WORKING PAPER

CHOICE LISTS AND 'STANDARD PATTERNS' OF RISK-TAKING

Ranoua Bouchouicha
Jilong Wu
Ferdinand M. Vieider

May 2023
2023/1068

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

Choice lists and 'standard patterns' of risk-taking*

Ranoua Bouchouicha[1], Jilong Wu[1], and Ferdinand M. Vieider[1,2]

[1] *RISLαβ, Department of Economics, Ghent University*
[2] *RISLαβ Africa, AIRESS, University Mohammed VI Polytechnic*

May 4, 2023

Abstract

The *fourfold pattern* of risk attitudes has been called 'the most distinctive implication of prospect theory'. It constitutes the central mechanism by which prospect theory (PT) explains the coexistence of gambling and insurance. Here, we compare risk-taking patterns obtained from certainty equivalents (*CEs*) to risk-taking patterns observed when presenting all single choices contained in the CE lists one-by-one in a binary choice setup. Choices obtained from CEs indicate a clear fourfold pattern. Binary choices, on the other hand, indicate risk aversion for small probability gains, and risk seeking for small probabilities losses—the opposite of what is predicted by the fourfold pattern. The use of CEs to measure PT parameters is often justified based on the fact that they avoid endogenous reference points, which have been documented by comparing CEs to probability equivalents (*PEs*). We show that loss aversion in a PT model can actually not account for this discrepancy, since the gap between CEs and PEs requires different loss aversion coefficients for each PE task. Our results thus question the applicability of PT beyond the restrictive realm of CEs, which are arguably a poor proxy for most real-world decisions.

1 Motivation

A number of 'standard patterns' have been documented in the literature on decision making under risk and uncertainty. The *fourfold patterns* of risk attitudes—risk seeking for small probability gains and large probability losses, risk aversion for small probability losses and large probability gains—constitutes one of the central features of prospect theory (PT; Kahneman and Tversky, 1979). This pattern constitutes the mechanism by which the model accounts for the coexistence of gambling and insurance—a foundational issue in decision making under risk (Vickrey, 1945; Friedman and Savage, 1948;

*Ranoua Bouchouicha and Ferdinand Vieider gratefully acknowledge funding from the Research Foundation—Flanders (FWO) under the project "Causal Determinants of Preferences" (G008021N). We are indebted to Olivier L'haridon and Ryan Oprea for helpful comments and discussions. All errors remain our own.

Markowitz, 1952). Tversky and Kahneman (1992) refer to the fourfold patterns as "the most distinctive implication of prospect theory" (p. 306).

A model, however, ought to be judged by its *predictive* performance. Here, we test PT's predictive performance on its home turf—choices between binary prospects and sure outcomes. We compare risk-taking patterns obtained based on choice lists to elicit certainty equivalents (*CEs*) to patterns obtained by presenting the single choices contained in the lists in a binary choice setup, where only a single choice is shown on each screen. The results we obtain differ widely between the two choice contexts. The list format produces the 'standard' fourfold pattern documented in the PT literature using CEs (e.g., Tversky and Kahneman, 1992; Gonzalez and Wu, 1999; Bruhin, Fehr-Duda and Epper, 2010; L'Haridon and Vieider, 2019). In binary choices, however, risk seeking for small probability gains and risk aversion for small probability losses disappear, while likelihood sensitivity increases. That is, we document not only large *quantitative* differences between the two contexts, but also *qualitative* differences—the fourfold pattern, which is strong for CEs, disappears in binary choice.

The intimate marriage between PT and CEs can be traced back to a seminal contribution by Hershey and Schoemaker (1985). Comparing risk attitudes inferred from CEs to risk attitudes inferred from probability equivalents (*PEs*), the authors documented a substantial increase in risk aversion using the latter. They explained this difference by loss aversion relative to an endogenous reference point constituted by the unvarying sure amount in PEs (see also Vieider, 2018). This arguably contributed in no small part to Tversky and Kahneman's (1992) choice to adopt CEs to measure the PT functionals. In particular, the feature of CEs whereby the sure amount varies in a list while the prospect is kept fixed is supposed to exogenously fix the reference point to 0—an essential condition for the identification of all of PT's separate components.

We thus further compare the patterns obtained from CEs to patterns obtained from PEs. Depicting the choice patterns using nonparametric decision weights in a dual-EU setup, we find the resulting functions to cross at large probabilities. Loss aversion coefficients needed to account for the discrepancy for different PE tasks range from about 6 (string loss aversion) to 0.13 (strong gain seeking), both of which are implausible values based on the accumulated evidence from over 600 existing data points (Brown, Imai, Vieider and Camerer, 2022). Contrary to popular wisdom, loss aversion with respect to the sure amount in PEs can thus *not* account for the difference between CEs and PEs.

One might be tempted to blame the differences between lists and binary choices on increased noise in binary choice. A close look at the data, however, reveals that the patterns we document eschew such simple explanations. While it is true that we find increased inconsistencies 'within lists' in binary choice, choice inconsistencies are on average concentrated just below the point of expected-value equivalence, thus pointing to a region of indifference as documented by Cubitt, Navarro-Martinez and Starmer (2015) and Agranov and Ortoleva (2017), rather than purely random choice patterns. What is more, measures of preference inconsistency 'between lists' akin to first-order stochastic dominance violations indicate higher levels of noise in choice lists. Arguably, binary choices thus produce *more* rather than *less* coherent choice patterns, although they also reveal some uncertainty about the precise point of indifference.

This study further contributes to a series of recent papers challenging PT's ability to account for empirical paradoxes that were once considered to fall within its remit. Sydnor (2010) used a numerical calibration exercise to show that PT parameters as measured in typical experiments are unable to account for the widespread overinsurance of modest risks. Oprea (2022) showed that the fourfold pattern observed with CEs is reproduced in choices between deterministic options that are represented with the same level of complexity as in choice lists under risk. Using a large representative subject pool, Chapman, Dean, Ortoleva, Snowberg and Camerer (2023) document that different measures of loss aversion are unrelated to the gap between willingness-to-accept and willingness-to-pay. Our contribution constitutes a possibly even bigger challenge to PT, since it suggests that—as long as we care about the predictive ability of a model beyond statistically fitting functional forms to data *ex post*—PT ought to be considered a theory based on CEs and made only to explain patterns arising from CEs. This creates serious questions about the applicability of PT to real word choice problems, which arguably resemble the binary choice setup much more than they do choice lists.

