Cognitive Frictions and Canonical Patterns in Risk-Taking*

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

[1] *RISLαβ, Department of Economics, Ghent University*
[2] *Haas School of Business, University of California, Berkeley*

June 15, 2026

Abstract

Using a comprehensive meta-analysis and new experiments, we show that identical risky choice options elicit systematically different behavior when presented one-by-one versus when bundled into choice lists. In particular, we find that the canonical fourfold pattern of risk attitudes in choice lists tends to disappear in binary choice, replaced by a very different pattern of risk-taking. No model of utility maximization can account for this fundamental change in behavior, given that, in our experiment, constituent choices are *identical* across choice contexts. This raises fundamental questions about the way we interpret prospect theory, which is centered on the fourfold pattern. We propose (and empirically test) a model based on noisy cognition that organizes our results, and argue that endogenizing parameters of descriptive models like prospect theory provides a promising way of improving their predictive accuracy.

1 Introduction

The "fourfold pattern of risk attitudes" is one of the most important empirical patterns in behavioral economics. Tversky and Kahneman (1992) describe it as "the most distinctive implication of prospect theory", and prospect theory (PT; Kahneman and Tversky, 1979, Tversky and Kahneman, 1992, Wakker, 2010), in turn, is unquestionably the field's most important theory of risk-attitudes. The fourfold pattern is a putative tendency for people to be risk averse for small probability losses and large probability gains, and risk seeking for small probability gains and large probability losses. For decades, this pattern has been

*Ranoua Bouchouicha and Ferdinand Vieider gratefully acknowledge funding from the Research Foundation—Flanders (FWO) under the project "Causal Determinants of Preferences" (G008021N). Previous versions of this manuscript were circulated under the titles "Is Prospect Theory Really a Theory of Choice?". We are indebted to Mohammed Abdellaoui, Ben Enke, Thomas Epper, Paul Feldman, Olivier L'Haridon, Pietro Ortoleva, Larry Samuelson, Christian Walter, and Songfa Zhong for helpful comments and discussions. All errors remain our own.

treated as a robust empirical regularity — one that PT was designed to explain. However, in an important paper, Harbaugh, Krause and Vesterlund (2010) called the universality of this pattern into question, reporting an experiment in which the fourfold pattern occurred in an explicit valuation elicitation task (using the BDM mechanism), but failed to appear when subjects are instead asked to directly choose between lotteries.

In this paper, we use multiple strands of evidence to show that while the fourfold pattern appears in direct standard choice list elicitations of certainty equivalents, it is typically replaced by very different behavior (closer to a *two*fold pattern) when subjects instead directly *choose between* lotteries. This instability of risky choice to superficial changes in the choice environment suggests that these patterns are likely a consequence of imperfect information processing rather than of the maximization of stable preferences.

Our first contribution is to show, using a meta-analysis of certainty equivalents imputed from all relevant past estimates, that this instability has always been present in the literature but has gone largely unnoticed: binary choice and choice lists produce systematically different patterns of choice. Second, we report an experiment designed to stress-test and interpret this observation by comparing choice list elicitations and binary choice problems that are *identical* to one another from the perspective of *any* standard theory of preferences (including PT). Again, over the typical range of probabilities studied in the literature, we find the same fundamentally different risk-taking behavior in choice list elicitations vs. direct choice, suggesting that factors other than standard preferences are likely responsible for the fourfold pattern in choice lists.

In contrast to Harbaugh, Krause and Vesterlund (2010), who conclude that binary choice is 'indistinguishable from random behavior', we find that the fourfold pattern devolves into a *more stable* pattern of behavior in binary choice — a conclusion that differs from Harbaugh, Krause and Vesterlund (2010) because of our use of a much richer set of choice stimuli.[1] While mere noise in binary choice can be rationalized by standard utility maximization models (e.g., via random utility and additive white noise specifications), our finding that risk-taking patterns change altogether in binary choice cannot, calling for an alternative theoretical explanation. As a step in this direction, we offer a cognitive explanation for risk-taking behavior that predicts a fourfold pattern in choice list elici-

[1] As we discuss below, our results are entirely consistent with Harbaugh, Krause and Vesterlund (2010), but we are able to detect a less noisy pattern in binary choice simply because we use a much richer set of binary choice stimuli that allows us a clearer window on patterns under binary choice.

tations but a different pattern in direct choice due to the way cognitive frictions interact with procedural details of these choice environments. We design and run a distinctive experimental test of this model and find strong evidence for this interpretation. We argue that these results have important implications for the way we interpret descriptive theories like PT and suggest that the modeling of cognitive frictions can be helpful in overcoming their predictive limitations.

Meta analysis. Meta-analytically re-evaluating data from decades of prior research, we show that the fourfold pattern does not typically occur in direct choice between lotteries, but nearly always arises in experiments that explicitly elicit certainty equivalents, e.g. when choices are *bundled* into choice lists for the purpose of elicitation.[2] Instead, direct binary choices between lotteries more often display a *twofold pattern* of risk attitudes over the same probability range: consistent risk aversion for gains, and consistent risk seeking for losses.[3] Despite being pervasive in prior data, this disjunction between binary choices and bundled choices has gone mostly unnoticed in the literature, with the important exception of Harbaugh, Krause and Vesterlund (2010), who (to our knowledge) were the first to report evidence that the fourfold pattern fails to appear in binary choice.

Motivated in part by Harbaugh, Krause and Vesterlund (2010)'s finding that the fourfold pattern obtains in valuation elicitations but not in binary choice tasks, we examine to what degree this same disjunction is reflected in the long prior literature estimating prospect theory parameters on experimental data. To investigate, we conduct a meta-analysis in which we search for all prior papers that (i) use data from standard choice list tasks or data from simple binary choices; and that also (ii) estimate PT parameters on those data. We then use the estimated PT parameters to predict certainty equivalents, allowing us to make inferences about the average patterns of risk-taking in the different

[2]The fourfold pattern has also been documented using valuation tasks, such as BDM willingness to pay elicitations, which are very similar to choice lists in many respects — see e.g. Chapman et al. (2023*a*) for the similarity between choice lists and valuations. We focus on choice lists both because they are the most common method in the literature for eliciting certainty equivalents, and because they are easier to compare to binary choice (i.e., because choice lists are nothing more than a bundled series of binary choices).

[3]It is important to emphasize that we do not claim that risk seeking for gains or risk aversion for losses can never occur in binary choice, no matter how small the probabilities or how large the outcomes. Such patterns have indeed been documented e.g. by Chark, Chew and Zhong (2020) and Ruggeri et al. (2020). Our point is rather that over typical probability ranges studied in the literature there is a strong dissimilarity in patterns of behavior across decision formats that cannot be accounted for in standard models: the fourfold pattern occurs robustly in one format (choice lists) but not in the other (binary choice).

papers. The results paint a surprisingly clear picture. Estimates from prior choice list tasks overwhelmingly produce evidence for the full fourfold pattern. By contrast estimates from prior binary choice tasks virtually never do. Instead, binary choice often produces a very different "twofold pattern" of global risk aversion for gains and risk-seeking for losses. This analysis suggests that we have in some sense "known" that the fourfold pattern doesn't really predict binary choice all along, but the literature has somehow largely been unaware of this fact.

We argue that this was missed largely because the literature has tended to emphasize estimates of the individual *theoretical* components of PT over the core *empirical* patterns that the theory was designed to explain. PT consists of two theoretical components that *jointly* determine risk attitudes: an inverse-S shaped probability weighting function and an S-shaped utility function (analogous to a reference-dependent version of a standard utility function). Estimates on both binary choice and choice list data have produced structural estimates consistent with inverse-S shaped probability weighting, but this is not sufficient to establish the fourfold pattern because of the confounding influence of utility curvature, which is typically estimated as much more severe in binary choice than in choice list tasks. The literature, however, has typically not put these estimates together to assess their joint implications for the (fourfold) pattern the theory predicts.

