Theory Dec. (2012) 72:359–386
DOI 10.1007/s11238-010-9234-3

Measuring the time stability of Prospect Theory preferences

Stefan Zeisberger · Dennis Vrecko · Thomas Langer

Published online: 26 November 2010
© Springer Science+Business Media, LLC. 2010

Abstract Prospect Theory (PT) is widely regarded as the most promising descriptive model for decision making under uncertainty. Various tests have corroborated the validity of the characteristic fourfold pattern of risk attitudes implied by the combination of probability weighting and value transformation. But is it also safe to assume stable PT preferences at the individual level? This is not only an empirical but also a conceptual question. Measuring the stability of preferences in a multi-parameter decision model such as PT is far more complex than evaluating single-parameter models such as Expected Utility Theory under the assumption of constant relative risk aversion. There exist considerable interdependencies among parameters such that allegedly diverging parameter combinations could in fact produce very similar preference structures. In this paper, we provide a theoretic framework for measuring the (temporal) stability of PT parameters. To illustrate our methodology, we further apply our approach to 86 subjects for whom we elicit PT parameters twice, with a time lag of 1 month. While documenting remarkable stability of parameter estimates at the aggregate level, we find that a third of the subjects show significant instability across sessions.

Keywords Prospect Theory · Time stability · Risk preferences · Experimental economics

JEL Classification D81 · C91

1 Introduction

The question of whether individual risk preferences are a stable individual trait is important for many economic decisions. The elicitation of individual risk preference

S. Zeisberger (✉) · D. Vrecko · T. Langer
Finance Center Münster, University of Münster, Universitätsstraße 14–16, 48143 Münster, Germany
e-mail: stefan.zeisberger@uzh.ch

parameters and their utilization for predictive or prescriptive purposes largely rests on the assumption that risk preferences are stable over time. A near-at-hand method for analyzing issues of temporal stability is to ask individuals to provide certainty equivalents (CEs) for the same lottery at different points in time. If the answers are interpreted in an Expected Utility Theory (EUT) framework and constant relative risk aversion (CRRA) is furthermore assumed, the analysis of preference stability is straightforward. From each CE an Arrow–Pratt coefficient of risk aversion can be determined to provide simple indicators of risk attitude and to address the question of temporal stability. If the problem is approached from a more descriptive point of view and the stability of risk attitudes is discussed in the light of Cumulative Prospect Theory,[1] things become trickier. In Prospect Theory (PT), widely regarded to be the most promising descriptive theory of decision making under uncertainty, risk attitudes follow from the shape of the value function as well as the shape of the probability weighting function and their interaction. A single CE can result from an infinite set of curvature and probability weighting parameter combinations. Assuming one-parameter functional specifications for value and weighting function, at least two lotteries are required to separate between curvature and probability weighting. Importantly, considerable interdependencies among parameters remain even for a large set of carefully chosen lotteries, and minor changes in CEs might lead to sizeable differences in the point estimates of the parameters. Against this background it seems important to understand whether different parameter combinations resulting from estimation sessions at different points in time actually represent similar or very distinct risk attitudes.

The central goal of this paper is to demonstrate the interaction effects among PT parameters and to develop a methodological framework for analyzing the temporal stability of risk attitudes under the notion of PT. The framework we introduce sorts out the measurement problems associated with CE stimuli by explicitly accounting for the interaction effects among parameters. To illustrate our methodology, we apply the procedure in an experimental study with 86 subjects and present initial evidence on the temporal stability of PT parameter estimations.

The flourishing literature on time stability scrutinizes to what extent individuals respond to risk in an intertemporally consistent manner.[2] To our knowledge, McGlothlin (1956) is the first paper on preference stability over time. He analyzes the stability of horse race bettors' aggregate behavior over the course of a racing day and finds a persistent overweighting of small probabilities. Grayson (1960) is the first to analyze temporal stability at the individual level. He elicits the utility curves of the owner of an oil exploration company before and after a 3-month interval, and finds greater risk aversion in the later curve. However, the individual was able to provide business-related reasons for his temporal instability (change in risk attitude) at

[1] See Wakker (2010) for a comprehensive overview on (cumulative) Prospect Theory. In the remainder of this text we will omit the "Cumulative" in front of Prospect Theory.

[2] In order to differentiate between time stability and regular/internal consistency in individual decision-making the test-retest interval is the deciding factor. Studies of time stability use a time lag of at least 1 month (usually 1–6 months) between the first and second run of the elicitation procedure. In contrast, assessments of a subject's regular consistency in decision-making work with shorter test-retest intervals, usually less than one week (we refer to Stott (2006) for an overview). Assessments of the internal consistency of decisions work without any time delay.

the end of the second interview session. Swalm (1966) interviews a group of business executives and elicits CEs of hypothetical gambles. He observes a "surprising stability" of the resulting corporate utility curves based on unreported data which he had collected.

But how does research on time stability fit into the broader picture? To identify related fields of research, another look at the history of the field proves to be useful. The starting point for today's work on preference stability, the early psychological literature on intertask consistency, is discussed by Slovic (1972). He observes a rapid decline in correlations between risk-taking measures (none of them utility-based) as their structural similarity decreases. The literature on preference stability strives to identify the determinants of this, at a first glance, unsatisfactory finding. A large part of the recent literature can be categorized in two dimensions: stability type (stability across domains, stability across elicitation methods, and stability over time) and measure/theory (survey questions, EUT, and PT). As can be seen from Table 1 there is—up to now—almost a complete lack of literature on time stability under PT. The only paper we know of that looks at time stability from a PT angle is the recent work by Baucells and Villasís (2010). Their analysis is, however, restricted to the question whether the reflection effect can be observed in simple lottery choices and whether this effect is stable over time (more about this paper in Sect. 2). The lack of more general examinations of the stability of PT preferences is not overly surprising. The analysis turns out to be complicated from a conceptual point of view. With our work we try to shed light on some of the issues.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature on preference stability over time. Section 3 is the most relevant part as it elaborates on the mechanics of the interaction effects within PT preferences and thus provides the theoretical background for the experimental analysis. Section 4 presents the design of the experimental study that we have run to illustrate the newly developed methodology of PT time stability analysis. In Sect. 5, we investigate the time stability of PT risk preferences based on the experimental data. Section 6 concludes with a discussion of practical implications and highlights areas of future research.

2 Literature on time stability

We restrict our literature review to the "stability over time" row of Table 1. First, we will consider studies using survey questions, move on to EUT, and finish with PT-related work on time stability. There has been a recent trend to investigate questions of long-term preference stability with panel-study data from the field (Brunnermeier and Nagel 2008; Malmendier and Nagel 2010). This approach is a vital complement to our research. It explores the predictability of changes in risk attitudes due to changes in observable exogenous factors, whereas we examine how the temporal stability of risk preferences can be judged in the absence of such exogenous shocks. Malmendier and Nagel (2010) use Survey of Consumer Finances data from 1964–2004 to investigate whether differences in individuals' experiences of

Table 1 Overview of recent literature on preference stability

	Survey (like) Questions	EUT[a]	PT
Stability across domains[b]	Weber et al. (2002)	Wolf and Pohlman (1983)	Camerer (1998)
	Guiso and Paiella (2006)	Eckel and Wilson (2004)	
	Lauriola et al. (2007)	Schechter (2007)	
	Nosic and Weber (2008)	Barseghyan et al. (2008)	
	Dohmen et al. (2009)		
Stability acrosselicitation methods[c,d]	Grable and Lytton (2001)	Isaac and James (2000)	Hershey and Schoemaker (1985)
	Kruse and Thompson (2003)	Bleichrodt et al. (2007)	Harbaugh et al. (2007)
	Anderson and Mellor (2009)	Berg et al. (2005)	Abdellaoui et al. (2007a)
	Dohmen et al. (2009)	James (2007)	Bleichrodt et al. (2007)
		Bruner (2008)	
		Deck et al. (2009)	
		Morone (2009)	
		Dave et al. (2010)	
Stability over time	Sahm (2007)	Wehrung et al. (1984)	Baucells and Villasís (2010)
	Brunnermeier and Nagel (2008)	Smidts (1997)	
	Vlaev et al. (2009)	Harrison et al. (2005)	
	Malmendier and Nagel (2010)	Andersen et al. (2008a,b)	
		Goldstein et al. (2008)	

An overview of recent literature on preference stability along two dimensions: stability type (stability across domains, stability across elicitation methods, and stability over time) and measure/theory [survey questions, Expected Utility Theory (EUT), and Prospect Theory (PT)]

[a] An extensive review of the developments in utility theory during the last decades is provided by Abdellaoui et al. (2007a)

[b] This field is closely related to framing effects. One may argue that different domains by themselves will induce differences in reference points. Comparisons of an aggregate level can be drawn from the application of PT to different fields

[c] Not across incentive levels (e.g., Camerer and Hogarth (1999); Holt and Laury (2002, 2005)) or implementation modes (e.g., von Gaudecker et al. (2008))

[d] For early work on the association between gain equivalences, CEs, and probability equivalences, see Bassler et al. (1973), Hershey et al. (1982), and Wehrung et al. (1984)

macro-economic shocks affect the willingness to take financial risks. They document a strong reaction of risk preferences to stock returns and inflation experienced over the course of an individual's life, even after controlling for socio-economic variables. Conversely, Brunnermeier and Nagel (2008) find that wealth fluctuations have hardly any influence on the proportion of liquid wealth invested in risky assets, suggesting that habits or subsistence levels do not cause individuals' risk aversion to be time-varying.[3] Their data stems from the Panel Study of Income Dynamics (1984–2003). Sahm (2007) corroborates this finding by analyzing Health and Retirement Study (1992–2002) data on hypothetical gambles over lifetime income. She observes that risk tolerance differs greatly across individuals, but is relatively stable for a particular individual. However, factors such as aging and business cycles can systematically affect an individual's risk tolerance. Documenting a significant correlation between the survey measure and actual risk taking (stock ownership), Sahm is convinced that preferences can be reliably measured at the individual level.