Context effects for measurements of risk attitudes have a long history in economics and psychology, going back at least to Slovic (1964), and periodically resurfacing in the literature under different forms and with different focus (Hershey and Schoemaker, 1985; Crosetto and Filippin, 2015; Vieider, Lefebvre, Bouchouicha, Chmura, Hakimov, Krawczyk and Martinsson, 2015; Mata, Frey, Richter, Schupp and Hertwig, 2018; Zhou and Hey, 2018). The power of the context effects we present here rests in the observation that we do not change the underlying elicitation format, but only whether choices are bundled into lists or not. We furthermore exclusively use lists involving certainty in one of the outcomes, which creates ideal conditions for the observation of likelihood distortions (Hershey, Kunreuther and Schoemaker, 1982; McCord and de Neufville, 1986).

Harbaugh, Krause and Vesterlund (2010) present what is likely the test that is closest to our contribution. Using pricing tasks, they reproduce the 'standard' fourfold pattern. Using binary choices between the same prospects and their expected values, however, they document patterns "indistinguishable from random choice" (p. 595). Our study differs from theirs in two major ways. One, in our case the choice format is truly *identical* between treatment conditions. Two, by presenting the entirety of choices contained in the lists in the binary choice setup, we can paint a much richer picture of the resulting choice patterns. This latter aspect is indeed crucial, and leads us to reach very different conclusions—far from being random, the choice patterns documented from binary choice are very coherent, albeit very different from the choice patterns found using CEs.

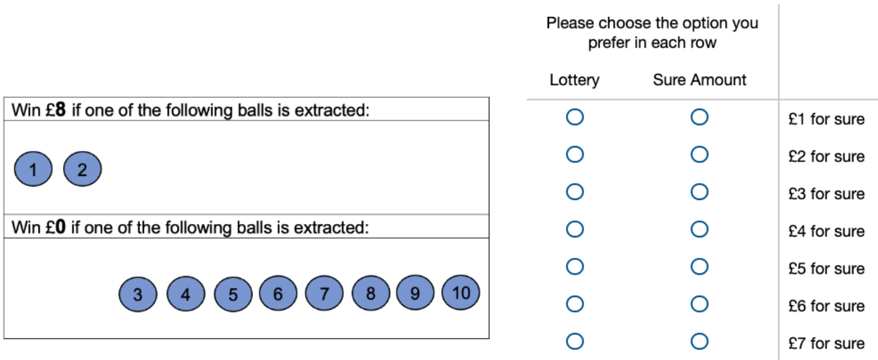
To be sure, PT parameters have been measured using elicitation formats other than CEs. Arguably, however, such elicitations have produced quantitatively and qualitatively different results, which have rarely been discussed explicitly, not least because the absence of experimental comparison treatments made it impossible to draw strong inferences on any differences. For instance, Abdellaoui (2000) presents a parameter-free measurement of the PT functionals where probability weighting is identified from PEs. The lists are, however, filled in by a bisection procedure, which replies on binary choice. Abdellaoui, Kemel, Panin and Vieider (2019) present weighting functions based on the choice list format of Holt and Laury (2002), again filled in by a bisection procedure. None of these papers finds risk taking for small probability gains. Other papers based on the same type of list without implementing bisection have typically documented S-shaped functions indicating *under*-weighting of small probabilities, and *over*-weighting of large probabilities (Andersen, Harrison, Lau and Rutström, 2014).

Some papers have also compared choice lists and binary choices from the same list, but these investigations were confined to a single list. Lévy-Garboua, Maafi, Masclet and Terracol (2012) compared choices in the task of Holt and Laury (2002) to the underlying binary decisions, and documented higher levels of risk aversion and noise in binary choices. Freeman, Halevy and Kneeland (2019) also compared risky choices obtained from a choice list to risky choices in a *single* binary choice, and found significantly more risk aversion in binary choice. They ascribed this effect to the random incentive mechanism—when incentives are random, they take away the nominal certainty from sure outcomes, which may result in increased risk seeking. Note that such a mechanism cannot account for our results, since the number of choices and hence the random incentives are constant across our treatments (see also Freeman and Mayraz, 2019, for a reinterpretation of the previous explanation). Our approach differs from these papers by presenting a whole series of choice lists. Importantly, this allows us to reach conclusions about the predictive performance of prospect theory, which are not discussed in these related papers. This also allows for a more nuanced assessment of noise, where the conclusions in our examination depend crucially on the type of noise being examined.

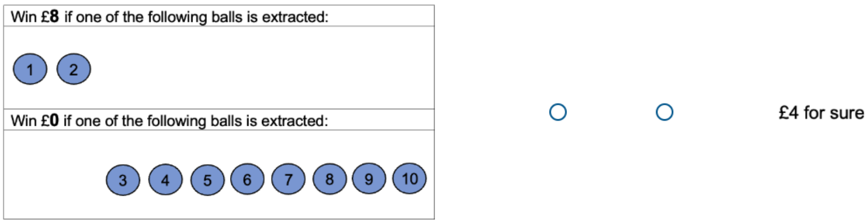
We remain explicitly agnostic as to the specific drivers of the context effects themselves. Candidate mechanisms include anchoring at the mid-point of a list (Andersson, Tyran, Wengström and Holm, 2016; Vieider, 2018), as well as more sophisticated models of contextual preferences. For instance, a specific class of range normalization models have modelled the normalization of choice options by the specific context provided by the two choice options (Louie, Grattan and Glimcher, 2011). Peterson, Bourgin, Agrawal, Reichman and Griffiths (2021) documented evidence for such context effects using a large data set and applying machine learning techniques. The results furthermore point to the promise of models of noisy coding when it comes to explaining choice behaviour under risk (Khaw, Li and Woodford, 2021; Vieider, 2021). Such models start from noisy mental signals about the specific characteristics of a choice situation. Early-generation models have exclusively modelled binary choices. However, adding signals about the context—such as e.g. the range of a choice list—could yield systematic predictions of how different information may change behaviour in otherwise identical choice problems, such as documented here.

2 Experiments

All experiments included in this paper were conducted online on Prolific UK within a short time span in the winter of 2022/23. Instructions were provided in short videos, which provided a machine-generated voice-over to slides illustrating the experimental tasks. Experiments involved identical choices packaged either in choice lists, or in presented in a binary choice setup. Figure 1 shows screenshots of what the choice environment looked like for certainty equivalents (top), and for the extrapolated binary choice tasks (bottom). Screens for other tasks looked similar, and are not shown. Importantly, we made sure that any visual display would be the same across the choice lists and binary choice formats, to avoid introducing any confounding factors. While the choice lists necessarily present the choices in an orderly fashion grouped into lists, the binary choices are presented in random order across screens.



(a) Certainty Equivalent (CE) Condition



(b) Binary Choice (BC) Condition

Figure 1: The screenshot of two choice conditions

Screenshots from the certainty equivalent experiment for gains. Panel (a) shows a screenshot of a typical choice list for a prospect yielding £8 with a 20% probability, or else 0. Panel (b) shows one binary choice extracted from that same list as it was presented in the binary choice condition.