Experimental evidence. Experiments deploying binary choices versus choice lists to estimate PT functionals in the prior literature have typically not been designed with an eye to comparability between the two formats. Differences in the meta-analytic patterns we document could thus in principle arise from differences in the number and types of choices subjects face in the two settings. To test whether such factors are driving this result, we design a new experiment that presents identical choices between a lottery and sure amounts of money, presented either one-per-screen (our 'Choice' tasks) or bundled into choice lists (what we call 'Equivalence' tasks).[4] Because these tasks include *identical*

[4]Throughout the paper, we reserve the term "choice list" for the type of tasks used to elicit certainty equivalents – the *monetary values* obtained for sure people consider to be equally good as playing the lottery. These indifferences are generally elicited in the prospect theory literature, including by Tversky and Kahneman (1992). However, there is also evidence that choice lists keeping the *sure payments* constant and vary the probability in a list produce very different findings on, e.g., probability weighting (e.g. Feldman and Ferraro, 2023; Shubatt and Yang, 2024). These studies show that the standard inverse-S shape of the probability weighting function measured with certainty equivalents reverses into an S-shaped function when using probability equivalents instead.

component lottery choices, they are indistinguishable under the lens of standard models of utility maximization, providing a maximally stringent test of the hypothesis that binary choice and choice lists produce fundamentally different patterns of behavior.

We find substantial differences in behavior across these two settings that are very consistent with our meta-analytic findings. In our choice list tasks we find highly conventional evidence of the fourfold pattern. In our binary choice tasks, by contrast, the fourfold pattern collapses: systematic risk seeking disappears for small probability gains and systematic risk aversion disappears for small probability losses. The fourfold pattern in choice lists thus tends to be replaced by a *twofold pattern in binary choice*, just as in our meta-analysis. Because our design consists of identical choices, this finding suggests that the differences between the two tasks are *entirely* driven by the difference in the presentation — a difference that should not matter under any standard utility-based model. A tempting explanation is that (perhaps) this occurs because binary choice is more difficult than choice lists, preventing a true, latent fourfold pattern from expressing itself due to an increase in noisy behavior. However, the data is inconsistent with such an explanation — if anything they suggest the opposite pattern of reduced noise in binary choice.

Implications. We view these findings as having two primary implications. First, our findings inform a long-running debate in the PT literature about whether the theory should be interpreted as a description of (i) a suite of cognitive adaptations and shortcuts that generate its distinctive patterns; or (ii) what economists would usually think of as a standard sort of preference — a reflection of stable tastes for risk with welfare implications. We interpret the systematic violation of procedural invariance shown by our re-examination of the literature — and confirmed in our new experiments — as providing evidence in favor of the cognitive shortcuts interpretation, (i).

Second, we view these findings as informing the question of whether there is value in complementing tractable descriptive theories like PT with models that more explicitly model the cognitive processes underlying the anomalies they describe. We view the fact that a key diagnostic pattern at the heart of the theory (the fourfold pattern) robustly varies with choice context (and disappears in a setting as important as direct binary choice) as strong evidence that pursuing such models may be in fact quite valuable. One way of seeing such cognitive models is indeed as a way to endogenously account for the origin of the descriptive parameters estimated in PT settings (Vieider, 2025). Seen through this lens, explicit models of the cognitive processes generating observed choice behavior hold the promise of restoring some of the predictive value to descriptive models that they would otherwise risk losing.

Theoretical Mechanism and Experimental Validation. Given this, we next propose and experimentally validate an explanation that builds on the idea that observed choice patterns may be the result of cognitive frictions in information processing rather than stable tastes for risk. Our explanation builds on a series of recent findings suggesting that likelihood-dependence in risk-taking (the ultimate foundation of the fourfold pattern) may arise from cognitive frictions, and the mechanism the mind uses to overcome these limitations (Robson, 2001; Netzer, 2009; Khaw, Li and Woodford, 2021; 2023; Vieider, 2024; Enke and Graeber, 2023; Frydman and Jin, 2023; Glimcher and Tymula, 2023; Barretto-García et al., 2023; Netzer et al., 2024; Oprea, 2024*b*; Oprea and Vieider, 2024). Such approaches are often inspired by findings in neuroscience, and rest on the premise that explicitly modeling the process by which choice options are assessed and decisions are formed can improve the predictive powers of models of decision-making.

Our model is based on a simple intuition. In binary choice, it is natural to separately assess the odds of winning in a lottery and the relative rewards offered by the different options, and to then trade off these dimensions to reach a decision (Vieider, 2024). Choice lists, on the other hand, seem to suggest a different process, whereby the fixed element in the list — the lottery in the case of certainty equivalent lists — is evaluated first. From a cognitive point of view, the problem then consists in finding the value of this lottery, and matching it with an equivalent sure amount (see also Khaw, Li and Woodford, 2023, for a similar argument). Even though the cognitive processes at work are the same, the different ways in which they are applied depending on the choice context will result in different predictions. We experimentally test these predictions by presenting information sequentially in a binary choice setting, thus trying to trigger the sequential evaluation mode of choice lists in a binary choice setting. As predicted by our model, choice patterns in such a sequential evaluation model converge towards those observed in choice list settings, thus providing supporting evidence for the key mechanism underlying our model.

Relation to prior work. Our work is most closely related to Harbaugh, Krause and

Vesterlund (2010), who compare willingness-to-pay (elicited using the BDM mechanism) for lotteries to binary choices between the same lotteries and their expected value. They recover the fourfold pattern in willingness-to-pay but find choice behavior that is "indistinguishable from random choice" in the lottery choices (p. 595). The paper is seminal for introducing evidence of differences between valuation tasks and binary choice in expressions of the fourfold pattern, and in showing that the fourfold pattern fails to manifest in binary choice. Relative to this paper, our contribution is to (i) use a large-scale meta-analysis that re-examines the sum-total of PT estimations from choice lists and binary choice to show that binary choice in fact produces a consistent two-fold pattern of risk-taking; (ii) run experiments with much richer choice stimuli that allow us to directly compare the certainty equivalents underlying binary choice to directly elicited certainty equivalents; (iii) use identical choice situations presented one-by-one vs packaged into choice list, producing an especially stringent test of the relationship between the two; and (iv) present and test a theoretical account of what drives differences in observed behavior. Although we view our results as perfectly consistent with theirs, because of our richer stimuli our conclusions also differ in one important sense from theirs: while they conclude that binary choice produces patterns 'indistinguishable from random behavior', we are able to show using our much richer array of choice tasks that in fact binary choice produces tends to produce a two-fold pattern that is overall *less* random than choices observed in choice lists.[5] This distinction matters: whereas random choice in binary choice tasks can be organized under a 'utility maximization plus white noise' model, the highly coherent patterns we document in binary choice cannot, thus raising important questions about the interpretation of such models.

Our work is more broadly related to a long literature in psychology and economics showing that decision under risk often fails *procedural invariance*: measured risk preferences often depend in systematic ways on the method of elicitation used. These effects have been documented at least since Slovic (1964), and have periodically resurfaced in the literature in many different contexts (e.g., Hershey and Schoemaker, 1985; Crosetto and Filippin, 2015; Mata et al., 2018; Friedman et al., 2017; Zhou and Hey, 2018; Friedman et al., 2022), often yielding evidence broadly consistent with ours. In a recent paper

[5]This difference in conclusion is due to the fact that while Harbaugh, Krause and Vesterlund (2010) examine only a single binary choice for each lottery, we study a large number of binary choices. This allows us to sharply measure the certainty equivalent rationalizing binary choices and conduct a sharper head-to-head comparison with direct elicitations.