Although the functional form of PT models usually comes with a larger number of parameters than their EUT counterparts, both preference theories try to capture the general risk propensity of an individual. Consequently, evidence of somewhat stable EUT parameter estimates would constitute a solid foundation for any research on PT preferences. Wehrung et al. (1984) investigate the stability of CRRA parameters over a 1-year interval for 90 business executives. The hypothetical investment decisions are geared towards both personal and corporate resources. They find a small but highly significant positive correlation ($\rho = 0.36$) for the personal risk measures, but no stability for business risk propensity. Smidts (1997) examines the long-run temporal stability of risk attitudes of Dutch farmers concerning the market price for potatoes. Subjects are asked to specify certainty equivalents for 50/50 lotteries (midpoint chaining technique). This task is repeated after 1 year. He observes an even stronger correlation ($\rho = 0.44$) for the Pratt–Arrow coefficient of absolute risk aversion. Harrison et al. (2005) test the stability of risk preferences by estimating CRRA coefficients at two distinct points in time with the Holt and Laury (2002) procedure. Using panel regressions they find no significant differences in risk aversion over a time span of 5–6 months. Andersen et al. (2008a) study the temporal stability of Holt/Laury-based CRRA coefficient estimates by revisiting subjects at five points in time over a 17-month horizon. The follow-up experiments were conducted with a time lag of 3, 5, 11, and 17 months. They document some variation in the elicited coefficients over time, but correlations with the initial experiment of 0.43, 0.46, 0.58, and 0.34 (all significant at the 5% level) suggest a general tendency of stability. Goldstein et al. (2008) measure the consistency of CRRA parameters with a time lag of 1 year for a highly diversified subject pool. Participants use the distribution builder (Sharpe et al. 2000) to generate desired return distributions in a fictitious retirement savings scenario. The authors come to relatively similar results; the obtained correlation coefficients are even slightly higher than the above reported values.

To our knowledge, no prior work has looked directly at the stability of individual preference parameters over time when these are characterized by a PT functional. The

[3] More precisely, the changes in risk aversion are not large enough for individuals to overcome their inertia and adjust the asset allocation accordingly.

work by Baucells and Villasís (2010) on the existence and time stability of the reflection effect (risk aversion in the gain domain and risk propensity in the loss domain) could be related to PT, because this specific pattern of risk attitude is predicted under standard assumptions about the curvature of the PT value function (and when neglecting probability weighting). However, the experimental design of Baucells and Villasís (2010) builds on only two lottery choices (one in the gain domain and one in the loss domain) at two different points of time. It can thus not deliver more detailed preference information than just the general tendency of risk aversion or propensity. Furthermore, the issue of probability weighting is explicitly blinded out in their work. However, probability weighting and its interaction with the PT value function turn out to be highly relevant when measuring PT preferences and their stability in a more thorough way. Nevertheless, the study of Baucells and Villasís (2010) delivers the interesting insight that not even the general pattern of risk attitudes for gains and losses seems to be particularly robust over time. About 40% of subjects showed the reflection effect pattern in the first session, and 53% in the second session 3 months later. However, for only 24% was the reflection effect present in both sessions.

3 Interaction effects among PT parameters

3.1 General considerations

Prospect Theory preferences can be measured and elicited in various ways. One important aspect is to decide whether a parametric or a non-parametric approach should be taken. The assumption of a specific parameterization "confounds the general test of the theory with that of the specific parametric form" (Tversky and Kahneman 1992, p. 311). Therefore non-parametric analyses of the various facets of PT preference structures have been developed and presented in the recent literature (Abdellaoui 2000; Abdellaoui et al. 2007b).[4] However, the goal of our paper is not to test the validity of general PT preference structures (Abdellaoui et al. 2007b), but to analyze the stability of specific PT preferences and to highlight the interaction of the PT components. For this purpose, it seems more appropriate to restrict the PT value and probability weighting functions to specific parametric forms and to analyze the interaction of the limited number of parameters that capture the key ingredients of PT evaluation. This holds all the more, as Gonzalez and Wu (1999) have shown that for standard parameterizations of PT—as we use them in the following—good parametric fits can be observed both for value and probability weighting function at the individual level.

A second important aspect is to decide whether to estimate the PT parameters simultaneously from a number of certainty equivalents (CE) or to make use of one of the advanced "chained" elicitation methods that have been developed and that allow an independent elicitation of loss aversion, diminishing sensitivity and probability weighting (Abdellaoui et al. 2007a,b; Erner et al. 2010). We decided to choose the former approach. As we explain in the following (Abdellaoui et al. 2007b), the anal-

[4] Abdellaoui et al. (2008) provide a thorough review of the pros and cons of non-parametric fitting.

ysis of the time stability of PT parameters is strongly driven by the assumpTIONS about response errors and the "surprising effects" of such errors on the goodness of fit for seemingly very different PT parameter combinations. The symmetry that can be induced by simply collecting CEs for a set of carefully designed lotteries allows analyzing the interaction of response errors and parameter estimates in a clean way and without too many ad hoc assumptions. For the chained methods the analysis of response error is much more complicated and inscrutable and many further assumptions on error calibration and propagation would be needed.[5] In general, one has to decide between the theoretical elegance of an elicitation procedure on the one hand and its practical applicability on the other. Blavatskyy (2006) comes to the conclusion that the theoretically optimal elicitation procedure uses a non-parametric three-stage design. While theoretically appealing, such sophisticated elicitation methods show limitations in practice (see, e.g., Erner et al. (2010). One crucial assumption is, for instance, that the magnitude of the response error is unrelated to the complexity of the experimental task.

3.2 PT decision maker providing perfect answers

We start our analysis by considering a PT decision maker who provides CEs for simple lotteries without making any response error. We assume that his PT preferences can be described by the functional forms proposed by Tversky and Kahneman (1992):

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases} \tag{1}$$

$$w^{+}(p) = \frac{p^{\delta^{+}}}{(p^{\delta^{+}} + (1-p)^{\delta^{+}})^{1/\delta^{+}}}, \quad w^{-}(p) = \frac{p^{\delta^{-}}}{(p^{\delta^{-}} + (1-p)^{\delta^{-}})^{1/\delta^{-}}}. \tag{2}$$

Other functional forms could have been used for the analysis without major impact on the conceptual approach and the results.[6]

Within this framework diminishing value sensitivity, probability weighting, and loss aversion jointly determine the risk attitude of an individual (see, for example, Schmidt and Zank 2008). Importantly, this creates interdependencies among the corresponding parameters that are important to take into account when these are jointly elicited via CEs.[7] To illustrate the interdependencies we will use a (α, δ^{+})coordinate system where each point in the coordinate system represents a distinct combination (α for the shape of the value function; δ^{+} for the shape of probability weighting).[8]

---

[5] See Erner et al. (2010) for a detailed analysis of response error and error propagation in such chained methods.

[6] For the value function, a two-part power specification is largely undisputed. For probability weighting, tests of various functional forms have produced equivocal results (Abdellaoui et al. 2008). To avoid overfitting and due to the general noisiness of experimental data, we rely on a one-parameter weighting function to measure time stability (see, for example, Wu et al. 2004).

[7] Luce (1999, p. 28), van de Kuilen and Wakker (2009, p. 2), Abdellaoui et al. (2008, pp. 30–31), and Gonzalez and Wu (1999, p. 157, fn. 11), commented by Stott (2006, p. 21).

[8] For the sake of brevity we restrict all illustrations to the gain domain. Yet, the same reasoning can directly be applied to the estimation of β and δ^{-} based on loss lotteries.

Similar to Bleichrodt et al. (2007) we restrict the parameter space to the range from 0.20 to 2.00 (for the graphical displays and for optimizing procedures). For any given lottery and (sensibly stated) CE, one can compute a set of parameter combinations that are in line with the observed decision behavior. In our (α, δ^+) coordinate system these combinations form a curved line of a particular shape. If we restrict our attention to binary lotteries with the lower outcome fixed at zero, the shapes of these compensation curves are solely determined by the probabilities of the upper outcome (p). The size x of the non-zero outcome is not relevant by contrast, because for power value functions CEs are proportional to outcome size as shown in Eq. 3.

$$\begin{aligned} \text{CE} &= \left(w^+(p)\cdot x^\alpha + w^+(1-p)\cdot 0^\alpha\right)^{\frac{1}{\alpha}} = \left(w^+(p)\right)^{\frac{1}{\alpha}}\cdot x \\ &= \left(\frac{p^{\delta+}}{\left(p^{\delta+}+(1-p)^{\delta+}\right)^{1/\delta+}}\right)^{\frac{1}{\alpha}}\cdot x \end{aligned} \tag{3}$$

For an illustration, assume that a PT decision maker with $\alpha = 0.88$ and $\delta^+ = 0.61$ (these are the median values estimated by Tversky and Kahneman 1992) faces a fifty-fifty gamble for a chance to win 60€ (again fixing the non-zero outcome at 0€). According to his parameters, the T/K median decision maker will state a CE of 22.43€. However, due to the compensating effects between the curvature and the probability weighting function, 22.43€ is also the CE response of an infinite number of other PT decision makers with different (α, δ^+) parameters. Figure 1a shows the compensation curve of (α, δ^+)combinations leading to the same CE of 22.43€.