Certainty equivalents for gains. We used a total of 21 certainty equivalents (a complete list is provided in Table 1 below), systematically varying probabilities p, and both the upper outcome x and lower outcome y in a systematic fashion between lists. The sure amount s changed in steps of £1 between the lower outcome and the upper outcome. In total, our stimuli consist of 274 binary choices. 327 subjects signed up for the experiment, but we dropped 26 of them who failed to correctly answer some simple comprehension checks after watching the instructional video. We thus ended up collecting valid responses from 301 individuals (CE: N=156; BC: N=145). The median subject took 40 minutes to finish the experiment (33 minutes for certainty equivalents; 50 minutes for binary choice). Each subject was compensated for their time according to

Prolific regulations. In addition, each subject had a 1/10 chance to play one randomly selected choice for real money.

Certainty equivalents for losses. We also elicited certainty equivalents for losses. The setup was identical to the one used for gains, except with negative instead of positive amounts. We implemented the experiment using hypothetical payoffs. Tests of the effect of real incentives are inconclusive. Particularly in the loss domain, real losses are typically deducted from an initial endowment, since implementing real losses falls foul of ethical guidelines. If (some) subjects integrate the endowment with the payouts, then implementing real incentives may actually *distort* measured behaviour and bias the inferences drawn, instead of providing true loss incentives. After excluding 10 subjects who did not pass some very basic comprehension tests, we were left with 201 subjects providing valid responses (CE: N=98; BC: N=103). A typical subject took 42 minutes to complete the experiment with 31 minutes for the CE condition, and 50 minutes for the BC condition.

Probability equivalents (gains only). The experiment eliciting probability equivalents for gains was conducted in a similar way to the one eliciting certainty equivalents for gains, except that probabilities varied within a choice list instead of the sure amount. The stimuli thus fixed the outcomes of the prospect, x and y, and the sure amount s. The stimuli were designed to resemble the ones for CEs from the perspective of an expected value maximizer (Table 4 provides a complete list). The probability varied within each given list between 0 and 1 in steps of 0.05. Subjects obtained a fixed participation fee of £7, and had a 1/10 chance to play one of their choices for real money. We collected valid responses from 187 subjects (PE: N=93; BC: N=94), with a median subject taking 34 minutes to complete the experiment.

3 Results

3.1 Certainty equivalents and 'the fourfold pattern'

We start by examining the certainty equivalent experiment for gains. Figure 2 shows some representative patterns emerging from the experiment. Panel A plots the probability of winning in a prospect offering £24 or else 0 against the choice proportion for the risky option, which under well-behaved preferences will approximate a normalized certainty equivalent.[1] Note that, here as well as throughout the paper, concepts such as 'certainty equivalents' have an inherently stochastic interpretation. Since we allow for multiple switching in choice lists and inconsistencies in binary choice, there is no guarantee that there will be a one-to-one mapping between choices and switching points. The measures we use are thus generated by tallying up choices for the risky option. As we will show farther below, however, on average the choice patterns are regular and well-behaved, so that this does not bias our inferences in any way.

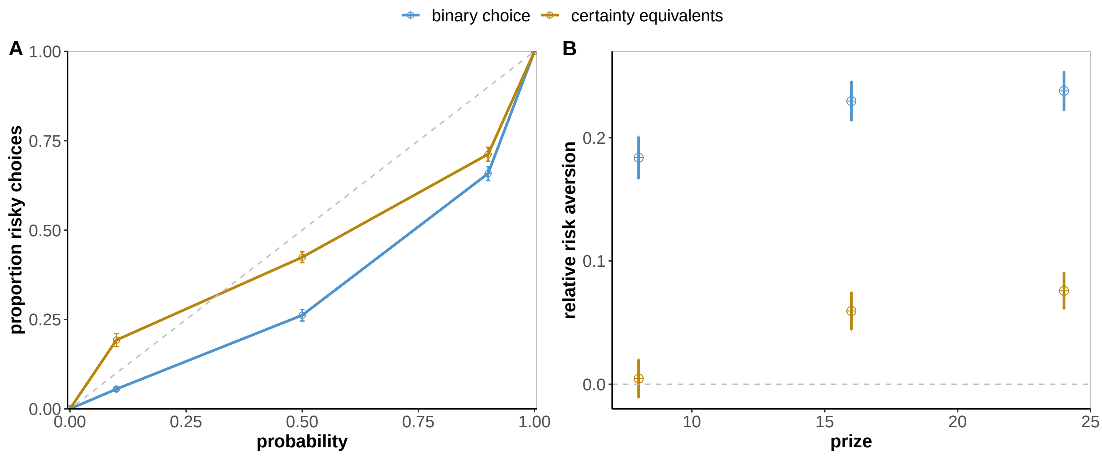


Figure 2: Nonparametric risk taking measures for certainty equivalents and binary choice for gains

Comparison between choice proportions for the risk option in identical choices that are either packaged in choice lists (CEs) or presented one-by-one in a binary choice setup (BC). Panel A compares how the number of choices for the risky option varies across the probability interval for prospects $(24, p; 0)$, with $p = \{0.1, 0.5, 0.9\}$. Panel B shows how a nonparametric index of relative risk aversion, $p - \widehat{s}$, changes as the stakes x increase for prospects $(x, 0.5; 0)$, for values of $x = \{8, 16, 24\}$, where $\widehat{s}$ indicates the derived "stochastic" certainty equivalence.

When the choices are collected in a choice list, the function in panel A exhibits the hallmark patterns documented in the PT literature. Small probabilities are overweighted, resulting in risk seeking. Moderate to large probabilities are underweighted, resulting in risk aversion. The combination of these two patterns results in pronounced *likelihood-insensitivity*. The picture obtained from presenting the same identical choices contained in the choice lists in a binary choice setup, however, is quite different. For one, we see a strong level effect, whereby the function we obtain is shifted downward, and thus shows a sizeable increase in risk aversion. The patterns are also qualitatively different from those obtained when using choice lists. In particular, even the smallest probability of $p = 0.1$ is now *under*weighed, resulting in risk aversion for small probability prospects. Large probabilities continue to be underweighted, and since the difference between the

[1] By 'normalized certainty equivalent' we mean in general $\frac{s-y}{x-y}$, where s in this case indicates a 'stochastic switching point'. Although our analysis is model-free, this measure can be interpreted as a decision weight in a prospect theory model with linear utility.

functions is much smaller for large probabilities, likelihood-insensitivity is much less pronounced. Using CEs, we thus replicate the 'standard' fourfold pattern documented in the PT literature based on CEs. Once we use binary choice, however, the fourfold pattern disappears and gives way to a uniformly convex function, which however keeps some modest degree of likelihood-insensitivity.