McGranaghan et al. (2024) show that binary choices and choice lists generate often very different evidence in favor of the common ratio effect, and argue that this is driven by the differential effects of noise in the two settings. Relative to much of this literature, the procedural invariance violations we document are particularly striking because they appear in choice lists and binary choices that consist of an *identical* set of constituent decisions.

Closely related is a literature on "preference reversals" which shows that apparent preferences over lotteries often flip when elicited via binary choice versus valuation (buying or selling prices of a lottery; Lichtenstein and Slovic, 1971; Grether and Plott, 1979). Modern versions of prospect theory allow researchers to explain such preference reversals via loss aversion, due to the differing implicit endowments in choice versus valuation tasks (Schmidt, Starmer and Sugden, 2008). The violations of procedural invariance we document involve identical choices without variation in endowments, meaning they cannot be explained in this way. Indeed, we show empirically that even allowing for endogenous reference points cannot account for our results via prospect theory.[6]

Methodologically, the closest work to ours is a recent series of papers that compares choice lists and choice behavior by (i) having subjects make decisions in multiple price lists and (ii) having subjects also make decisions in the "stacked" choices of the lists individually and in a random order. Lévy-Garboua et al. (2012) compared choices in the task of Holt and Laury (2002) to the underlying binary decisions, and documented higher levels of risk aversion and increased 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 (however, see Freeman and Mayraz, 2019, for an account that suggests an alternative explanation). Revisiting this issue, Brown and Healy (2018) conclude that incentive-compatibility is guaranteed by randomly paying one of several choices presented in a binary choice mechanism in random order. These papers examine only one multiple price list and one collection of corresponding binary choice problems. One of our main contributions relative to these papers is to study *multiple* choice lists, and multiple corresponding collections of binary

[6] Shubatt and Yang (2024) offer a cognitive explanation for preference reversals, rooted in the noise generated by the complexity of comparing lotteries — an explanation that is related to our model and is broadly consistent with many of our findings.

Choice tasks. We believe this allows us to paint a much richer picture of the factors driving decisions, and to use the resulting patterns to test implications for PT.

Finally, our paper relates to a growing literature in economics that attempts to re-interpret anomalies under the lens of domain-general cognitive frictions rather than domain-specific preferences or errors (Enke, 2024; Enke et al., 2024; Oprea, 2024*a*). Most relevant for our purposes is a growing literature on *noisy cognition*, which explains anomalies as growing out of the imprecise ways constrained brains represent information and the biases produced by Bayesian responses to this noise. These models can explain many aspects of prospect theory, including probability weighting (Vieider, 2024; Khaw, Li and Woodford, 2023; Frydman and Jin, 2023; Glimcher and Tymula, 2023; Netzer et al., 2024) and tests of these models have been highly empirically successful (Natenzon, 2019; Prat-Carrabin and Woodford, 2022; Enke and Graeber, 2023; Barretto-García et al., 2023; Oprea and Vieider, 2024). We use a model in this class to explain our main findings, and test that model using a novel experiment, adding to this accumulation of evidence.

2 Reassessing the Literature

We take advantage of the fact that PT parameters have been estimated repeatedly in a long prior literature from both choice lists (by far the most commonly used method for eliciting certainty equivalents) and binary choices to contrast regularities across the two formats. We limit ourselves to choice lists and binary choice because — at least in principle — they are both made up of the same component parts, i.e. individual choices between lotteries with different variances and expected values. This will allow us to assess the generality of differences between choice lists and binary choice, as well as allow us to better interpret the plausible drivers of such differences.

In Section 2.1 we collect all relevant prospect theory estimates from the prior literature and use these estimates to assess (using imputed certainty equivalents implied by these estimates) whether the fourfold pattern occurred in choice lists but not binary choice. This systematic re-examination of the literature confirms one key finding by Harbaugh, Krause and Vesterlund (2010): that the fourfold pattern does not occur in binary choice, but does occur in choice lists. The rich binary choice stimuli typical of this literature also enables us to further interpret this finding: at least over the typical probability ranges examined in the literature, binary choice tends to produce a *twofold pattern of risk-taking* — risk aversion for gains, and risk seeking for losses. In Section 4 we discuss why the prospect theory literature has mostly not noticed that the theory's "most distinctive implication" typically does not appear in direct lottery choice.

2.1 Meta-analysis

In order to meta-analyze evidence on the fourfold pattern from the prior literature, we collect estimates of the parameters of PT functionals from dozens of prior papers. Most of the prior literature relevant to the fourfold pattern summarize their data by reporting estimates of PT parameters including those from (i) a probability weighting function and (ii) a utility function. Our approach is therefore to calculate certainty equivalents based on these past estimates and examine how often these calculations show evidence of the pattern in both choice list tasks and binary choice tasks.

Inclusion criteria. We began by collecting all PT estimates from the prior literature that can be used to contrast the fourfold pattern in choice lists and binary choices. Our inclusion criteria were that 1) estimates should stem either from a pure choice list setup or a pure binary choice setup (thus excluding hybrid elicitation mechanisms such as choice lists filled in based on a bisection procedure)[7]; and 2) for the paper to present an estimation of PT parameters. The latter criterion ensures that we can use the reported PT parameters to infer a predicted CE for a given probability and outcome, and thus to have comparable quantities. We conducted a literature search in the spring of 2023. The search procedures followed closely those of the meta-analysis on loss aversion of Brown et al. (2024). We then read through the abstracts and excluded papers that clearly did not meet our criteria. We subsequently read all papers that had passed this initial stage, and encoded the PT parameters.

Analysis Approach. For each paper, we use PT estimates to calculate a predicted certainty equivalent (CE), $\widehat{c} = u^{-1}[w(p)u(x)]$, where u designates the utility function, w the probability weighting function, and u^{-1} the inverse of the utility function. In what follows, we use $x = 100$ and probabilities $p = 0.1$ and $p = 0.9$ to calculate normalized

[7]We decided to include papers that used a list format, but used 2 or more stages to zoom in on the precise CE. This allowed us to include e.g. the seminal papers of Tversky and Kahneman (1992) and Gonzalez and Wu (1999) in the analysis.

certainty equivalents $\widehat{c}/x$ at low and high probabilities, but the qualitative results do not change much for different monetary outcomes or even more extreme probabilities (see Online Appendix B).[8]

We obtain standard errors for the predicted CEs by applying a bootstrap procedure. We assume the parameters to be normally distributed around their mean estimate with variance equal to the squared standard error of the parameter (as customary in meta-analysis). By drawing repeatedly from the uncertainty intervals surrounding the parameters and using the draws to calculate a vector of CEs, we can obtain an approximation of the standard errors surrounding our mean estimate of the predicted CE. Online Appendix B contains the details of the procedure and presents an overview of the imputed CEs for different studies.

Obtaining standard errors allows us to apply Bayesian meta-analytic procedures to the data, weighing each observation by the inverse of its variance (see e.g. Brown et al., 2024). This procedure is of course only as good as the data we feed into it. The bootstrapped standard errors especially could be affected by differences in estimation methods, and functional and modeling assumptions across studies. Our calculation procedure also implicitly assumes that errors across parameters are independent, and may over-estimate the standard errors of the calculated CEs if that assumption does not hold. To counteract these issues, we will present a number of robustness checks. In particular, we will present simple averages of the point estimates, and approximate standard errors by making them proportional to the inverse of the square root of the sample size.