In general, for each upper outcome probability, there exists one distinct compensation curve with (α, δ^+) combinations that lead to the same CE as the true parameter combination. Compensation curves for similar (gain) probabilities exhibit a similar shape and reflect the interaction effects between α and δ^+. While it is not surprising that a combination of two preference parameters α_{true} and δ^+_{true} cannot be deduced from a single indifference statement, the particular shape of the compensation curves will prove to be important in the following. Figure 2 displays respective curves of a T/K median decision maker for gain lotteries with winning probabilities of 20,

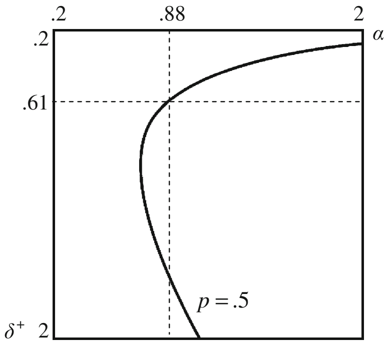


Fig. 1 Compensation curve of $(\alpha; \delta^+)$ combinations for a single 50/50 lottery. The compensation curve of parameter combinations leading to the same CE as the (perfect) Tversky/Kahneman median decision maker's CE [calculated with $(\alpha_{\text{true}}; \delta^+_{\text{true}}) = (0.88; 0.61)$] for a two-outcome lottery with one positive and one zero outcome is shown. The positive outcome's probability p equals 0.5

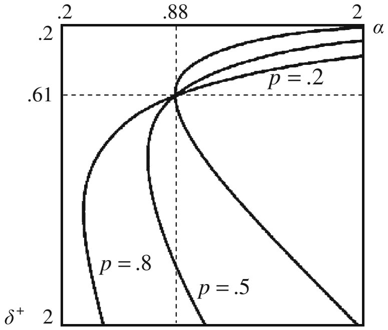


Fig. 2 Compensation curves of (α; δ^+) combinations for single lotteries. The compensation curves of parameter combinations leading to the same CE as the Tversky/Kahneman median decision maker's CE for two-outcome lotteries [calculated with (α_{true}; δ^+_{true}) = (0.88; 0.61)] are shown. The upper (i.e., non-zero) outcome's probability p equals 0.2, 0.5, and 0.8, respectively

50, and 80%. Most importantly, the compensation curves show a relatively similar shape despite the use of different probabilities. The true (α, δ^+) combination (0.88, 0.61) is placed on each curve and it is the only parameter combination explaining all indifference statements.

3.3 PT decision maker with response error

The analysis of Sect. 3.2 assumed the decision maker to be able to state his true CEs without any response error. In such a scenario, two lotteries with different gain probabilities are sufficient to elicit the decision maker's true (α, δ^+) parameters. If a larger set of lotteries is used, the unique (α, δ^+) combination will just be reinforced. When integrating a response error, however, the compensation curves introduced in Sect. 3.1 usually do not cross in one point and need to be replaced by a (sensibly chosen) goodness-of-fit measure. For this purpose we use maximum likelihood (ML) estimation and analyze how likely the stated CEs are generated by various (α, δ^+) combinations if an error component is present in the evaluation process. We assume the error term to be normally distributed on the stated CE scale. As the employed set of lotteries has different outcome sizes we further assume that there is a constant relation c between the size of the error term and lottery i's standard deviation.[9] If n CEs are elicited, the (log-)likelihood function is thus given for any (α, δ^+) combination as:

$$\text{LL} = \sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sqrt{2\pi}c\sigma_i}\cdot\exp\left(-\frac{1}{2}\left(\frac{\text{CE}_{\text{stated},i}-\text{CE}_i(\alpha,\delta^+)}{c\sigma_i}\right)\right)^2\right)\right] \qquad (4)$$

[9] The specification of c is important for the upcoming discussion of temporal stability of preferences. We will explain in Sect. 4, when we discuss our experimental design and findings, how we came to assume a value of 19,8% for c. In this section, we will also use this specific value of c for the graphical displays.

The ability of each (α, δ^+) combination to explain an observed set of CE answers can be illustrated by "(log-) likelihood topographies". To generate the topographies we apply a brute-force approach and estimate the likelihood for any (α, δ^+) combination on a grid within a sensible range. The restriction of the parameter space to the range from 0.20 to 2.00 and a grid size of 0.01 implies the calculation of $181 \times 181 = 32{,}761$ likelihoods. Higher levels in the topography denote better (α, δ^+) fits, i.e., a higher likelihood that a specific (α, δ^+) combination has generated the given set of CE responses. Figure 3a illustrates the log-likelihood topography of the T/K median decision maker, perfectly reporting his true CEs for gain lotteries with winning probabilities of 20, 50, and 80% when the true CEs are augmented with an error component calibrated to fit our experimental data.[10] Again, a main direction of the interaction effects between α and δ^+ is clearly visible. Adjacent to the optimal solution of $\alpha = 0.88/\delta^+ = 0.61$, a variety of parameter combinations is able to provide fits that explain the decision-making behavior of the T/K median decision maker almost equally well. This insight recommends implementing lottery sets with larger ranges of probabilities to achieve a larger variation in the compensation curves. Although interaction effects can be partly offset by this approach, sizeable interaction effects remain. This holds true even if probabilities range from 1 to 99%, as can be seen in Fig. 3b where a log-likelihood topography is plotted for the extensive set of 28 lotteries used by Tversky and Kahneman (1992). These flat (likelihood) maxima not only impede the estimation of PT parameters at a single point in time but also necessitate a joint analysis of α and δ^+ for measuring PT time stability.

3.4 The time stability of PT preferences

As we have seen, interaction effects between value function and probability weighting function are systematic. Hence, if we elicit PT preferences at two different points in time, a decision maker can exhibit nearly time-stable preferences if his point estimates for α and δ^+ move alongside the main direction of compensation. Therefore an isolated analysis of α and δ^+ would prove to be highly error-prone. For a joint analysis of the temporal stability of value and probability weighting function parameters, we have to use the information about their joint likelihood distribution in a likelihood ratio (LR) test instead. The LR test measures whether the two samples of elicited parameter sets, one for each experimental session, come from the same population. The null hypothesis of the LR test is time stability; hence significance levels report the probability of a subject being incorrectly classified as time-instable. This probability results from a ML parameter estimation that combines the subject's CEs of both sessions. As indicated before, the p values will depend on the size of the scaling factor c that we use during the ML estimation procedure. The larger the error term is assumed, the more likely CE deviations are explained by noise rather than time-instable behavior. Thus, we will defer the further discussion of time stability issues to Sect. 5 where we demonstrate the procedure based on real experimental data and

[10] More specifically, the likelihood topography is created by the use of the nine gain lotteries of our experiment (see Sect. 4 for details).

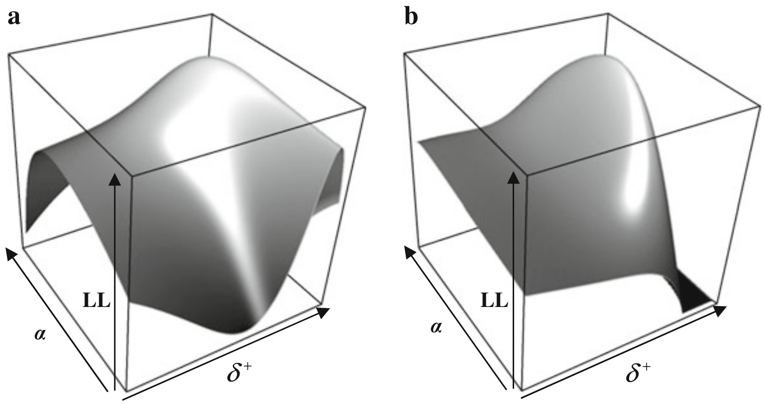


Fig. 3 Log-likelihood topographies for Tversky/Kahneman median decision maker. The figure illustrates the G/L model's log-likelihood (LL) topography for the Tversky/Kahneman median decision maker $(\alpha_{\text{true}}; \delta^{+}_{\text{true}}) = (0.88; 0.61)$, reporting his true CEs for all lotteries. The (relative) log-likelihood of all parameter combinations within the grid is indicated by grayscales, a lighter grayscale indicates a higher log-likelihood. **a** Based on the nine gain lotteries of this experiment, **b** features the extensive set of 28 lotteries with nine different probabilities ranging from 1 to 99% used by Tversky and Kahneman (1992). For optical reasons, the profiles are cut off at a log-likelihood of −58 (−300) in **a** (**b**), i.e., all values below this threshold are displayed as if they were lying exactly at this level

determine the *c* endogenously. Importantly, other (not Tversky/Kahneman) functional forms of PT preferences pose the same challenge: they come with their own likelihood topographies. As a consequence, interaction effects also play an important role when various functional specifications of PT preferences are comparatively tested.

4 Experiment and parameter estimation

4.1 Experimental design

The theoretical considerations of Sect. 3 are now illustrated by an experimental study in which PT parameters are elicited twice. To enable a test of preference stability over time, our experiment comprises two sessions with a time lag of 1 month. The first session took place in December 2007, the second in January 2008. Overall, 86 advanced undergraduate students from the University of Münster, Germany, took part in both sessions. The average age of participants was 23.8 years (median: 23 years) at the time of the first session and 22% of subjects were female. The experiment was computer-based and instructions were provided on screen. To be brief in general, a substantial part of the instructions was geared towards achieving a full understanding of stating a CE for a two-outcome lottery. Additionally, multiple exemplary answers and their consequences ("what-if" scenarios) were discussed to guarantee that subjects answered according to their true preferences. Subjects were also encouraged to ask the experimenter for help in case they did not fully understand the instructions (we refer to Appendix A for the full instructions).

Table 2 Overview of lotteries in elicitation procedure

Gain lotteries									
Upper outcome	+5€	+5€	+5€	+20€	+20€	+20€	+60€	+60€	+60€
Probability (%)	20	50	80	20	50	80	20	50	80
Loss lotteries									
Lower outcome	−5€	−5€	−5€	−20€	−20€	−20€	−60€	−60€	−60€
Probability (%)	20	50	80	20	50	80	20	50	80
Mixed lotteries									
Upper outcome	+5€	+5€	+20€	+20€	+60€	+60€			
Probability (%)	50	50	50	50	50	50			
Lower Outcome	−20€	−60€	−5€	−60€	−5€	−20€			

An overview of the 24 risky lotteries used in the elicitation procedure grouped by type (gain, loss, and mixed)

In each of the two sessions, a set of 24 lotteries was presented to subjects. These were divided into three categories: nine gain, nine loss, and six mixed lotteries.[11] All lotteries were constructed from the same set of outcomes and probabilities. The outcomes were ±5€, ±20€, and ±60€, the associated probabilities 20, 50, and 80%. Each of the nine gain and loss lotteries combined one of the three outcomes with one of the three probabilities. The remaining probability mass was attributed to the second outcome fixed at 0€. All six mixed lotteries had equal chances for gains and losses. Table 2 presents an overview of all lotteries used in the experiment.

