 094.
Hey, John D., Andrea Morone, and Ulrich Schmidt (2009). "Noise and Bias in Eliciting Preferences." *Journal of Risk and Uncertainty*, 39, 213–235.
Köbberling, Veronika and Peter P. Wakker (2005). "An Index of Loss Aversion." *Journal of Economic Theory*, 122, 119–131.
Krause, Jan, Patrick Ring, and Ulrich Schmidt (2020). "Hyperopic Loss Aversion." SSRN Working Paper No. 3582050.
Kreps, David M. and Evan L. Porteus (1978). "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica*, 46, 185–200.
Lampe, Immanuel and Daniel Würtenberger (2020). "Loss Aversion and the Demand for Index Insurance." *Journal of Economic Behavior & Organization*, 180, 678–693.
Li, Chen, Kirsten I. M. Rohde, and Peter P. Wakker (2024). "The Deceptive Beauty of Monotonicity, and the Million-Dollar Question: Row-First or Column-First Aggregation?" Rotterdam: Working Paper, Erasmus University Rotterdam.

Matousek, Jindrich, Tomas Havranek, and Zuzana Irsova (2022). "Individual Discount Rates: A Meta-Analysis of Experimental Evidence." *Experimental Economics*, 25, 318–358.
Meng, Juanjuan and Xi Weng (2017). "Can Prospect Theory Explain the Disposition Effect? A New Perspective on Reference Points." *Management Science*, 64, 3331–3351.
Noussair, Charles N., Stefan T. Trautmann, and Gijs Van de Kuilen (2014). "Higher Order Risk Attitudes, Demographics, and Financial Decisions." *Review of Economic Studies*, 81, 325–355.
Prelec, Drazen et al. (1998). "The Probability Weighting Function." *Econometrica*, 66, 497–528.
Rohde, Kirsten I. M. and Xiao Yu (2024). "Intertemporal Correlation Aversion-a Model-Free Measurement." *Management Science*, 70, 3493–3509.
Scherpenzeel, Annette C. and Marcel Das (2010). "'True' Longitudinal and Probability-Based Internet Panels: Evidence From the Netherlands."In *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, edited by Marcel Das, Peter Ester, and Lars Kaczmirek. Taylor & Francis, pp. 77–104.
Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen (2010). "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review*, 100(1), 557–71.
Tversky, Amos and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5, 297–323.
Umer, Hamza (2023). "Effectiveness of Random Payment in Experiments: A Meta-Analysis of Dictator Games." *Journal of Economic Psychology*, 96, 102608.
van Bilsen, Servaas and Roger J. A. Laeven (2020). "Dynamic Consumption and Portfolio Choice under Prospect Theory." *Insurance: Mathematics and Economics*, 91, 224–237.
van Bilsen, Servaas, Roger J. A. Laeven, and Theo E. Nijman (2020). "Consumption and Portfolio Choice under Loss Aversion and Endogenous Updating of the Reference Level." *Management Science*, 66, 3799–4358.
Wakker, Peter, Richard Thaler, and Amos Tversky (1997). "Probabilistic Insurance." *Journal of Risk and Uncertainty*, 15, 7–28.

Supplementary Data

Supplementary data are available at *JEEA* online.

Journal of the European Economic Association 2026 24(2):701–735
https://doi.org/10.1093/jeea/jvaf037