Results. As expected, we find robust evidence of the fourfold pattern in prior choice list studies. Figures 1 and 2, Panel A, show estimates for Gains for $p = 0.1$ and $p = 0.9$ respectively. At $p = 0.1$ all but a handful of estimates fall far to the right of 0.1, indicating risk seeking, with an average meta-analytic normalized certainty equivalent of 0.181, with a 95% credible interval of $[0.162, 0.201]$. At $p = 0.9$ we find the opposite, with all estimates falling to the left of 0.9 with an average meta-analytic normalized certainty equivalent of 0.721, with a 95% credible interval of $[0.693, 0.748]$. Panel A in

[8] The reason for using probabilities of 0.1 and 0.9 is that these are often the most extreme probabilities included in the studies. Extrapolations to more extreme probabilities should thus be consumed with some caution, though our results are still robust to smaller probability 0.05 (see Table B.5). Likewise \$100 was an amount often used in the early literature. Since we use CRRA coefficients in our computations, which are indeed estimated in the great majority of papers, the monetary amounts used do not have much influence on our conclusions.

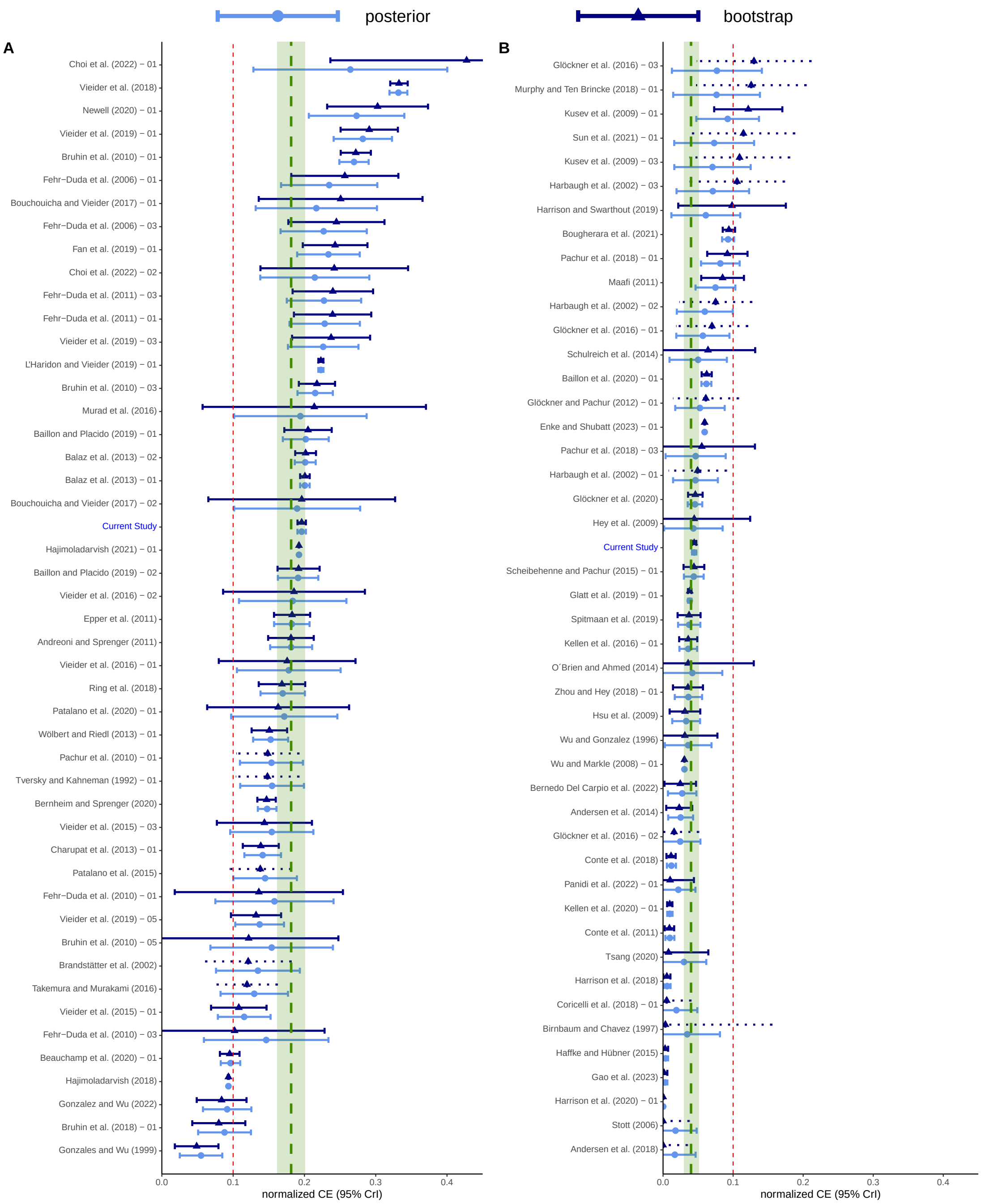


Figure 1: Forest plot of inferred CEs for a wager $(100, 0.1)$, gains
Forest plot of calculated CEs, normalized as $\frac{\widehat{c}}{100}$ to be directly comparable to the probability of winning, and 95% credible intervals. Panel A summarizes CEs inferred from choice list studies using elicited certainty equivalents to measure risk attitudes, whereas panel B shows CEs inferred from studies using binary choices to obtain PT parameters. The navy blue triangles indicate calculated data points, while the light blue squares indicate the meta-analytic posterior. The vertical dashed lines with the shaded green region surrounding it represents the meta-analytic average with its 95% credible interval. Dotted confidence intervals stem from studies which did not report statistical information for the reported PT parameters, and for which we thus had to impute the SEs as missing data (see online appendix for details).

Figures 3 and 4 shows that each of these risk postures "flip" in choice lists studies under Losses. At $p = 0.1$ (pictured in Figure 3), the meta-analytic mean for the 25 studies