For each of the 24 lotteries, subjects were asked to state their CE. Subjects were first confronted with the gain lotteries, followed by the loss lotteries, and finally, the mixed ones. Within each category the lotteries were presented in a randomized order. The CEs had to be entered via a keyboard with a maximum of two decimal places (Eurocent).[12] Only CEs within the outcome range of a lottery were accepted. Otherwise an error message was displayed and subjects were prompted to re-enter their CE.

To warrant incentive compatibility, every tenth subject received a variable payment in both experimental sessions. The group of "winners" was drawn independently for both runs. In either session the elicitation procedure was the second part of a larger experiment.[13] We used the first part of both experiments to generate earnings of at least 60€ for all subjects before the elicitation questions. The total payment equaled the sum of the earnings of the first and the second part of the experiment. The gains of the first part were reduced if there was a loss incurred in the second part. Consequently, this linked setup allows for an incentive-compatible elicitation of preferences

[11] Three additional gain and loss lotteries with gain/loss probabilities of 5% were removed from the dataset because the corresponding answers showed disproportionally high noise, probably due to rounding behavior of subjects, as CEs were relatively low. An inclusion of these additional lotteries, however, does not qualitatively change the results presented in the paper.

[12] We refer to Appendix B for a sample screen layout.

[13] The first part of the December session comprised a set of investment decisions targeted at myopic loss aversion. In the January session, the preceding investment decisions were used to test a distribution builder-like application.

even for loss and mixed lotteries. In both sessions, the elicitation part lasted about 30 min including reading the instructions. The average length of the total experiment was 50 min for the December and 60 min for the January session. For the payout of the elicitation part, one of the lotteries was selected randomly and played by applying the Becker–DeGroot–Marschak (1964) mechanism. Owing to their balanced structure, the expected value of the 30 lotteries equaled 0€. The variable payment for the whole experiment ranged from 35.26€ to 242.32€. Additionally, each participant earned a fixed payment of 4€ for the December and 10€ for the January session. All payment details were clearly communicated to the subjects at the beginning of each experimental session.

4.2 Parameter estimation technique

As explained in Sect. 3.3, we rely on ML estimation to determine the individual PT parameters. To meet both goals of this paper (illustrating our methodology and presenting initial evidence on the temporal stability of PT parameters), we work with two different model specifications: gains/losses only (G/L) and full. The G/L model is particularly suited to illustrate the methodological framework for analyzing the temporal stability of risk attitudes under PT. It determines likelihoods for combinations of α and δ^+ based on the nine gain lotteries ($i = 1..9$), as well as for combinations of β and δ^- based on their loss counterparts. This is done for all subjects j and separately for session s ($s = 1 \ldots 2$), see Eq. 5a for gain lotteries. In a second step, the loss aversion parameter λ is estimated over the six mixed lotteries using the other parameters as input. This segmented approach allows for a separate test of PT components on time stability, and hence, benefits the interpretability of results.

The full model allows classification of subjects according to the stability of their overall decision-making behavior over time under PT. It jointly estimates all five PT parameters (α, δ^+, β, δ^-, and λ) and includes the entire set of 24 lotteries. In contrast to the G/L model, curvature and probability weighting estimates are also influenced by the responses to the mixed lotteries. The log-likelihood of Eq. 5b is maximized with a gradient algorithm.

$$\mathrm{LL}^{\mathrm{gain}}_{G/L,j,s} = \sum_{i=1}^{9}\left[\ln\left(\frac{1}{\sqrt{2\pi}c\sigma_i}\cdot\exp\left(-\frac{1}{2}\left(\frac{\mathrm{CE}_{\mathrm{stated},i,j,s}-\mathrm{CE}_{i,j,s}(\alpha,\delta^+)}{c\sigma_i}\right)\right)^2\right)\right] \tag{5a}$$

$$\mathrm{LL}_{\mathrm{full},j,s} = \sum_{i=1}^{24}\left[\ln\left(\frac{1}{\sqrt{2\pi}c\sigma_i}\cdot\exp\left(-\frac{1}{2}\left(\frac{\mathrm{CE}_{\mathrm{stated},i,j,s}-\mathrm{CE}_{i,j,s}(\alpha,\delta^+,\beta,\delta^-,\lambda)}{c\sigma_i}\right)\right)^2\right)\right] \tag{5b}$$

For both model specifications the error term σ_i is assumed to be normally distributed on the stated CE scale. The parameter c captures the constant relation between the size of the error term and lottery i's standard deviation[14] and is calibrated to maximize the like-

[14] Other weighting schemes are possible. As a robustness check, we also estimated parameters by assuming lottery-triple specific outcome weighted error terms. Results applying this standard error specification are very similar.

Table 3 Prospect Theory parameter estimations G/L model

	α	δ^+	β	δ^-	λ
Panel A: parameters 1st session					
25%	0.89	0.75	0.80	0.65	0.77
50% (median)	0.99	0.86	0.88	0.83	1.46
75%	1.11	0.98	1.01	0.95	2.84
Panel B: parameters 2nd session					
25%	0.88	0.76	0.80	0.66	0.63
50% (median)	1.00	0.87	0.93	0.75	1.23
75%	1.13	0.97	1.02	0.88	2.80
Panel C: parameters both session (pooled)					
25%	0.89	0.74	0.80	0.66	0.78
50% (median)	0.98	0.89	0.88	0.77	1.41
75%	1.08	0.94	1.00	0.90	2.63

Quartiles of PT parameter point estimates (G/L model) of the 73 subjects for the first (Panel A), second (Panel B), and the combined estimation for both sessions assuming time-consistent decision making behavior (Panel C) are shown

lihood across all subjects $j = 1 \ldots J$: $\max[\sum_{j=1}^{J} (\sum_{s=1}^{2} (\mathrm{LL}_{G/L,j,s}^{\text{gain}} + \mathrm{LL}_{G/L,j,s}^{\text{loss}}))]$.[15] The resulting (optimal) factor amounts to $c = 0.198$. Hence, we assume an error term sized to 19.8% of the corresponding lottery's standard deviation. The magnitude of the random error has to be kept in mind when interpreting the results concerning time stability. Yet, it has no influence on the point estimates, i.e., all parameter point estimates are independent of c.

5 Parameter estimates and their stability over time

5.1 Findings on the group level

To maintain the focus of the analyses, we exclude subjects who had difficulty understanding the experimental task or were insufficiently motivated. This is accomplished by a pair-wise comparison of lotteries to detect first-order stochastic dominance violations.[16] By applying this criterion we exclude 13 of the 86 subjects, i.e., 15%.

[15] A theoretically appealing alternative to this method is represented by introducing lottery-specific error terms via an additional session directly after the first one, i.e., without (significant) time delay. Differences in stated CEs between these two sessions can only result from noise (rather than temporal instability). The standard deviation of the error term could thus be set equal to the standard deviation of CEs between these two sessions (pooled over all subjects). Such a procedure, however, implies that subjects must be confronted with three sessions in total. In order to attempt to complete the experimental task within acceptable bounds we applied a two-session procedure.

[16] We regard any CE answer that is strictly greater than the stated CE for a (first-order) stochastically dominating lottery as inconsistent. In our experiment, a stochastically dominating lottery is a lottery with an equal or higher gain probability and an equal or higher (non-zero) outcome, while one of these two inequalities needs to be strict. In any of the four scenarios (two domains, two experimental sessions), we allow for one of the nine stated CEs to be varied arbitrarily to resolve apparent inconsistencies. If, in at least

Table 4 Prospect Theory parameter estimations full model

	α	δ^+	β	δ^-	λ
Panel A: parameters 1st session					
25%	0.92	0.75	0.82	0.66	0.75
50% (median)	1.00	0.86	0.91	0.82	1.42
75%	1.12	0.98	1.04	0.95	2.62
Panel B: parameters 2nd session					
25%	0.90	0.75	0.82	0.66	0.72
50% (median)	0.99	0.86	0.93	0.74	1.37
75%	1.13	0.97	1.03	0.89	2.41
Panel C: parameters both session (pooled)					
25%	0.91	0.74	0.81	0.64	0.78
50% (median)	0.99	0.88	0.90	0.76	1.37
75%	1.09	0.92	1.01	0.89	2.42

Quartiles of PT parameter point estimates (full model) of the 73 subjects for the first (Panel A), second (Panel B), and the combined estimation for both sessions (Panel C) are shown

Seventy-three subjects remain in the dataset. All following analyses are based upon this restricted group. We also used the restricted dataset to determine the scaling factor c. Using the ML estimation technique described in Sect. 4.1 (G/L model), we calculate parameters that are within the range of related studies, see Table 3 for quartile results.[17] In the gain domain our median estimates for curvature are equal to 0.99 and 1.00 for the two experimental sessions, indicating linearity of the value function. The interquartile range covers values between 0.88 and 1.13 for both sessions. In the loss domain we observe βs of 0.86 in the first and 0.93 in the second session, indicating the typical convex shape of the value function in the loss domain.[18] Sixty-five percent of the subjects in the first and 70% in the second session have β values lower than one.

We find subjects to conduct moderate probability weighting—slightly more pronounced in the loss domain—with median values of 0.86/0.87 in the gain and 0.83/0.75 in the loss domain (first/second session). Thus, we observe slightly less pronounced median probability weighting than other studies, e.g., Abdellaoui (2000). In the four domain/time combinations, a fraction of 79 to 92% of subjects shows the characteristic inverse S-shaped probability weighting, i.e., δ values of lower than one. Median estimates for the loss aversion coefficient λ equal 1.46 and 1.23 for first and second session.

Using the full model to jointly estimate all five PT parameters, the results are very similar (see Table 4). The majority of subjects show only marginal differences in parameter estimates: for 72% of subjects the α/β estimate remains

one of the four scenarios, more than one CE must be varied to relieve the set of nine CE answers from any inconsistency, the corresponding subject is excluded from the analyses.

17 A comprehensive overview of the gain domain is presented by Stott (2006), Table 5. Abdellaoui et al. (2008) survey loss aversion (see Table 1).