Table 1: Choice proportions of risky option by treatment for CEs over gains

Task	CE	BC	p-value	Task	CE	BC	p-value
(16, 0.2; 0)	0.27	0.11	<0.01	(17, 0.5; 4)	0.43	0.34	<0.01
(16, 0.3; 0)	0.32	0.13	<0.01	(24, 0.1; 0)	0.19	0.06	<0.01
(16, 0.5; 0)	0.44	0.27	<0.01	(24, 0.5; 0)	0.42	0.26	<0.01
(16, 0.7; 0)	0.55	0.46	<0.01	(24, 0.9; 0)	0.71	0.66	<0.01
(16, 0.8; 0)	0.62	0.56	<0.01	(24, 0.4; 12)	0.44	0.31	<0.01
(16, 0.8; 0)	0.62	0.57	<0.01	(24, 0.6; 12)	0.47	0.43	<0.01
(16, 0.1; 4)	0.20	0.11	<0.01	(8, 0.2; 0)	0.32	0.12	<0.01
(16, 0.5; 4)	0.43	0.33	<0.01	(8, 0.5; 0)	0.50	0.32	<0.01
(16, 0.9; 4)	0.70	0.68	0.15	(8, 0.8; 0)	0.67	0.57	<0.01
(15, 0.5; 4)	0.45	0.35	<0.01	**(8, 0.8; 0)**	0.69	0.59	<0.01
(16, 0.5; 5)	0.42	0.35	<0.01				

List of choice tasks with choice proportions per treatment condition. The indicated p-values are based on two-sided Wilcoxon rank sum tests. CE: certainty equivalent set up. BC: binary choice set up.

Table 1 shows the choice proportions for the two treatment conditions by task, and provides nonparametric tests on the differences in calculated choice proportions. It indicates that the patterns displayed in the graph are quite typical, with all but one comparison showing significant differences between certainty equivalents and binary choice. Note further that none of these patterns can be explained 'simply' by noise. Even though the discrete choice experiment takes longer than the CE experiment, subjects did not know about the differential lengths of the experiments ex ante, and the differences we observe are identical if we restrict our analysis of the discrete choice data only to the first half of the experiment. The following section will provide a more in-depth analysis of noise in the two treatment conditions.

Panel B in Figure 2 shows a measure of relative risk aversion, that is given by the choice proportion of the risky option subtracted from the probability of winning. Note that in the case of $y = 0$, a normalized certainty equivalent subtracted from the probability allows us to capture changes in relative risk aversion in the sense of Arrow-Pratt. We once again see the same level effect as in panel A for all stake levels. Patterns for CEs indicate the typical increasing relative risk aversion (IRRA) documented in the literature (Holt and Laury, 2002; Fehr-Duda, Bruhin, Epper and Schubert, 2010; Bouchouicha and Vieider, 2017; Di Falco and Vieider, 2022). A similar pattern of IRRA is also observed in binary choice, thus pointing to the robustness of the phenomenon.

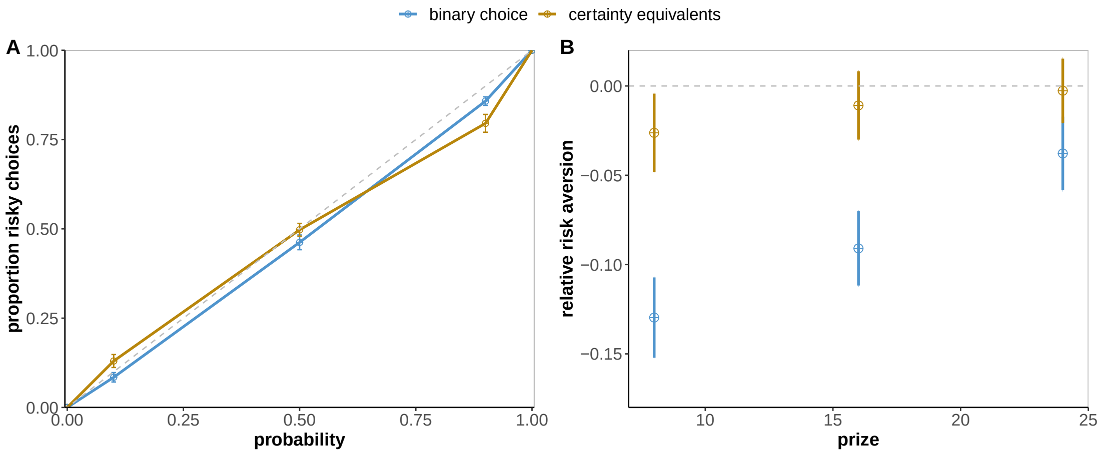


Figure 3: Nonparametric risk taking measures for certainty equivalents and discrete choice for losses

Figure 3 shows the equivalent results for losses, and Table 2 provides summary measures and tests of all choice tasks. Once again, the risk aversion for small probabilities observed in the choice list format, which conforms to the prediction of the fourfold pattern, turns into risk *seeking* in binary choice format, thus contradicting the prediction of the fourfold pattern. This pattern of slight risk seeking is indeed observed across the probability interval, whereas the function derived from CEs shows the typical inverse-S shaped pattern (the pattern is not very pronounced, but this is not unusual for losses—see e.g. L'Haridon and Vieider, 2019 for evidence from 30 different countries). Panel B shows changes in relative risk aversion with stake size. Most measures are negative, indicating a tendency towards risk seeking. Using CEs, we observe a pattern resembling the constant relative risk aversion documented for losses in previous studies (Fehr-Duda et al., 2010; Bouchouicha and Vieider, 2017). In binary choice, however, we observe clear evidence for increasing relative risk aversion. In other words, whereas using CEs the patterns for losses tend to differ from the ones for gains, as also documented in the previous literature, in BC we find convergent evidence for increasing relative risk aversion.

Table 2: Choice proportions of safe option by treatment for CEs over losses

Tasks	CE	BC	p-value	Tasks	CE	BC	p-value
(-16, 0.2; 0)	0.23	0.14	$<$0.01	(-17, 0.5; -4)	0.43	0.36	$<$0.01
(-16, 0.3; 0)	0.32	0.20	$<$0.01	(-24, 0.1; 0)	0.13	0.08	$<$0.01
(-16, 0.5; 0)	0.49	0.41	$<$0.01	(-24, 0.5; 0)	0.50	0.46	0.02
(-16, 0.7; 0)	0.64	0.72	$<$0.01	(-24, 0.9; 0)	0.80	0.86	$<$0.01
(-16, 0.8; 0)	0.69	0.78	$<$0.01	(-24, 0.4; -12)	0.34	0.26	$<$0.01
(-16, 0.8; 0)	0.69	0.78	$<$0.01	(-24, 0.6; -12)	0.49	0.44	0.05
(-16, 0.1; -4)	0.11	0.09	0.27	(-8, 0.2; 0)	0.17	0.13	0.05
(-16, 0.5; -4)	0.42	0.34	$<$0.01	(-8, 0.5; 0)	0.47	0.37	$<$0.01
(-16, 0.9; -4)	0.76	0.80	$<$0.01	(-8, 0.8; 0)	0.70	0.75	0.04
(-15, 0.5; -4)	0.44	0.34	$<$0.01	**(-8, 0.8; 0)**	0.68	0.75	$<$0.01
(-16, 0.5; -5)	0.40	0.34	$<$0.01				

List of choice tasks with choice proportions per treatment condition. The indicated p-values are based on two-sided Wilcoxon rank sum tests. CE: certainty equivalent set up. BC: binary choice set up.