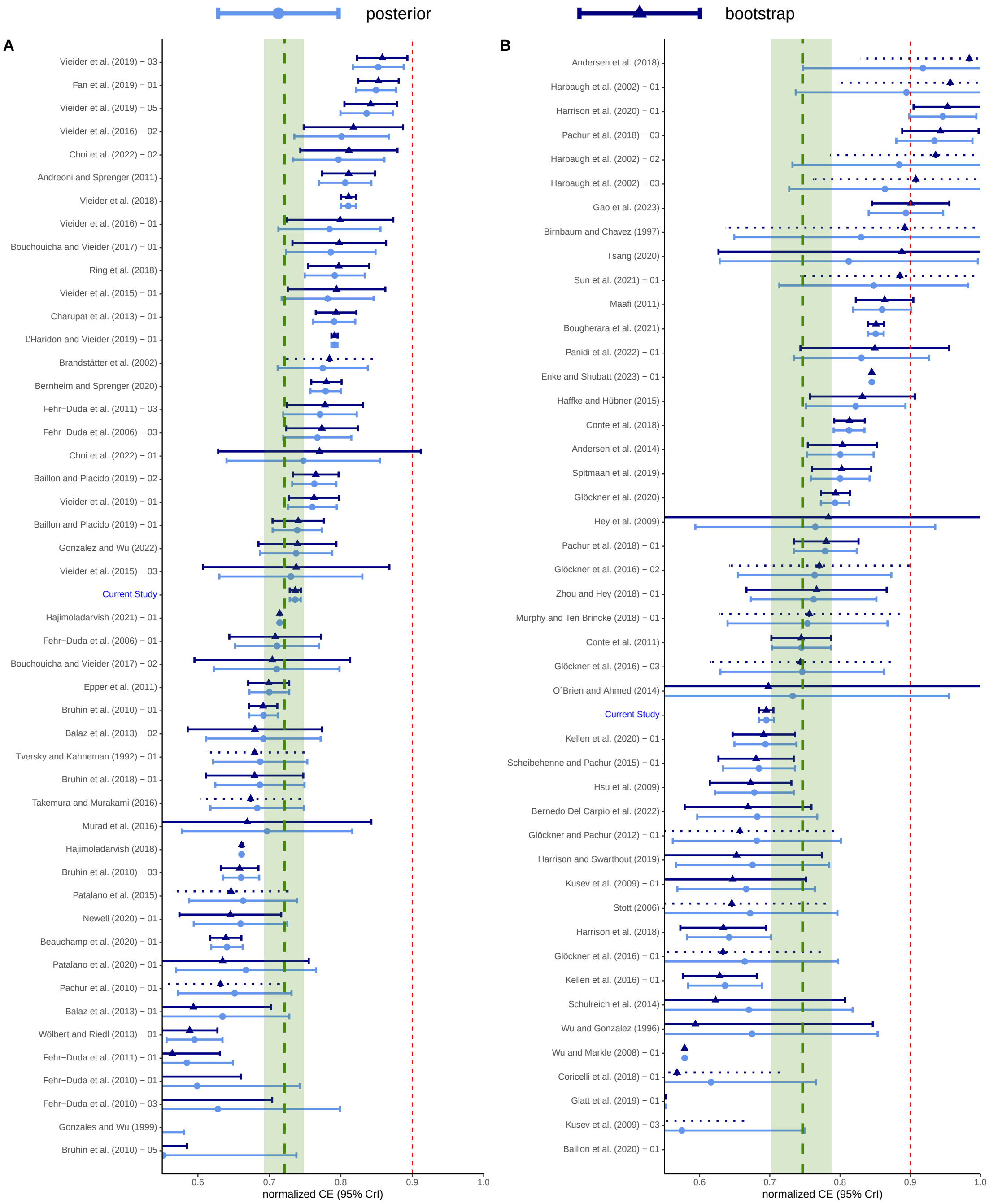


Figure 2: Forest plot of inferred CEs for a wager $(100, 0.9)$, gains

Forest plot of calculated CEs, normalized as $\frac{\widehat{c}}{100}$ to be directly comparable to the probability of winning, and 95% credible intervals. Panel A summarizes CEs inferred from choice list studies using elicited certainty equivalents to measure risk attitudes, whereas panel B shows CEs inferred from studies using binary choices to obtain PT parameters. The navy blue triangles indicate calculated data points, while the light blue squares indicate the meta-analytic posterior. The vertical dashed lines with the shaded green region surrounding it represents the meta-analytic average with its 95% credible interval. Dotted confidence intervals stem from studies which did not report statistical information for the reported PT parameters, and for which we thus had to impute the SEs as missing data (see online appendix for details).

again falls far to the right of 0.1, which in losses now indicate *risk aversion* (mean: 0.213; CrI: [0.171 , 0.256]). At $p = 0.9$ (pictured in Figures 4), the meta-analytic mean for the

25 studies flips again, now falling far to the left of 0.9, indicating *risk seeking* in losses (mean: 0.764; CrI: $[0.737, 0.791]$). Thus, in prior choice list studies, with few exceptions, we calculate certainty equivalents indicating risk seeking for low- and risk aversion for high-probability gains, and the reverse in each case for losses. In elicitations using choice lists, the fourfold pattern appears to be extremely robust.

Our main finding is that this pattern collapses in binary choice studies. Panel B in each Figure plots data from these studies. In Figure 1 we find that, in violation of the fourfold pattern, the meta-analytic mean is 0.040 with a credible interval of $[0.030, 0.051]$, indicating a normalized certainty equivalent less than 0.1 and therefore *risk aversion*. Indeed, almost all studies yield *risk averse* certainty equivalents, with no statistically significant results indicating the fourfold pattern's predictions of risk seeking. This risk aversion remains at 0.9 in Figure 2 (mean: 0.746; CrI: $[0.703, 0.787]$), suggesting that unlike in choice lists subjects display *uniform risk aversion* in the gains domain in prior experiments using binary choices. In Losses, at $p = 0.1$ (Figure 3) the meta-analytic mean is 0.075 with a credible interval of $[0.054, 0.097]$, indicating modest risk seeking. Though the data are more mixed here, only one study shows statistically significant evidence of the risk-aversion predicted by the fourfold pattern, whereas 8 are significant in the opposite direction. At $p = 0.9$ (Figure 4) we continue to find evidence of risk seeking in losses in the meta-analytic mean (mean: 0.784; CrI: $[0.713, 0.842]$), suggesting that subjects are reliably risk seeking in losses regardless of probability.

Thus, overall, the fourfold pattern tends to collapse towards a twofold pattern in binary choice: risk aversion in gains and risk seeking in losses.

Robustness. Meta-analysis can suffer from data problems, meaning robustness checks are especially important to verify our statistical findings. Information on standard errors is not always reported or available in past studies. Even when such information is available, the statistical evidence provided may be inaccurate. This holds all the more in our case, since we need to simulate the standard errors of the calculated CEs from a combination of PT parameters. To test the robustness of the results, we can instead calculate the simple means of the CEs calculated from the parameter point-estimates. The mean normalized CE calculated from choice list studies for gains at $p = 0.1$ is 0.186 (median 0.182) and at $p = 0.9$ is 0.707 (median 0.726). The mean normalized CE from choice list studies for losses is 0.231 (median 0.230) at $p = 0.1$ and 0.774 (median 0.765)

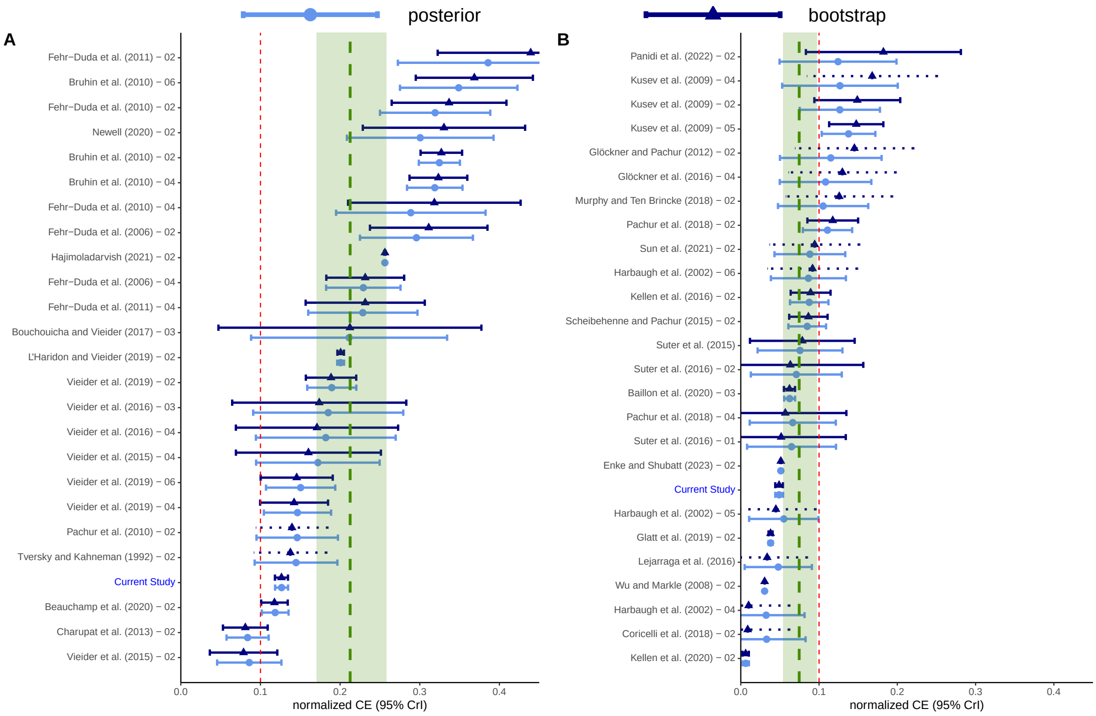


Figure 3: Forest plot of inferred CEs for a wager $(100, 0.1)$, losses

Forest plot of calculated CEs, normalized as $\frac{\widehat{c}}{100}$ to be directly comparable to the probability of winning, and 95% credible intervals. Panel A summarizes CEs inferred from choice list studies using elicited certainty equivalents to measure risk attitudes, whereas panel B shows CEs inferred from studies using binary choices to obtain PT parameters. The navy blue triangles indicate calculated data points, while the light blue squares indicate the meta-analytic posterior. The vertical dashed lines with the shaded green region surrounding it represents the meta-analytic average with its 95% credible interval. Dotted confidence intervals stem from studies which did not report statistical information for the reported PT parameters, and for which we thus had to impute the SEs as missing data (see online appendix for details).