18 Booij and van de Kuilen (2009) and Abdellaoui et al. (2007a) also observe the curvature to be more pronounced in the domain of losses. Abdellaoui (2005) document just the opposite.

within 0.05 of the G/L model value, for δ^+/δ^- an even larger fraction of the differences (96%) remain below this threshold. The output of both models is even more alike when it comes to median estimates (across all subjects). Over all four curvature and probability weighting parameters, the difference in median estimates is at most 0.04, although the loss aversion coefficient differs by 0.14. Overall, the median (and also quartile) coefficients on all PT parameters are remarkably stable in both the G/L and the full model. Furthermore, our parameter estimates are comparable to previously published values.[19] All individual full model parameter estimates for both experimental sessions are provided in Table 5.

5.2 Time stability at the individual level

To jointly analyze the temporal stability of value and probability weighting function parameters, we have to use the information about their joint likelihood distribution in a LR test. The null hypothesis of the LR test is time stability; hence significance levels report the probability of a subject being incorrectly classified as time-instable. The numerator of the test statistic corresponds to the probability of observing the stated CEs under the null. This probability results from a ML parameter estimation that combines the subject's CEs of both sessions to a total of 18 (G/L model). The denominator corresponds to the probability of observing the stated CEs under the alternative hypothesis, i.e., the likelihoods of separated parameter estimations for sessions one and two are multiplied. As indicated before, the p values depend on the size of the scaling factor c determined during the ML estimation procedure.[20] The larger the error term assumed, the more likely CE deviations are explained by noise rather than time-instable behavior. As shown in Sect. 4.1, we estimated the c at 19.8% of a lottery's standard deviation. The use of a smaller standard error would decrease reported significance levels, i.e., increase the fraction of subjects classified as time-instable (for a given significance level) and *vice versa*. Visually speaking, a smaller standard error would increase the slope in the likelihood topographies.

At a significance level of $p = 5\%$, 12 (27) of the 73 subjects, i.e., 16% (37%) are classified as time-instable in the gain (loss) domain. Overall, 33 subjects (45%) are instable in at least one of the two domains. Individual p values for the gain and

[19] We also classified subjects according to the stated CEs in relation to the expected value of the lotteries, a similar approach to the weak definition of risk aversion by Chateauneuf and Cohen (1994). The majority of subjects shows characteristic PT pattern, i.e., showing risk averse behavior in the gain and risk seeking behavior in the loss domain. Comparing these general risk attitudes between two sessions constitutes an alternative approach of measuring time stability.

[20] Since the LR test statistic is only asymptotically χ^2 distributed, we use a bootstrap procedure to determine the statistical significance of subjects' temporal instability. For each lottery we randomly generate 10,000 normally distributed CE answers with means equal to the G/L model's point estimates (pooled estimates over both sessions) and standard deviations equal to the standard deviation of the lottery and compute 10,000 LRs. The resulting quasi-continuous distribution of LRs is then compared to the actual likelihood to determine the p value. As the resulting distribution of LRs depends on the estimated parameters, the bootstrap procedure is to be conducted for each parameter combination and subject, respectively.

Table 5 Individual Prospect Theory preference parameters

Subj. ID	1st session					2nd session					Both sessions (pooled)					Time stability (G/L) (Bootstrapped p value)	
	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	Gains	Losses
Median	1.00	0.86	0.91	0.82	1.42	0.99	0.86	0.93	0.74	1.37	0.99	0.88	0.90	0.76	1.37		
1	0.94	0.71	1.05	0.58	0.75	0.85	0.74	0.59	0.69	2.54	0.89	0.72	0.82	0.61	1.40	0.922	0.007
2	0.86	0.98	0.86	0.89	1.16	1.11	0.56	0.80	0.63	0.52	0.93	0.74	0.80	0.74	1.30	0.008	0.018
3	1.47	1.12	0.72	0.90	11.58	1.29	1.20	0.95	0.96	1.60	1.40	1.18	0.81	0.93	5.43	0.341	0.092
4	0.96	0.59	0.93	0.87	0.62	0.91	0.73	1.01	0.84	0.20	0.91	0.67	0.96	0.86	0.39	0.211	0.891
5	0.92	1.07	0.84	0.93	0.76	1.49	0.83	0.92	0.66	17.16	1.11	0.92	0.84	0.79	1.95	0.000	0.157
6	0.93	0.92	0.87	0.85	1.32	0.88	0.86	0.82	0.52	2.53	0.91	0.89	0.80	0.67	2.16	0.672	0.000
7	0.87	0.75	1.05	0.56	0.70	0.89	0.61	1.44	0.49	0.13	0.87	0.68	1.22	0.53	0.31	0.416	0.194
8	0.76	0.82	1.25	0.81	0.14	0.91	0.98	1.22	0.72	0.59	0.82	0.88	1.22	0.76	0.29	0.300	0.741
9	1.00	1.00	1.00	1.00	0.99	0.97	0.90	1.01	0.99	0.78	0.99	0.95	1.00	0.99	0.92	0.848	0.996
10	1.03	1.01	1.11	1.00	0.85	0.92	0.98	0.79	1.09	1.57	0.97	0.99	0.93	1.02	1.19	0.909	0.104
11	1.00	0.89	1.88	2.02	0.13	0.89	0.70	1.15	0.89	0.56	0.94	0.78	1.17	0.94	0.61	0.294	0.720
12	1.04	0.68	0.93	0.46	2.26	1.43	2.02	0.92	0.91	2.41	0.96	0.83	0.88	0.64	1.39	0.076	0.000
13	0.94	0.93	1.04	0.99	1.00	0.99	0.90	1.02	0.97	0.84	0.96	0.91	1.03	0.98	0.87	0.985	0.939
14	1.08	0.72	0.89	0.78	2.06	1.24	0.77	0.73	0.66	5.49	1.16	0.74	0.81	0.71	3.32	0.530	0.050
15	0.99	0.47	0.64	0.50	2.35	1.00	0.61	0.87	0.67	0.20	0.97	0.54	0.73	0.58	0.71	0.059	0.000
16	1.27	0.66	0.82	0.46	11.17	0.90	0.89	0.81	0.72	0.88	1.04	0.77	0.77	0.59	2.72	0.052	0.001
17	1.20	0.94	1.04	1.10	2.22	1.18	0.72	2.02	0.44	0.14	1.16	0.81	1.34	0.65	0.68	0.370	0.000
18	0.83	0.94	0.83	0.88	1.02	0.90	0.83	0.87	0.90	0.91	0.86	0.88	0.85	0.89	0.94	0.818	0.977
19	0.85	0.89	1.23	1.19	0.23	0.91	0.88	1.10	0.68	0.96	0.88	0.88	1.09	0.86	0.61	0.817	0.018
20	1.09	0.52	0.42	0.72	5.02	1.26	0.57	0.65	0.73	24.29	1.12	0.56	0.54	0.71	8.55	0.151	0.000

Table 5 continued

	1st session					2nd session					Both sessions (pooled)					Time stability (G/L)	
Subj. ID	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	(Bootstrapped p value)	
Median	1.00	0.86	0.91	0.82	1.42	0.99	0.86	0.93	0.74	1.37	0.99	0.88	0.90	0.76	1.37	Gains	Losses
21	0.96	0.59	0.72	0.50	1.91	1.08	0.78	0.74	0.51	3.58	1.01	0.67	0.72	0.51	2.56	0.035	0.822
22	1.14	0.84	0.85	0.74	3.67	1.58	0.51	0.89	0.48	5.06	1.27	0.66	0.83	0.61	3.99	0.015	0.003
23	0.94	0.64	1.06	0.92	0.71	1.13	0.90	1.00	0.66	3.32	1.02	0.74	1.01	0.77	1.59	0.010	0.068
24	0.84	0.95	0.89	0.91	0.70	0.88	0.87	1.00	0.91	0.72	0.86	0.91	0.95	0.91	0.70	0.893	0.728
25	1.09	0.85	1.39	0.67	1.03	1.07	0.78	1.02	0.54	0.72	1.07	0.82	1.19	0.58	1.16	0.738	0.001
26	1.27	1.16	0.86	0.73	5.45	1.09	0.95	1.08	0.78	1.19	1.17	1.04	0.96	0.76	2.66	0.729	0.117
27	0.93	0.98	1.06	1.00	0.68	0.79	0.97	1.45	2.02	0.16	0.87	0.97	1.03	1.14	0.61	0.451	0.283
28	1.19	0.85	0.87	0.86	2.92	0.96	0.97	0.92	0.81	2.20	1.06	0.91	0.90	0.83	2.36	0.410	0.681
29	1.02	0.98	0.91	0.96	1.78	1.00	1.00	1.00	1.00	1.00	1.01	0.99	0.95	0.98	1.37	0.961	0.897
30	1.50	0.52	1.14	0.31	4.13	1.36	0.75	0.80	0.56	10.94	1.41	0.62	0.87	0.45	8.17	0.074	0.007
31	0.89	0.79	1.65	0.47	0.10	1.82	2.02	1.36	0.66	2.16	0.99	0.91	1.49	0.56	0.21	0.126	0.310
32	1.01	0.84	1.11	0.53	0.43	0.97	0.92	1.12	0.55	0.36	0.99	0.88	1.12	0.54	0.38	0.842	0.946
33	1.37	0.78	0.71	0.69	13.41	0.90	0.56	0.93	0.61	0.38	1.09	0.65	0.80	0.65	1.79	0.000	0.389
34	0.91	0.95	0.91	0.95	0.99	0.93	0.94	0.87	0.90	1.29	0.92	0.95	0.89	0.92	1.13	0.992	0.933
35	0.94	0.93	1.01	0.94	1.27	0.93	0.75	1.13	0.72	0.32	0.91	0.84	1.05	0.82	0.57	0.457	0.262
36	0.86	0.86	2.02	1.95	0.03	0.99	1.01	1.03	0.84	2.31	0.91	0.91	1.18	0.99	0.62	0.229	0.019
37	1.71	0.70	0.59	0.67	56.99	1.29	0.72	0.90	0.37	9.18	1.52	0.70	0.65	0.54	27.34	0.016	0.023
38	0.97	0.96	1.04	0.99	0.85	0.87	0.86	1.06	0.70	0.63	0.92	0.90	1.03	0.82	0.77	0.597	0.189
39	0.99	0.86	0.71	0.56	2.30	1.29	0.98	0.71	0.85	3.49	1.13	0.92	0.68	0.70	3.22	0.079	0.004
40	0.91	0.80	0.62	0.71	4.69	0.75	0.88	0.66	0.56	1.79	0.84	0.83	0.64	0.63	3.15	0.457	0.226