3.2 Decision noise

One may be tempted to blame the difference in behaviour we have observed between choice lists and binary choice on differences in noise arising from the two formats. As we show in this section this would, however, run the risk of being overly simplistic.

We indeed observe an increased frequency of 'multiple switching' within any given task in binary choice compared to certainty equivalents, as shown in figure 4. The patterns are similar for gains, shown in panel A, and for losses, shown in panel B. While the overall switching rates are higher in binary choice, they do follow a highly regular pattern. In particular, the switching rates in binary choice exhibit a marked peak around expected value differences indicating risk aversion for gains, and slight risk seeking for losses. The switching rates for CEs indicate that some multiple switching is going on in choice lists, too, but that this multiple switching is much less frequent (11.6% of subjects in CEs versus 100% of subjects in BC for gains; and 12.0% of subjects in CEs versus 100% of subjects in BC for losses exhibit this pattern in at least one list), and again quite regular. The inconsistencies we observe within tasks are thus best interpreted as indicating regions of indifference such as documented by Cubitt et al. (2015) and Agranov and Ortoleva (2017), and consistent with the notion of cognitive uncertainty of Enke and Graeber (2019). The choice list format seems to dramatically reduce these effects. Such reduced noisiness may be seen as an asset from the vantage point of a deterministic model such as PT—an additional aspect that may have contributed to the marriage between PT and CEs. Artificially reducing this type of noise may, however, affect the proper identification of decision noise, which plays a central role in stochastic models of choice (Khaw et al., 2021; Vieider, 2021). Note furthermore that reducing noise within-list may well induce additional noisiness under different forms.

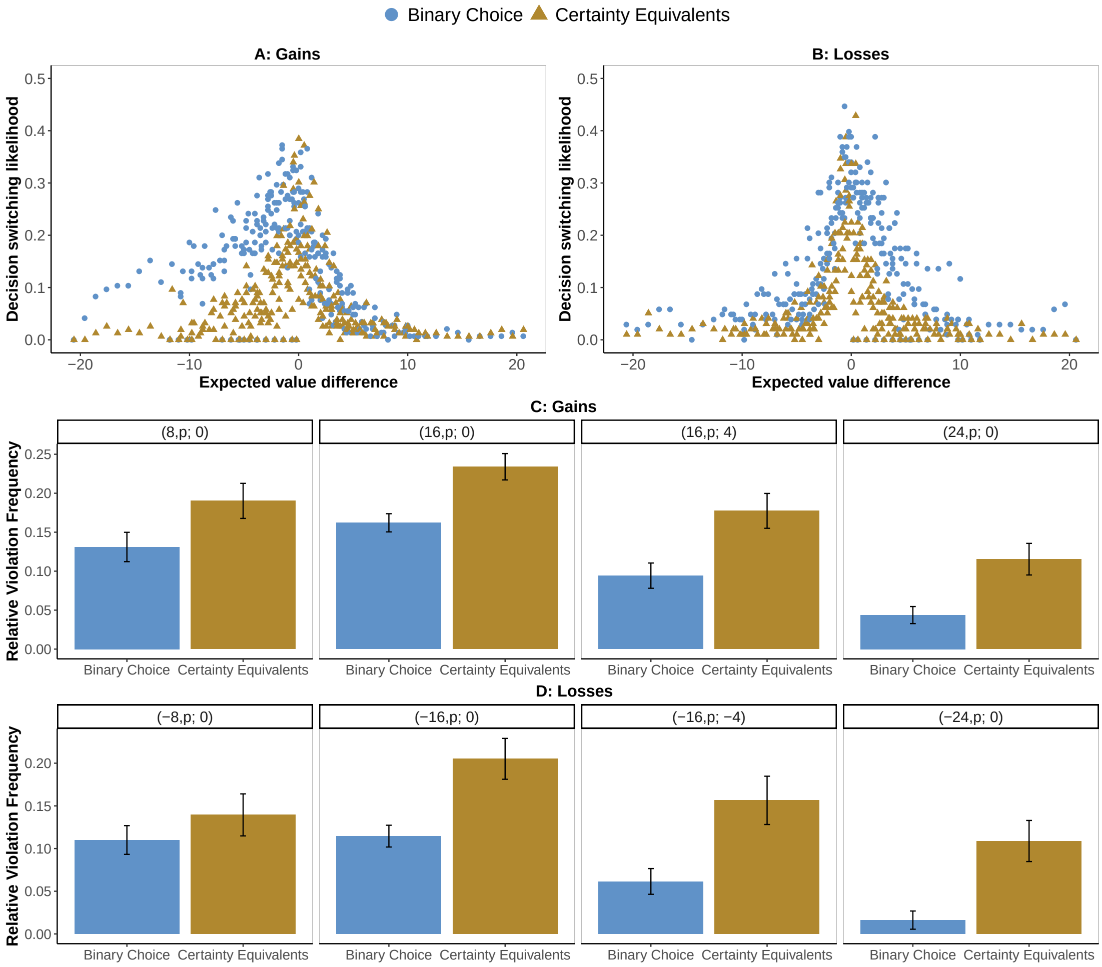


Figure 4: Noise in lists versus discrete choice

Panels A and B compare subjects' tendency to switch from one option to another as the sure amount increases. The expected value difference on the x-axis is calculated as the sure amount minus the expect value of the prospect, and the y-axis shows switching frequencies with each increase in the sure outcome. Panels C and D examine first-order stochastic dominance violations between lists as the probabilty increases for identical outcomes. These plots report the relative violation frequency, which is the violation frequency divided by the number of all possible violations.