at $p = 0.9$. The calculated CE for gains from binary choice studies is 0.045 (median 0.036) at $p = 0.1$ and 0.752 (median 0.773) at $p = 0.9$. The mean relative CE for losses is 0.079 (median 0.069) at $p = 0.1$ and 0.765 (median 0.800) at $p = 0.9$. Thus our main findings continue to strongly hold. While there is overwhelming evidence for the fourfold pattern in PT estimates obtained from direct choice list elicitation, estimates based on binary choice indicate something closer to a twofold pattern – risk aversion for gains, and risk seeking for losses. In Online Appendix B we report two additional robustness checks. Following the recommendation of Furukawa et al. (2006), we use the average standard deviation in studies in which we can obtain it, and then divide this standard deviation by the square root of the sample size to derive standard errors. All of our results are robust to this additional robustness check. Furthermore, we conduct a series of meta-regressions, controlling for number of outcomes in the lottery, whether incentives are real or hypothetical, whether one of the two options consisted of a sure payment (in binary choice), and a variety of study and estimation characteristics. Once again, our

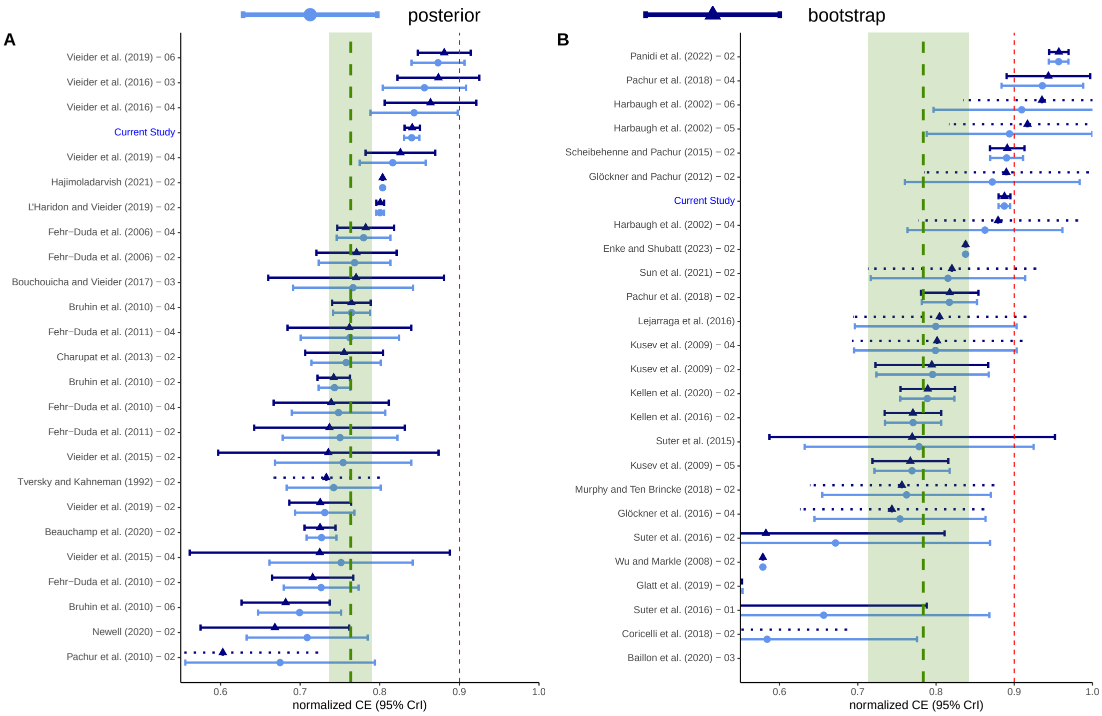


Figure 4: Forest plot of inferred CEs for a wager $(100, 0.9)$, losses

Forest plot of calculated CEs, normalized as $\frac{\widehat{c}}{100}$ to be directly comparable to the probability of winning, and 95% credible intervals. Panel A summarizes CEs inferred from choice list studies using elicited certainty equivalents to measure risk attitudes, whereas panel B shows CEs inferred from studies using binary choices to obtain PT parameters. The navy blue triangles indicate calculated data points, while the light blue squares indicate the meta-analytic posterior. The vertical dashed lines with the shaded green region surrounding it represents the meta-analytic average with its 95% credible interval. Dotted confidence intervals stem from studies which did not report statistical information for the reported PT parameters, and for which we thus had to impute the SEs as missing data (see online appendix for details).

key results are unaffected when controlling for such potential differences between studies, suggesting that they are not driven by systematic differences in procedures or estimation techniques in choice lists versus binary choice.

Likelihood insensitivity. Finally, we complement our primary analysis of the fourfold pattern with a second, subtler prediction about risky choice made by PT: likelihood insensitivity. While the fourfold pattern describes PT's main prediction about how risk postures change with unmixed lottery characteristics, likelihood insensitivity describes how the intensity of preferences for or against risk change as probabilities do. PT predicts that subjects' apparent risk preferences respond to increases in probabilities sluggishly, with the intensity of preferences changing less quickly than probabilities themselves do. If a utility function is estimated assuming expected utility theory, the degree of curvature of that function will thus be dependent on the fixed probability used to estimate it (Hershey, Kunreuther and Schoemaker, 1982). Likelihood insensitivity is a necessary (but not sufficient) condition for the fourfold pattern to occur: if subjects are risk seeking for small probability gains (risk averse for small probability losses), their insensitivity to probabilities ensures that this preference 'flips' as probabilities get larger.

To gather a transparent, non-parametric measure of likelihood sensitivity, we calculate the change in the rate at which subjects select the risky lottery at probabilities "mirrored" around 0.5 at $p = 0.9$ and at $p = 0.1$, at $p = 0.8$ and at $p = 0.2$, and at $p = 0.7$ and at $p = 0.3$. We then normalize these individual differences by the true difference in probabilities to put them on a common scale, and average them to get a subject-wise index of likelihood sensitivity.[9] A subject who weights probabilities linearly (as in, e.g., expected utility theory) will have an index of 1. Subjects with indexes below 1 are likelihood-insensitive, so that their relative risk aversion increases in probabilities for gains (decreases in probabilities for losses), and above 1 likelihood-oversensitive (relative risk aversion decreasing in probabilities for gains, and increasing in probabilities for losses).

Figure 5 plots empirical CDFs of the likelihood-sensitivity index for Gains (left panel) and Losses (right panel) from previous studies. In both cases, perfect likelihood sensitivity would predict a difference of 1, but estimates of sensitivities based on choice lists are *universally* lower than this. By comparison, in binary choice estimated sensitivities are substantially *higher*, with a significant number of studies showing the reversed pattern of *over-sensitivity* to probabilities. The difference in likelihood-sensitivity across choice lists and binary choice is highly statistically significant based on Wilcoxon tests ($p < 0.001$). Thus, the prior literature suggests that not only the fourfold pattern but also the subtler phenomenon of likelihood insensitivity is much weaker in binary choice than in choice lists. Once again, then, we find that a core descriptive element of prospect theory is robust in choice lists but weakens in binary choice. As with the fourfold pattern, the likelihood insensitivity predicted by prospect theory is a substantially better description of how subjects fill in choice lists than how they directly choose between them.