Table 5 continued

	1st session					2nd session					Both sessions (pooled)					Time stability (G/L)	
Subj. ID	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	(Bootstrapped p value)	
Median	1.00	0.86	0.91	0.82	1.42	0.99	0.86	0.93	0.74	1.37	0.99	0.88	0.90	0.76	1.37	Gains	Losses
41	0.96	1.02	0.85	0.85	1.93	0.96	0.86	0.76	0.86	1.91	0.96	0.93	0.81	0.85	1.94	0.748	0.944
42	1.16	0.72	0.80	0.63	2.11	1.23	0.87	0.63	0.92	8.01	1.19	0.79	0.70	0.76	3.94	0.212	0.162
43	0.90	0.83	1.20	1.12	0.44	1.07	0.63	0.88	0.65	2.35	0.98	0.72	0.97	0.80	1.24	0.379	0.000
44	1.45	1.28	0.70	0.92	16.11	2.02	1.19	0.70	0.84	48.38	1.74	1.26	0.70	0.88	28.16	0.057	0.862
45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	1.000
46	1.39	0.98	0.63	0.72	15.93	1.33	0.86	0.75	0.88	8.46	1.36	0.91	0.69	0.79	11.81	0.876	0.014
47	0.84	0.84	0.75	0.69	1.35	0.88	0.85	0.75	0.68	1.99	0.86	0.85	0.75	0.68	1.62	0.650	0.915
48	0.43	0.49	0.90	0.41	0.54	1.35	0.62	0.83	0.37	3.99	0.89	0.50	0.92	0.37	1.12	0.000	0.406
49	1.08	0.96	0.79	0.45	2.38	1.27	0.84	0.91	0.81	5.37	1.17	0.89	0.80	0.61	3.93	0.376	0.000
50	1.01	0.89	0.80	0.96	2.46	1.06	1.08	1.16	0.62	1.70	1.03	0.97	0.93	0.77	2.22	0.686	0.010
51	1.27	1.18	0.96	0.99	3.70	0.79	0.77	0.82	0.71	0.81	0.95	0.92	0.88	0.83	1.22	0.000	0.022
52	0.92	0.92	0.90	0.89	0.90	0.82	0.90	1.28	1.16	0.22	0.87	0.91	1.05	1.00	0.50	0.843	0.040
53	0.99	1.00	0.98	0.99	1.42	0.97	0.99	0.99	0.96	0.90	0.98	0.99	0.98	0.98	1.18	0.999	0.977
54	0.79	0.83	1.08	1.08	0.33	0.83	0.83	0.99	0.81	0.59	0.81	0.83	1.01	0.91	0.49	0.993	0.382
55	0.90	0.84	0.94	0.52	1.52	0.90	0.94	0.85	0.60	1.57	0.90	0.89	0.89	0.56	1.58	0.809	0.710
56	1.18	0.63	0.73	0.74	4.43	0.75	0.99	1.03	0.77	0.23	0.94	0.78	0.84	0.76	1.23	0.017	0.010
57	1.42	0.89	0.91	0.76	8.94	1.25	0.84	1.03	0.46	3.32	1.33	0.86	0.92	0.61	5.53	0.464	0.003
58	1.03	0.80	0.85	0.82	2.01	1.01	0.87	0.86	0.88	1.07	1.01	0.83	0.86	0.85	1.35	0.883	0.919
59	1.12	0.96	0.93	0.79	1.42	0.94	0.88	0.74	0.73	1.94	1.02	0.92	0.83	0.75	1.68	0.343	0.062
60	0.76	0.75	1.02	0.84	0.59	0.88	0.57	0.99	0.85	0.47	0.81	0.65	1.00	0.84	0.56	0.180	0.977

Table 5 continued

	1st session					2nd session					Both sessions (pooled)					Time stability (G/L)	
Subj. ID	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	α	δ^+	β	δ^-	λ	(Bootstrapped p value)	
Median	1.00	0.86	0.91	0.82	1.42	0.99	0.86	0.93	0.74	1.37	0.99	0.88	0.90	0.76	1.37	Gains	Losses
61	1.28	0.67	0.83	0.82	6.42	0.91	0.75	0.82	0.69	1.03	1.07	0.71	0.81	0.76	2.42	0.055	0.446
62	0.95	0.70	0.74	0.82	2.30	0.79	0.78	0.58	0.67	2.07	0.86	0.74	0.65	0.74	2.20	0.559	0.016
63	1.02	1.01	1.01	1.00	0.60	1.00	1.00	0.99	1.00	1.05	1.00	1.00	1.01	1.00	0.71	0.999	0.997
64	0.93	0.78	0.89	0.82	0.80	1.11	0.84	0.98	0.74	1.65	1.01	0.81	0.93	0.78	1.07	0.139	0.661
65	0.98	0.71	0.75	0.56	3.67	0.82	0.75	0.71	0.68	1.29	0.90	0.73	0.72	0.62	2.34	0.594	0.377
66	1.08	1.02	1.05	0.95	1.16	0.94	0.94	1.07	0.92	0.64	1.01	0.98	1.06	0.93	0.86	0.558	0.986
67	0.95	0.93	1.03	0.90	0.73	0.99	1.00	1.04	0.94	0.80	0.97	0.96	1.03	0.92	0.78	0.890	0.958
68	1.05	0.99	0.84	1.02	1.82	1.46	1.13	0.94	0.84	9.78	1.22	1.03	0.88	0.92	3.67	0.038	0.441
69	1.10	1.08	0.92	0.76	2.44	1.06	1.03	1.03	0.76	1.74	1.09	1.07	0.97	0.76	2.22	0.827	0.811
70	1.23	1.04	0.78	0.62	49.53	0.82	0.81	1.01	0.89	0.50	0.98	0.90	0.85	0.73	3.17	0.002	0.001
71	0.86	0.81	1.25	0.64	0.46	1.10	0.79	1.15	0.86	1.41	0.97	0.80	1.18	0.74	0.79	0.158	0.250
72	1.20	0.75	0.97	0.51	2.62	1.12	0.64	0.91	0.72	1.37	1.16	0.69	0.92	0.61	1.95	0.415	0.041
73	1.02	1.01	0.96	0.66	1.70	1.01	0.92	0.87	0.83	1.97	1.02	0.96	0.90	0.74	1.95	0.918	0.436

Individual CPT parameters of 73 subjects for both experimental sessions separately and pooled (full model) are shown. Additionally, bootstrapped p values for time stability in the gain and loss domain are displayed (G/L model) loss domain are presented in Table 5. To illustrate the conceptual background of the analysis, Fig. 4 graphs the likelihood profiles for two representative subjects, one of which is regarded as almost perfectly time-stable (left panel), the other as time-instable (right panel). Although the point estimate deviations for the two subjects are relatively similar (i.e., they show similar deviations of both α and δ^+), the judgment of time stability differs considerably.[21] This can be seen by comparing the likelihood profiles in the upper two rows of Fig. 4. All likelihood profiles show the above-described interaction effects between α and δ^+. Evidently, the likelihood profiles of the subject being classified as time-instable differ substantially.

The variation distance (VD) of likelihood topographies extends the graphical approach to the measurement of time stability and thus represents a visualization of the LR test. It is defined as the distance between two normalized likelihood profiles. For this purpose we normalize the sum of likelihoods over the entire estimation range (α, δ^+ from 0.2 to 2). In our analysis, the VD quantifies the relative overlap of first and second session's profiles. A value of zero implies that likelihood profiles are identical, and hence indicates perfect time stability. A VD of one represents disjoint likelihood profiles and an absolute lack of time stability. Figure 4e and f shows overlapping likelihoods as an illustration of the VD for the same two aforementioned representative subjects. Apparently, the subject being classified as time-stable shows much higher consistency. Her overlapping likelihood greatly exceeds that of the time instable subject. The respective VDs are 37.5% for the time-stable and 98.8% for the time-instable subject.

The results for the full model, which classifies subjects according to the stability of their overall decision-making behavior, are very similar. According to the LR test 24, out of the 73 subjects (33%) show a temporal instability (at a p value of 5%). In other words, a third of our subjects are significantly instable PT decision makers.

6 Conclusion

Prior research has focused on the stability of preferences when these are characterized by single-parameter EUT models. Measuring the time stability of multi-parameter decision models such as (cumulative) PT, where the risk attitude is determined by the shape of the value function as well as the shape of the probability weighting function and their interaction, is far more complex. We show that these interaction effects not only impede the elicitation of PT parameters at a single point in time (flat maxima in the estimation procedure), but also highlight the necessity for a holistic analysis of time stability. We present a methodological framework to analyze the temporal stability of preferences in a multi-parameter model that accounts for these interaction effects among parameters. Applying our framework, we are able to classify subjects according to the time stability of their decision-making behavior under PT.

In our experimental study with 86 subjects we elicit parameters within the range of related studies. We find the majority of participants show characteristic PT patterns,

[21] The two examples are subjects 23 (time-instable) and 26 (time-stable), see Table 5.