Such alternative forms of noise do indeed show up *between* tasks in the form of first-order stochastic dominance violations. Such violations are systematically more frequent for CEs than in binary choice across all possible tasks in which stochastic dominance can be violated, as shown in panel C for gains and in panel D for losses. While the patterns we observe are clear, there are several possible explanations. One possibility

Table 3: Check of Repeated Stimuli

	Gain Domain		Loss Domain	
prospect	Direct Choice	Choice List	Direct Choice	Choice List
(8, 0.8; 0)	0.772	0.487	0.688	0.610
	(0.697, 0.831)	(0.358, 0.599)	(0.571, 0.778)	(0.468, 0.721)
(16, 0.8; 0)	0.915	0.695	0.854	0.528
	(0.883, 0.938)	(0.598, 0.764)	(0.791, 0.899)	(0.368, 0.657)

Note: The test is Pearson's Correlation test. 95% confidence interval is reported in parentheses. The result remains essentially unchanged when applying Spearman's Correlation test.

is that the context effects driving the deviation in response patterns in lists from those observed in binary choice induce more randomness in switching points, i.e. they may distract from the characteristics of each single choice and thus induce higher levels of randomness in choice between lists. In this sense, the artificial reduction of within-task switching behaviour may create some randomness in the switching point in lists.

Finally, we can examine the test-retest reliability as a measure of choice consistency. To have a measure that is consistent across settings, we repeated two entire lists, $(8, 0.8; 0)$ and $(16, 0.8; 0)$ in the CEs for gains and $(-8, 0.8; 0)$ and $(-16, 0.8; 0)$ in CEs for losses. Table 3 reports the correlation coefficients of the proportion of safe choices for the repeated lists. We find reasonably large and significantly positive correlations between responses in repeated stimuli in both conditions. The correlations are, however, much stronger in the binary choice condition. This confirms that subjects' responses are quite stable when they answer those binary choices separately, and rather less stable when packaged in a choice list format. Whatever the exact explanation, the results show that the differences between presentation formats we present defy a simple error narrative whereby binary choice is just 'noisier'.

3.3 Endogenous reference points and loss aversion

A key reason why CEs are the tool of choice to elicit PT parameters is that they allow to exogenously fix the reference point to 0. This is indeed essential if one wants to separately identify all the different components of PT, since in the presence of endogenous reference points all prospects would become mixed (see Baillon, Bleichrodt and Spinu, 2020, for a study disentangling these effects). Hershey and Schoemaker (1985) famously showed that varying a probability in a list while keeping the sure amount fixed could create a risk of endogenous reference dependence. Given that in such a case all gambles become mixed, PT can no longer be fully identified.[2] In particular, it would no longer be possible to separately identify reference-dependence and rank-dependence (i.e., loss aversion and optimism/pessimism for gains and losses). One possibility is then that in binary choice the sure outcome may also act as an endogenous reference point, which would be troublesome for the identification of the full array of PT parameters.

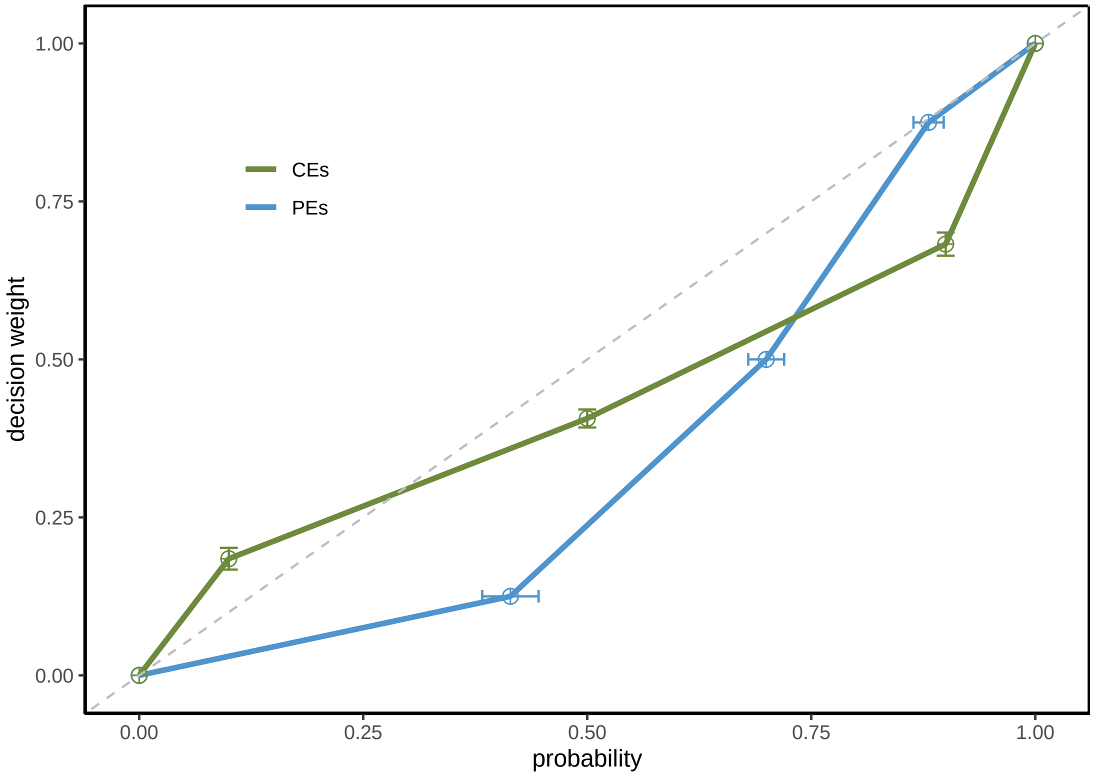


Figure 5: CEs versus PEs

Nonparametric dual-EU decision weights obtained from certainty equivalent (CE) lists and Probability equivalent (PE) lists. The patterns shown are obtained from lists offering £24 or else 0. In the CE lists, the sure amount is varied from £1 to £23 for probabilities $p = \{0.1, 0.5, 0.9\}$. In the PE lists, probabilities are varied with sure outcomes $s = \{3, 12, 21\}$.

Figure 5 compares the choice patterns obtained using CEs while varying the sure outcome within a list and probabilities between lists, to the choice patterns obtained using PEs while varying probabilities within a list and the sure outcome between lists. Both were measured in the same experimental setting, and using lists that ought to induce similar switching points for an expected value maximizer. While the CE patterns

[2] To cite but one example, Choi, Fisman, Gale and Kariv (2007) and Choi, Kariv, Müller and Silverman (2014) investigate allocations between two risky assets under budget constraints. Equal allocations, which occur frequently in the data, can be explained by either rank-dependence or loss aversion, but the two are not separately identified. This points to an over-specification of PT when it comes to its applicability to real-world choices.

exhibit the iconic inverse-S shape, the risk attitude patterns obtained from PEs appear to exhibit an S-shape over the probability interval. The stark contrast between the two risk taking patterns appears to indicate an effect whereby subjects take the range of the list being shown to them as an additional element informing the decision process.