Discussion. Our meta-analysis paints a clear picture. Choice list designs have consistently produced a fourfold pattern of risk attitudes, just as predicted by prospect theory.

[9]Let $\pi(p)$ be the normalized certainty equivalent,i.e. c/x. The index is then calculated as $\frac{\pi(p_h)-\pi(p_l)}{p_h-p_l}$, where p_h and p_l stand for the high and low probability, and are always chosen to be symmetric around 0.5.

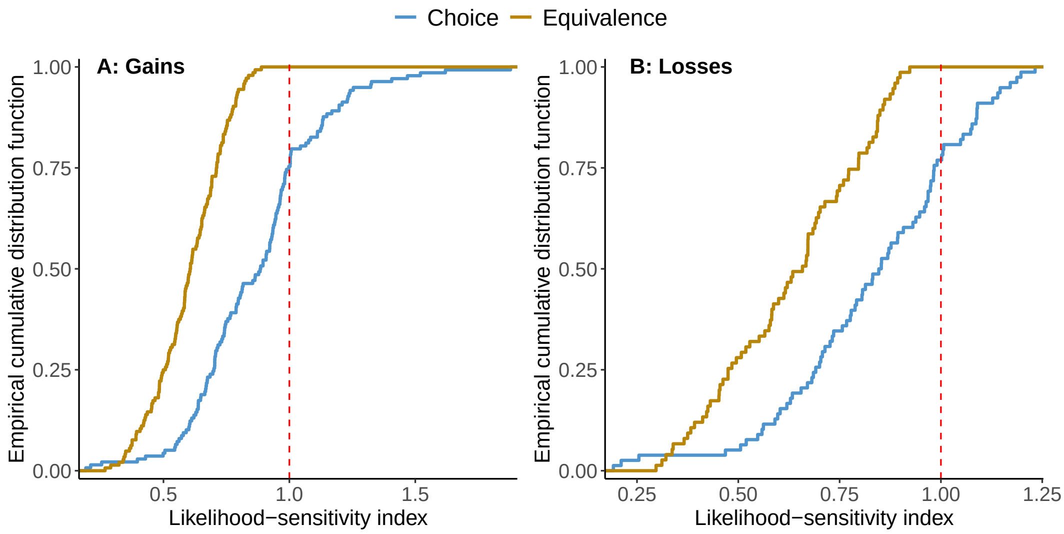
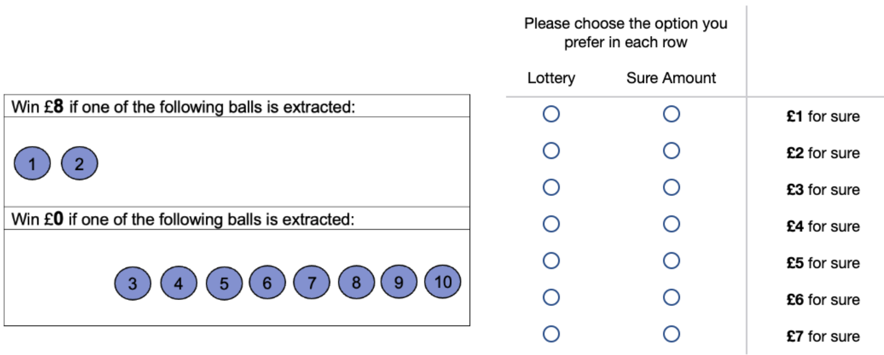


Figure 5: Empirical cumulative distribution function of likelihood-sensitivity
Empirical cumulative distribution function of the likelihood-sensitivity index, calculated as the average of the differences in normalized CEs for the pairs $(X, 0.9)$ and $(X, 0.1)$, $(X, 0.8)$ and $(X, 0.2)$, and $(X, 0.7)$ and $(X, 0.3)$. Normalization occurs by the division of the true probability difference, which means that subjects who are perfectly sensitive to probabilities (the EUT case) will have an index of 1. **Panel A** presents scatter plots of observations for the gain wager $X = 100$, derived from studies employing either certainty equivalents or binary choice methods. **Panel B** displays observations for the loss wager $X = -100$. Vertical dashed lines indicate the point of perfect likelihood-sensitivity.

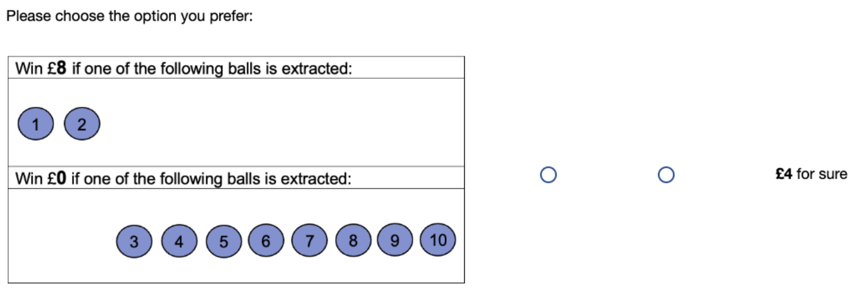
Binary choice tasks, on the other hand, tend not to produce the fourfold pattern and instead tend to produce something closer to a *twofold pattern* of risk-taking over the same probability range. This raises fundamental questions about the interpretation and predictive ability of PT when moving between different choice environments. One potential shortcoming of the evidence presented so far, however, is that it is not transparently causal: the evidence we have presented does not rely on a homogenous pool of subjects being prospectively randomized into different conditions. Optimization of experimental designs for the recovery of PT parameters may furthermore mean that the choices faced by subjects in choice list tasks and in binary choice tasks may not be truly identical. To test the robustness of our meta-analytic results to both of these potential shortcomings, we thus next report a new, randomized experiment.

3 Experiment

To measure the causal effect of presenting choices one-by-one in a binary choice setting versus collecting them into choice lists, we run an experiment. In particular, we study subjects' decisions between (i) a collection of two-outcome lotteries that pay some high amount x with probability p, and some lower amount $y < x$ otherwise, and (ii) a series of sure amounts of money that range from the low lottery outcome y to the high outcome x in steps of £1. The only difference between our binary choice and choice list treatments is whether these choices are presented one-by-one in a binary choice paradigm, or whether these same choices are instead collected into choice lists.



(a) *Equivalence* Condition



(b) *Choice* Condition

Figure 6: Screenshots of the two treatments

Screenshots from the Gain treatments. Panel (a) shows a screenshot of a typical Equivalence list for a lottery 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 Choice treatment.

Equivalence tasks: The most common method for directly eliciting certainty equivalents for lotteries is via choice lists. Choice lists are (effectively) a series of binary choices "stacked" on top of one another; Figure 6 shows an example (a screenshot from our experiment). On the left is a description of the lottery, and on the right is an ascending sequence of certain monetary payments. The subject makes a choice in each "row". By examining the sure payment amount at which the subject switches from preferring the lottery to preferring the sure payment, the researcher obtains an interval estimate of the subject's certainty equivalent for the lottery.[10] Each subject in our experiment is

[10]In ur design, subjects were allowed to choose independently in each row, which can give rise to

assigned 21 choice lists (parameters are described in Online Appendix A.1, allowing us to assess the fourfold pattern of risk attitudes and some related diagnostic questions (see below).

Choice tasks: In our binary choice tasks (shown in the bottom panel of Figure 6), subjects observe a lottery and a single sure amount, and simply choose the one they prefer. Our binary choice tasks always consist of a choice between a lottery and a sure payment, as in the example at the bottom of Figure 6. Each subject is assigned 274 binary choice tasks.

Comparability: The key to our design is that we selected our Choice tasks to include *all* of the individual choices embedded in each of the 21 MPLs of our Equivalence tasks. That is, we simply took the individual rows of each choice list, transformed each into a stand-alone binary choice task between the lottery and one of the sure amounts in each case, and assigned all of them to subjects. Subjects were assigned these Choice tasks in an order that randomly mixed both the lotteries and the certain payments. However, from a payoff perspective, subjects make identical decisions under an identical payoff scheme in our choice list and binary choice treatments. Under virtually any theory of utility maximization, including prospect theory, they should therefore yield identical patterns of behavior.