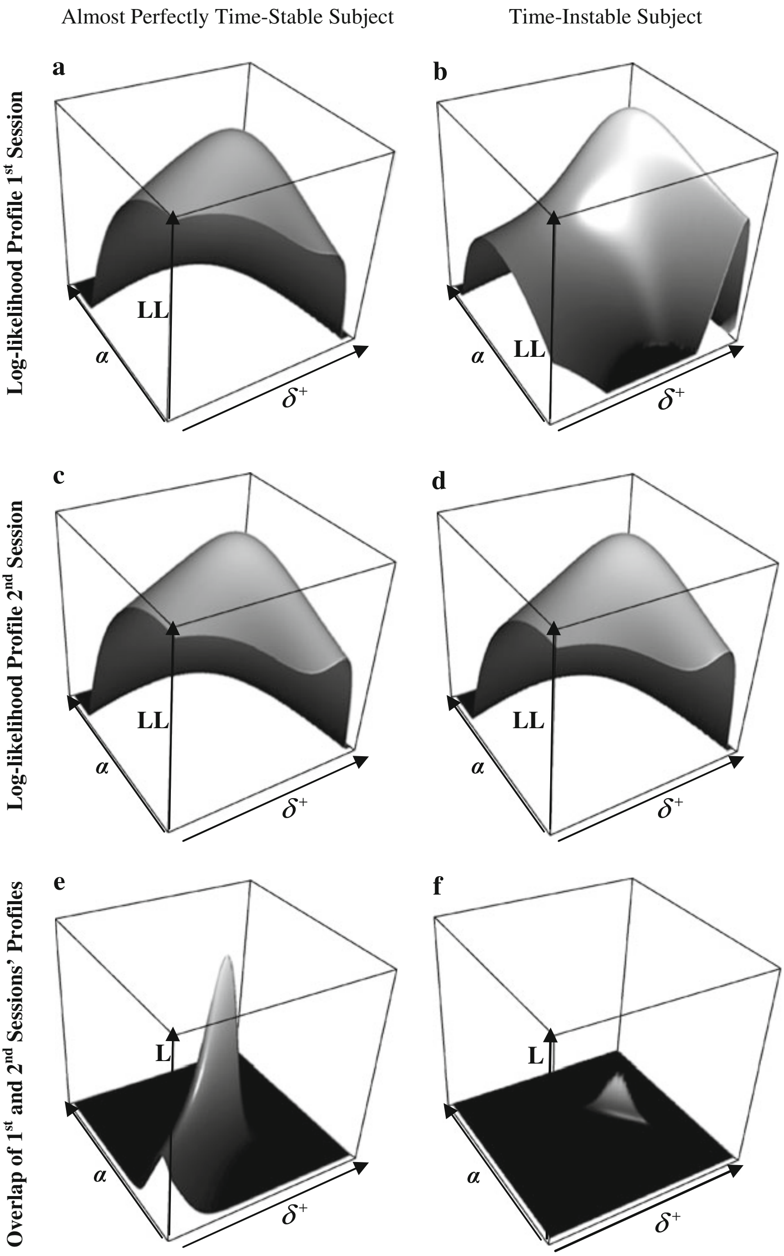


Fig. 4 Log-likelihood topographies for two representative subjects. The figure illustrates the G/L model's log-likelihood (LL) topographies for two representative subjects. **a**, **b** Based on the stated CEs of the first experimental session; **c**, **d** based on second session results. The resulting overlap, presented in **e** and **f**, is calculated by taking the element-wise (likelihood!) minimum over both sessions. Please note, for optical reasons, the log-likelihood profiles (**a**–**d**) are cut off at −58, i.e., all values below this threshold are displayed as if they were lying exactly at this level. Profiles in **e**, **f** are not cut off as they display likelihood minimums

i.e., diminishing value sensitivity, loss aversion and an inversely S-shaped probability weighting function. Relating to time stability, our results underline that PT is well-suited to describe preferences on an aggregated level as we observe remarkably stable median parameter estimates. With regard to individual preferences, however, our study reveals that one-third of the subjects show significant instability over time. Two caveats apply to our experimental analysis. First, it has to be kept in mind that the proportion of participants being classified as time-instable depends on the size of the error term that accounts for noise in the individuals' answers. Second, a time lag of 1 month might not be sufficiently large to reveal temporal instabilities. Thus, the experimental results can only provide initial evidence for temporal (in)stability of PT preferences. Yet, our analysis demonstrates how the interaction effects within the PT parameter elicitation procedure unfold.

An interesting field for future research will be to learn more about the reasons why we observe such high instability at the individual level. To discriminate more exactly between noise and temporal instability, an extended experimental design, featuring elicitation sessions with and without time delay, could be an effective means. The introduction of session delays exceeding 1 month represents an additional option to learn more about the underlying dynamics of the changes in decision-making behavior.

Acknowledgements We are indebted to Maik Dierkes, Craig Fox, Alexander Klos, Krzysztof Kontek, Mark Trede, and participants of the SPUDM22 conference in Rovereto (Italy, 2009), the Thurgau Experimental Economics Meeting in Kreuzlingen (Switzerland, 2010), the FUR conference in Newcastle (Great Britain, 2010), the Finance Center and econometrics research seminars at the University of Münster (Germany) for valuable comments and insights as well as to Sebastian Dyck and Martin Vanauer for their assistance in running the experiments.

Appendix A: Instructions (translated from German)

You will now be shown the instructions for the second part of the experiment. This part will take approx. 30–40 min (incl. the time necessary to read these instructions).

In this part of the experiment 30 lotteries will be presented to you. You will be asked to state your certainty equivalent (CE) for each of these lotteries. The CE is the amount that leaves you exactly indifferent between that amount and the lottery. For lotteries with positive outcomes the CE is exactly the amount for which it does not matter to you whether you receive it or are allowed to take part in the lottery. For lotteries with negative outcomes the CE is exactly the amount for which it does not matter to you whether you pay it or have to take part in the lottery. Let's look at two examples for clarification.

First example: assume that you are presented with the following gain lottery: with a probability of 50% you will win 50€, and with a probability of 50% you will receive nothing. You have to state your CE for this lottery. Let us assume you state a CE of 20€. This would imply that it does not matter to you if you receive the 20€ for sure or are allowed to take part in the lottery. Furthermore it would imply that:

- For every amount of money larger than 20€, you would prefer the certain money amount. Hence, you would prefer receiving 21€ over being allowed to take part in the lottery.

- For every amount of money smaller than 20€, you would prefer the lottery. Hence, you would prefer being allowed to take part in the lottery over receiving 19€.

Please note that the 20€ mentioned above are just an example and should by no means be considered as a correct answer. There are no correct or false answers. Your stated CE only has to be between 0€ and 50€. Besides that, only your personal preferences determine your CE.

Second example: assume that you are presented with the following loss lottery: with a probability of 50% you will lose 50€, and with a probability of 50% you won't lose anything. You have to state your CE. Let us assume you state a CE of −30€. This would imply that it does not matter to you if you have to pay 30€ for sure or have to take part in the lottery. Furthermore, this would imply that:

- For every amount of money larger than −30€ (i.e., losses smaller than 30€), you would prefer the certain money amount. Hence, you would prefer paying 29€ over having to take part in the lottery.
- For every amount of money smaller than −30€ (i.e., losses larger than 30€) you would prefer the lottery. Hence, you would prefer having to take part in the lottery over paying 31€.

Please note that the −30€ mentioned above are also just an example. The CE that you state only has to be between −50€ and 0€, depending on your own preferences.

A note on negative lotteries: naturally, your CEs have to be negative. Please have this in mind when entering your CE (i.e., don't forget to type the "–" in front of the number).

At the end of the experiment's second part, you will be presented with mixed lotteries. These are lotteries that will result either in a gain or a loss for you (thus, one outcome is positive, the other outcome is negative). Please state your CE here, too. For the mixed lotteries, the sign is very important. Therefore, please note when entering your CE:

- If your CE is negative you will be indifferent between paying that amount and taking part in the lottery. In this case, please enter a negative value.
- If your CE is positive you will be indifferent to either receiving that amount or taking part the lottery. In this case, please enter a positive value. Thus, for the mixed lotteries your CE might be positive in some cases and negative in others.

You will probably have to take some time to determine which amount makes you indifferent towards a lottery. It is very important that you take enough time to think about your decision carefully.

It is essential for us that you think about your decisions carefully and state your true preferences. On the one hand we try to draw conclusions for our research from your decisions. Hence, you support the chair and your department. On the other hand your monetary compensation depends on your decisions. The BDM-mechanism (which you already know from the first part of the experiment) will again ensure that you answer according to your true preferences.

Again, the objective of the BDM-mechanism is that your stated CE matches your true CE. Stated differently: by applying the BDM-mechanism it cannot be beneficial for you to state a CE (for strategic or other reasons) that is higher or lower than your true CE. The BDM-mechanism in this part of the experiment works as follows: a random number (i.e., particularly independent of your given CE) is drawn between the lower and the higher outcome of the lottery. When comparing this random number to your given CE two cases can be distinguished:

- This random amount is larger than (or equal to) your stated CE. In this case, you receive the randomly drawn amount (not your CE!).
- The random amount is smaller than your given CE. In this case, you will take part in the lottery.

Please note that you cannot change your CE after the random number has been drawn. Whether you take part in the lottery or receive the amount depends only on whether your initially stated CE is higher or lower than the randomly drawn amount. On the next page you will see why this mechanism ensures that you should always state your true CE.

An example shall clarify why giving your true CE is always in your best interest: Assume your true CE is 20€, but (for whatever reason) you stated a CE of 30€. If the randomly drawn amount is 25€, for example, you will take part in the lottery (because the amount is lower than your stated CE). You actually would have preferred receiving the 25€ over taking part in the lottery, though.

Now assume that your true CE is 30€, but (for whatever reason) you stated a CE of only 20€. If the randomly drawn amount is 25€, you will receive the 25€ (because the amount is higher than your stated CE). You would actually have preferred taking part in the lottery over receiving “only” 25€, though.

In both examples, random amounts lower than 20€ or higher than 30€ would not influence the situation. In these cases, you would not have an advantage or disadvantage, regardless of whether you answer according to your true preferences or not.

These explanations are supposed to illustrate that no situation can occur in which you would have an advantage from stating a CE different from your true CE. To the contrary, you can only have a disadvantage (as described in the examples). It is therefore advisable always to state your true CE.