Table 4: Choice proportions of safe option by treatment for PEs over gains

Tasks	PE	BC	p-value	Tasks	PE	BC	p-value
(16, p; 0) V.S. 2	0.39	0.52	<0.01	(24, p; 0) V.S. 3	0.41	0.54	<0.01
(16, p; 0) V.S. 8	0.72	0.82	<0.01	(24, p; 0) V.S. 12	0.71	0.82	<0.01
(16, p; 0) V.S. 13	0.86	0.93	<0.01	(24, p; 0) V.S. 21	0.90	0.96	<0.01
(16, p; 0) V.S. 13	0.84	0.94	<0.01	(24, p; 12) V.S. 17	0.50	0.40	<0.01
(16, p; 4) V.S. 6	0.37	0.25	<0.01	(8, p; 0) V.S. 1	0.36	0.49	<0.01
(16, p; 4) V.S. 10	0.64	0.71	<0.01	(8, p; 0) V.S. 4	0.70	0.82	<0.01
(16, p; 4) V.S. 14	0.80	0.86	<0.01	(8, p; 0) V.S. 6	0.83	0.92	<0.01

List of choice tasks with choice proportions per treatment condition. The indicated p-values are based on two-sided Wilcoxon rank sum tests.

Documenting similar patterns, Hershey and Schoemaker (1985) concluded that the patterns obtained using PEs could be obtained by PT with an endogenous reference point equal to the salient sure amount s, which does not change within a given list. That is, they simply rescaled the PT equation as follows:

$$u(0) = w^+(p)u(x-s) - \lambda w^-(1-p)u(s),$$

where u is a reference-dependent utility function, w^+ and w^- the probability weighting functions for gains and losses, respectively, and λ captures loss aversion. Given that we have all the elements to identify utility curvature and probability weighting from CEs, we can easily infer the loss aversion coefficient that would explain the discrepancy between CEs and PEs shown in figure 5 from the following equation:

$$\lambda = \frac{w^+(p)}{w^-(1-p)} \frac{u(x-s)}{u(s)}. \tag{1}$$

Even just eyeballing the figure tells us that we would need $\lambda \gg 1$ to explain the first PE data point to the left, $\lambda \approx 1$ for the second data point, and $\lambda < 1$ for the right-most decision weight obtained from PEs. The exact parameters will of course depend on assumptions made about functional forms and errors, but a 'standard' PT implementation (see Appendix A) yields $\lambda_{s=3} = 6.37$ [6.16 ; 6.95], $\lambda_{s=12} = 0.82$ [0.81 ; 0.84], and $\lambda_{s=21} = 0.13$ [0.12 ; 0.13]. It is thus clear that loss aversion cannot account for the discrepancy, since the loss aversion coefficient in PT is supposed to be constant. Note that, although we obtained these sepcific numbers based on the prospects paying £24 or 0 shown in figure 5, the patterns obtained for other prospects while systematically varying the sure outcome s are very similar.

4 Discussion and Conclusion

We have documented systematic differences between risk attitudes elicited in choice lists, and risk attitudes elicited in identical choices extrapolated from the lists to be randomly presented in a binary choice setup. The choice lists data reproduce the 'standard' fourfold pattern of risk attitudes documented in the prospect theory literature using certainty equivalents, and which serves as the latter's main vehicle to explain the coexistence of insurance and lottery play. The binary choice data, however, paint a rather different picture. We now find risk *aversion* for small probability gains, and risk *seeking* for small probability losses—the exact opposite of what is predicted by the 'fourfold pattern'. This constitutes a major challenge for PT, since the discrepancies occur on its home turf of choices between binary prospects and sure amounts of money.

One possible explanation, which has often been used as a justification for the use of certainty equivalents to measure prospect theory functionals, is that certainty equivalents help to exogenously fix the reference point to 0. Other tasks on the other hand may induce endogenous reference points, thus increasing risk aversion. We have shown that the discrepancy between certainty equivalent tasks and probability equivalent tasks cannot actually be explained by such an account, contrary to the prevailing view in the literature. This does of course not exclude that such endogenous reference-dependence may occur in binary choice. If that were truly the explanation of the discrepancy, however, that would point to an even more troubling conclusion. Given that real world decisions tend to resemble binary choices much more than certainty equivalents (or indeed pricing tasks—outside of auctions or experiments, it is rare to be asked to declare a maximum buying price), this would imply that prospect theory is over-specified, since this would inevitably result in collinearity between rank-dependence and loss aversion.

The patterns we have shown also defy a simple noise narrative, whereby the patterns we document under binary choice could be simply blamed on 'increased noise levels'. Comparing pricing tasks to discrete choices between gambles and their expected values, Harbaugh et al. (2010) concluded that while pricing showed the standard fourfold pattern, the binary choice tasks yielded decision patterns "indistinguishable from random choice". Our much richer choice setup paints quite a different picture. Far from being akin to random choice, discrete choice patterns seem in many ways more coherent and regular than the patterns observed in choice lists. We choose to see this as a good sign—presenting simple binary choice actually appears to focus attention on the essential aspects of the decision problem.

Appendix A: Recovering loss aversion from PE tasks

We estimate a simple aggregate PT model from the data for CEs, by letting $u(x) = \frac{1-exp(-\rho x)}{\rho}$, with different parameters for gains and losses, entered in terms of absolute amounts. Using an exponential utility specification is motivated by the IRRA patterns observed in the nonparametric data, and avoids issues in the identification of loss aversion when different utility coefficients are estimated for gains and losses using CRRA functions (Köbberling and Wakker, 2005). The probability weighting function is $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$, again with different parameters for gains and losses. We estimate the model using Bayesian techniques in Stan. The priors used for the parameters are mildly regularizing, i.e. they are uninformative in the sense of being centred on neutral values ($\rho = 0$, $\delta = \gamma = 1$), and they are diffuse, in the sense that the standard deviation is chosen in a way as to include a large range of parameters into the possible range (e.g., for γ and δ, 95% of the probability mass is allocated to the interval between 0 and 7).

Subsequently, we calculate the choice objects p, $x - s$ and s for each of the three PE tasks shown in Figure 5 (estimates obtained from other PE tasks yield similar results). We then inject the PT parameters estimated from CEs. Importantly, we do so using the entire vector of posterior draws for each of the parameters, which allows us to take the uncertainty in the parameter estimates into account, and thus to obtain credibility intervals for the estimates of loss aversion, as reported in the main text.