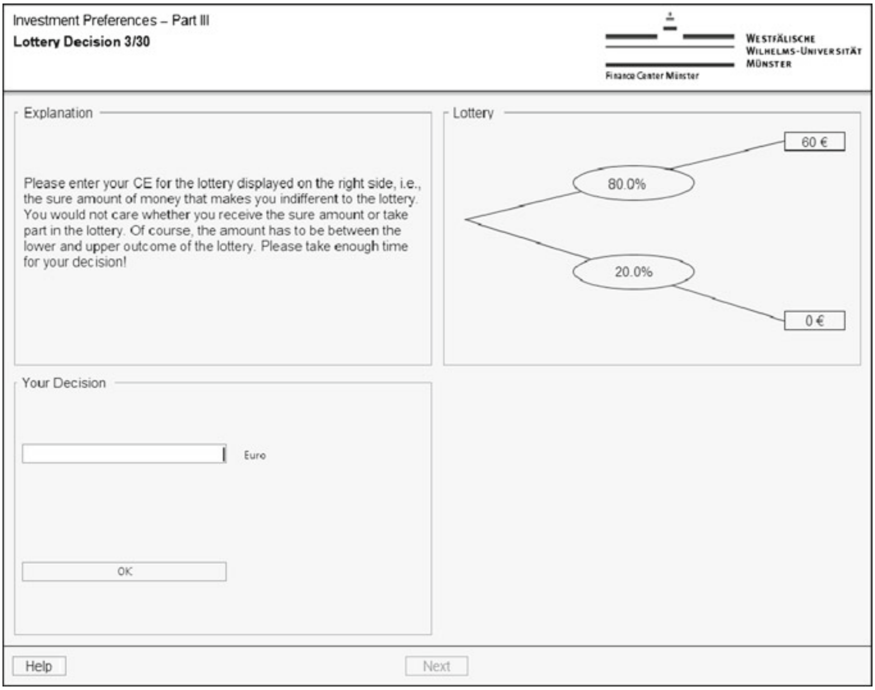
All numbers given above do not represent any recommendation. They only serve as an example. Your own CEs depend only on your preferences and don’t even have to be close to the numbers used above (and of course depend on the lotteries).

As already stated, we will draw exactly one-tenth of all experimental participants who will then receive (or possibly also lose) real money according to their decisions. For this purpose, one of your decisions (stated CE) will be chosen. It will then be determined by the BDM-mechanism whether you will receive the certain amount or take part in the lottery. In the latter case, the lottery will be played. This payment for the drawn participants will not take place before the end of the whole experiment (participants can be present).

When making decisions please always keep in mind that you are dealing with real money and that you might sustain real losses in this second part of the experiment.

If you still have any questions please don’t hesitate to ask the supervisor of the experiment. Otherwise you may click on “Next” to start the second part of the experiment.

Appendix B: Experiment example screen



An example screen as shown in the experiment (translated from German).

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science, 46*(11), 1497–1512.

Abdellaoui, M., Vossmann, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science, 51*(9), 1384–1399.

Abdellaoui, M., Barrios, C., & Wakker, P. (2007a). Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *Journal of Econometrics, 138*(1), 356–378.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007b). Loss aversion under prospect theory: A parameter-free measurement. *Management Science, 53*(10), 1659–1674.

Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion in prospect theory. *Journal of Risk and Uncertainty, 36*(3), 245–266.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008a). Lost in state space: Are preferences stable?. *International Economic Review, 49*(3), 1091–1112.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008b). Eliciting risk and time preferences. *Econometrica, 76*(3), 583–618.

Anderson, L., & Mellor, J. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty, 39*(2), 137–160.

Barseghyan, L., Price, J., & Teitelbaum, J. (2008). Are risk preferences stable across contexts? Evidence from insurance data. Working Paper, Cornell University. *American Economic Review* (forthcoming).

Bassler, J. F., MacCrimmon, K. R., & Stanbury, W. T. (1973). *Risk attitudes of business executives*. Fourth International Conference on Subjective Probability, Utility, and Decision Making. Rome, Italy.

Baucells, M., & Villasís, A. (2010). Stability of risk preferences and the reflection effect of prospect theory. *Theory and Decision, 68*(1-2), 193–211.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226–232.

Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences, 102*, 4209–4214.

Blavatskyy, P. (2006). Error propagation in the elicitation of utility and probability weighting functions. *Theory and Decision, 60*(2), 315–334.

Bleichrodt, H., Abellan-Perpinan, J. M., Pinto-Prades, J. L., & Mendez-Martinez, I. (2007). Resolving inconsistencies in utility measurement under risk: Tests of generalizations of expected utility. *Management Science, 53*(3), 469–482.

Booij, A., & van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology, 30*(4), 651–666.

Bruner, D. (2008). Changing the probability versus changing the reward. *Experimental Economics, 12*(4), 367–385.

Brunnermeier, M. K., & Nagel, S. (2008). Do Wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals. *American Economic Review, 98*(3), 713–736.

Camerer, C. (1998). Prospect theory in the wild: Evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames*. Cambridge: Cambridge University Press.

Camerer, C., & Hogarth, R. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1-3), 7–42.

Chateauneuf, A., & Cohen, M. (1994). Risk seeking with diminishing marginal utility in a non-expected utility model. *Journal of Risk and Uncertainty, 9*(1), 77–91.

Dave, C., Eckel, C., Johnson, C., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Working Paper*, University of Texas at Dallas, University of Arizona, University of Massachusetts Amherst.

Deck, C., Lee, J., Reyes, J., & Rosen, C. (2009). Measuring risk attitudes controlling for personality traits. *Working Paper*, Florida International University.

Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. (2009). Individual risk attitudes: Measurement, determinants and behavioral consequences. Working Paper, Maastricht University. *Journal of the European Economic Association* (forthcoming).

Eckel, C., & Wilson, R. (2004). Is trust a risky decision?. *Journal of Economic Behavior & Organization, 55*(4), 447–465.

Erner, C., Klos, A., & Langer, T. (2010). Can prospect theory be used to predict investor's willingness to pay? *Working Paper*, University of Muenster.

Goldstein, D., Johnson, E., & Sharpe, W. (2008). Choosing outcomes versus choosing products: Consumer-focused retirement investment advice. *Journal of Consumer Research, 35*(3), 440–456.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*(1), 129–166.

Grable, J., & Lytton, R. (2001). Assessing the concurrent validity of the SCF risk tolerance question. *Financial Counseling and Planning, 12*(2), 43–53.

Grayson, C. (1960). *Decisions under uncertainty: Drilling decisions by oil and gas operators*. Boston: Harvard University Graduate School of Business.

Guiso, L., & Paiella, M. (2006). The role of risk aversion in predicting individual behavior. In P.-A. Chiappori & Christian Gollier (Eds.), *Insurance: Theoretical analysis and policy implications*. Boston: MIT Press.

Harbaugh, W., Krause, K., & Vesterlund, L. (2007). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal, 120*(545), 595–611.

Harrison, G., Johnson, E., McInnes, M., & Rutström, E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters, 1*(1), 31–35.

Hershey, J., Kunreuther, H., & Schoemaker, P. (1982). Bias in assessment procedures for utility functions. *Management Science, 28*(8), 936–954.

Hershey, J., & Schoemaker, P. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science, 31*(10), 1213–1231.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655.

Holt, C., & Laury, S. (2005). Risk aversion and incentives: New data without order effects. *American Economic Review, 95*(3), 902–912.

Isaac, M., & James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty, 20*(2), 177–187.

James, D. (2007). Stability of risk preference parameter estimates within the Becker-DeGroot-Marschak procedure. *Experimental Economics, 10*(2), 123–141.

Kruse, J., & Thompson, M. (2003). Valuing low probability risk: Survey and experimental evidence. *Journal of Economic Behavior & Organization, 50*(4), 495–505.

Lauriola, M., Levin, I., & Hart, S. (2007). Common and distinct factors in decision making under ambiguity and risk: A psychometric study of individual differences. *Organizational Behavior and Human Decision Processes, 104*(2), 130–149.

Luce, D. (1999). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malmendier, U., & Nagel, S. (2010). Depression babies: Do macroeconomic experiences affect risk-taking? Working Paper, NBER. *Quarterly Journal of Economics* (forthcoming).

McGlothlin, W. (1956). Stability of choices among uncertain alternatives. *The American Journal of Psychology, 69*(4), 604–615.

Morone, A. (2009). On price data elicitation: A laboratory investigation. *Working Paper*, Munich Personal RePEc Archive.

Nosic, A., & Weber, M. (2008). How risky do I invest: The role of risk attitudes, risk perceptions and overconfidence. *Decision Analysis, 7*(3), 282–301.

Sahm, C. (2007). How much does risk tolerance change? *Working Paper*, Federal Reserve Board.

Schechter, L. (2007). Risk aversion and expected-utility theory: A calibration exercise. *Journal of Risk and Uncertainty, 35*(1), 67–76.

Schmidt, U., & Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science, 54*(1), 208–216.

Sharpe, W., Goldstein, D., & Blythe, P. (2000). The distribution builder: A tool for inferring investor preferences. *Working Paper*, Stanford University.

Slovic, P. (1972). Information processing, situation specificity, and the generality of risk-taking behavior. *Journal of Personality and Social Psychology, 22*(1), 128–134.

Smidts, A. (1997). The relationship between risk attitude and strength of preference: A test of intrinsic risk attitude. *Management Science, 43*(3), 357–370.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32*(2), 101–130.

Swalm, R. (1966). Utility theory—insights into risk taking. *Harvard Business Review, 44*(6), 123–136.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

van de Kuilen, G., & Wakker, P. (2009). The midweight method to measure attitudes towards risk and ambiguity. *Working Paper*, Tilburg University, Erasmus University.

Vlaev, I., Chater, N., & Stewart, N. (2009). Dimensionality of risk perception: Factors affecting consumer understanding and evaluation of financial risk. *Journal of Behavioral Finance, 10*(3), 158–181.

von Gaudecker, H., van Soest, A., & Wengstrom, E. (2008). Selection and mode effects in risk preference elicitation experiments. *Working Paper*, Tilburg University, Center for Economic Research.

Wakker, P. (2010) *Prospect Theory: For risk and ambiguity.* Cambridge University Press.

Weber, E., Blais, A. -R., & Betz, N. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*(4), 263–290.

Wehrung, D., MacCrimmon, K., & Brothers, K. (1984). Utility assessment: Domains, stability, and equivalence procedures. *INFOR, 22*(2), 98–115.

Wolf, C., & Pohlman, L. (1983). The recovery of risk preferences from actual choices. *Econometrica, 51*(3), 843–850.

Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In N. Harvey & D. Koehler, *The Blackwell handbook of judgment and decision making*. Oxford: Oxford University Press.