References

Abdellaoui, Mohammed (2000) 'Parameter-Free Elicitation of Utility and Probability Weighting Functions.' *Management Science* 46(11), 1497–1512

Abdellaoui, Mohammed, Emmanuel Kemel, Amma Panin, and Ferdinand M. Vieider (2019) 'Measuring risk and time preferences in an integrated framework.' *Games and Economic Behavior* 115, 459–469

Agranov, Marina, and Pietro Ortoleva (2017) 'Stochastic choice and preferences for randomization.' *Journal of Political Economy* 125(1), 40–68

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström (2014) 'Discounting behavior: A reconsideration.' *European Economic Review* 71, 15–33

Andersson, Ola, Jean-Robert Tyran, Erik Wengström, and Håkan J. Holm (2016) 'Risk Aversion Relates to Cognitive Ability: Preferences or Noise?' *Journal of the European Economic Association* 14(5), 1129–1154

Baillon, Aurélien, Han Bleichrodt, and Vitalie Spinu (2020) 'Searching for the reference point.' *Management Science* 66(1), 93–112

Bouchouicha, Ranoua, and Ferdinand M. Vieider (2017) 'Accommodating stake effects under prospect theory.' *Journal of Risk and Uncertainty* 55(1), 1–28

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer (2022) 'Meta-analysis of empirical estimates of loss aversion.' *Journal of Economic Literature, forthcoming*

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010) 'Risk and Rationality: Uncovering Heterogeneity in Probability Distortion.' *Econometrica* 78(4), 1375–1412

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer (2023) 'Willingness to accept, willingness to pay, and loss aversion.' Technical Report, National Bureau of Economic Research

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv (2007) 'Consistency and Heterogeneity of Individual Behavior under Uncertainty.' *The American Economic Review* 97(5), 1921–1938

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman (2014) 'Who Is (More) Rational?' *American Economic Review* 104(6), 1518–1550

Crosetto, Paolo, and Antonio Filippin (2015) 'A theoretical and experimental appraisal of four risk elicitation methods.' *Experimental Economics* pp. 1–29

Cubitt, Robin P, Daniel Navarro-Martinez, and Chris Starmer (2015) 'On preference imprecision.' *Journal of Risk and Uncertainty* 50(1), 1–34

Di Falco, Salvatore, and Ferdinand M. Vieider (2022) 'Environmental adaptation of risk preferences.' *Economic Journal, forthcoming* 132(648), 2737–2766

Enke, Benjamin, and Thomas Graeber (2019) 'Cognitive uncertainty.' Technical Report, National Bureau of Economic Research

Fehr-Duda, Helga, Adrian Bruhin, Thomas F. Epper, and Renate Schubert (2010) 'Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size.' *Journal of Risk and Uncertainty* 40(2), 147–180

Freeman, David J, and Guy Mayraz (2019) 'Why choice lists increase risk taking.' *Experimental Economics* 22(1), 131–154

Freeman, David J, Yoram Halevy, and Terri Kneeland (2019) 'Eliciting risk preferences using choice lists.' *Quantitative Economics* 10(1), 217–237

Friedman, Milton, and L. J. Savage (1948) 'The Utility Analysis of Choices Involving Risk.' *Journal of Political Economy* 56(4), 279–304

Gonzalez, Richard, and George Wu (1999) 'On the Shape of the Probability Weighting Function.' *Cognitive Psychology* 38, 129–166

Harbaugh, William T, Kate Krause, and Lise Vesterlund (2010) 'The fourfold pattern of risk attitudes in choice and pricing tasks.' *The Economic Journal* 120(545), 595–611

Hershey, John C., and Paul J. H. Schoemaker (1985) 'Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?' *Management Science* 31(10), 1213–1231

Hershey, John C., Howard C. Kunreuther, and Paul J. H. Schoemaker (1982) 'Sources of Bias in Assessment Procedures for Utility Functions.' *Management Science* 28(8), 936–954

Holt, Charles A., and Susan K. Laury (2002) 'Risk Aversion and Incentive Effects.' *American Economic Review* 92(5), 1644–1655

Kahneman, Daniel, and Amos Tversky (1979) 'Prospect Theory: An Analysis of Decision under Risk.' *Econometrica* 47(2), 263 – 291

Khaw, Mel Win, Ziang Li, and Michael Woodford (2021) 'Cognitive imprecision and small-stakes risk aversion.' *The Review of Economic Studies* 88(4), 1979–2013

Köbberling, Veronika, and Peter P. Wakker (2005) 'An index of loss aversion.' *Journal* *of Economic Theory* 122(1), 119 – 131

Lévy-Garboua, Louis, Hela Maafi, David Masclet, and Antoine Terracol (2012) 'Risk aversion and framing effects.' *Experimental Economics* 15(1), 128–144

L'Haridon, Olivier, and Ferdinand M. Vieider (2019) 'All over the map: A worldwide comparison of risk preferences.' *Quantitative Economics* 10, 185–215

Louie, Kenway, Lauren E Grattan, and Paul W Glimcher (2011) 'Reward value-based gain control: divisive normalization in parietal cortex.' *Journal of Neuroscience* 31(29), 10627–10639

Markowitz, Harry (1952) 'The Utility of Wealth.' *Journal of Political Economy* 60(2), 151–158

Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig (2018) 'Risk preference: A view from psychology.' *The Journal of Economic Perspectives* 32(2), 155–172

McCord, Mark, and Richard de Neufville (1986) '"Lottery Equivalents": Reduction of the Certainty Effect Problem in Utility Assessment.' *Management Science* 32(1), 56–60

Oprea, Ryan (2022) 'Simplicity equivalents.' *Working Paper*

Peterson, Joshua C, David D Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L Griffiths (2021) 'Using large-scale experiments and machine learning to discover theories of human decision-making.' *Science* 372(6547), 1209–1214

Slovic, Paul (1964) 'Assessment of risk taking behavior.' *Psychological Bulletin* 61(3), 220–233

Sydnor, Justin (2010) '(Over)insuring Modest Risks.' *American Economic Journal: Applied Economics* 2(4), 177–199

Tversky, Amos, and Daniel Kahneman (1992) 'Advances in Prospect Theory: Cumulative Representation of Uncertainty.' *Journal of Risk and Uncertainty* 5, 297–323

Vickrey, William (1945) 'Measuring Marginal Utility by Reactions to Risk.' *Econometrica* 13(4), 319

Vieider, Ferdinand M (2018) 'Violence and risk preference: Experimental evidence from afghanistan: Comment.' *American Economic Review* 108(8), 2366–82

Vieider, Ferdinand M. (2021) 'Decisions under uncertainty as bayesian inference.' Working Paper WP 21/1022 of Faculty of Economics and Business Administration, Ghent University, Belgium 21/1022, Ghent University, Faculty of Economics and Business

Administration

Vieider, Ferdinand M., Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk, and Peter Martinsson (2015) 'Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries.' *Journal of the European Economic Association* 13(3), 421–452

Zhou, Wenting, and John Hey (2018) 'Context matters.' *Experimental economics* 21, 723–756