Choice stimuli: The primary goal of the experiment is to contrast evidence of the fourfold pattern of risk-taking under binary choice and choice lists. To measure the pattern, we vary the probability p, the outcome payoffs x and y, and the gain/loss framing in an orthogonal fashion. We do this using a design with both between- and within-subjects variation. Within-subject, for every subject we vary p between 0.1 and 0.9, and we introduce variation in both x and y to allow us to estimate PT functionals. Between-subject we run a Gain treatment in which $x, y \geq 0$ and a Loss treatment in which $x, y \leq 0$. Parameters are provided for both treatments in Online Appendix Tables A.1 and A.2. Our Gain treatments are incentivized while we follow much of the literature (Wakker and Deneffe, 1996; Abdellaoui, 2000) by using hypothetical incentives in our Loss treatment — this allows us to avoid common concerns in the literature that integrating payoffs with the initial endowments required to incentivize losses might create distortions in measure-

'multiple switching'. We thus use a stochastic approximation to the certainty equivalent given by the choice proportions in any given list—see also below. Notice, however, that the results we report are robust to a vareity of different ways of analyzing the data.

ment that would confound our inferences. Results from our meta-analysis suggest that expressions of the fourfold pattern are not different in hypothetical versus incentivized experiments in the literature. In a meta-analysis examining all existing tests randomly allocating hypothetical choices versus real incentives, Li and Vieider (2025) document an average (median, modal) effect size close to zero. A partial exception is trials involving losses, where the evidence strongly suggests that providing endowments triggers house money effects.

Understanding the Design: Relative to some related exercises from the literature, our design has three characteristics that (to our knowledge) have not been combined and that we believe allow for an especially crisp answer to our motivating questions:

- First, we designed the binary choice and choice list tasks to measure behavior over identical choice primitives. Instead of giving subjects one binary choice task for each lottery, we gave them a rich set, sufficient to infer preferences at the same level of detail as in choice lists. This produces a particularly strong test of our null hypothesis, and allows us to expand upon the findings of Harbaugh, Krause and Vesterlund (2010), who presented only one single choice task per lottery (between the lottery itself and its expected value).
- Second, our use of choice lists (rather than, e.g., BDM value elicitations) allows us to make our binary choice and choice list tasks truly identical under standard theories. Subjects are literally making the exact same set of choices in the two environments, making our test again especially strong.
- Third, unlike recent papers that follow our basic design strategy (i.e., to contrast choice lists with individual binary choices that reproduce the individual rows of those lists; Lévy-Garboua et al., 2012; Freeman, Halevy and Kneeland, 2019; Freeman and Mayraz, 2019), we study a large number of distinct choice lists within-subject in our choice list tasks and a number of decomposed choice lists in our binary choice tasks, while implementing an *identical* payoff mechanism. This means we can, for the first time with such a design, really assess and contrast the performance of prospect theory predictions at the subject level. By following Brown and Healy (2018) and randomly selecting one choice to be paid (from a series of binary choices presented in random order), we make our choice elicitation

incentive-compatible.

These design elements give us an unusually interpretable answer to our main question — and one that arguably works against our hypothesis by maximizing the chances of finding similar behavior in choice lists and binary choices.

Implementation Details: 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.[11] The main experiment is a between-subjects 2×2 design crossing $\{Equivalence, Choice\} \times \{Gains, Losses\}$. In our Gains treatments 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 (Equivalence: N=156; Choice: N=145). The median subject took 40 minutes to finish the experiment (33 minutes in Equivalence; 50 minutes in 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. In our Losses treatments, excluding 10 subjects who did not pass some very basic comprehension tests, we were left with 201 subjects providing valid responses (Equivalence: N=98; Choice: N=103). A typical subject took 42 minutes to complete the experiment (31 minutes for the choice list treatment, 50 minutes for the binary choice treatment).

3.1 Results

Figure 7 plots the main results, focusing on lotteries yielding a prize of £24 (a loss of £24) or else 0.[12] On the y-axis we plot choice proportions for the lottery (choice proportions for the sure amount for losses). Given that in our design sure amounts vary

[11] The full video instructions are available at `https://www.youtube.com/@RislabUgent`. The slides are included in Online Appendix D.

[12] Like most of the literature, we focus on lotteries with lower payments of $y = 0$ in our reduced form analysis because of tendencies towards decreasing absolute risk aversion in lottery choice (Bouchouicha and Vieider, 2017*a*) — tendencies which tend to modestly compress these patterns towards risk neutrality. Our structural analysis in Section 4, which makes use also of lotteries with $y > 0$ (included in our design to allow for proper identification of utility curvature in these estimates), reveals patterns that are very similar to this reduced form analysis. In Online Appendix A.1 we report results for each of the lotteries individually and find that (i) there are indeed statistically significant differences between Choice and Equivalence for most individual lotteries, and (ii) most are statistically different from risk neutrality. As expected, the exceptions to (ii) are concentrated in lotteries with non-zero lower payments where there is a stronger tendency towards risk neutrality in Choice.

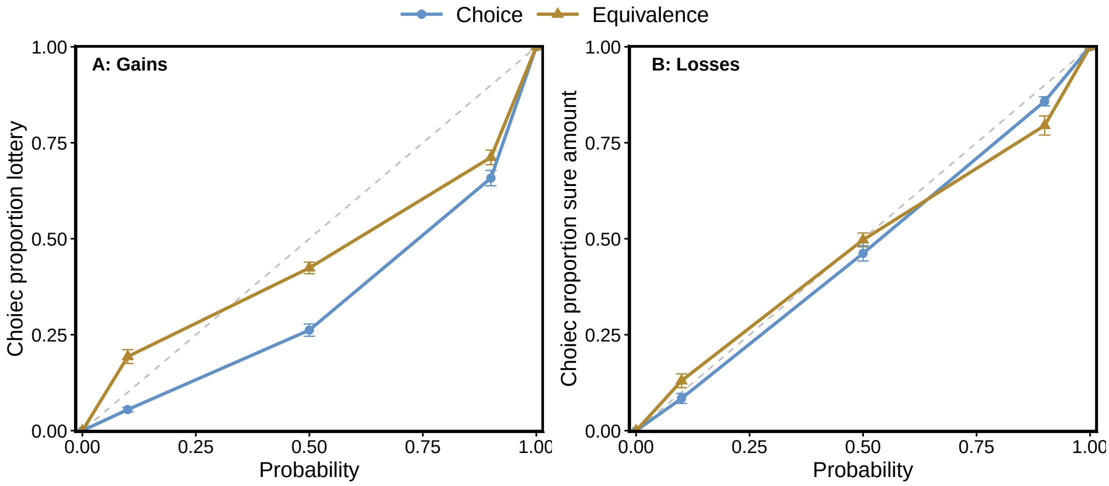


Figure 7: Nonparametric risk-taking measures evaluating the fourfold pattern by treatment.
The figure plots choice proportions of the lottery (gains, Panel A) or of the sure amount (losses, Panel B). Tables A.1 and A.2 in the Online Appendix provide the corresponding statistics for the entire set of individual lotteries included in the experiment.

in constant steps of £1 between the low outcome y and high outcome x of the lottery (the smaller and larger loss for Losses), such choice proportions approximate certainty equivalents when choice are well-behaved (see below for alternative ways of analyzing the data).[13] This allows us to directly compare the choice proportions to the risk neutral benchmark (in this case simply p). In Section 4 below (e.g., Figure 10) we complement this reduced-form analysis with certainty equivalents derived from structural estimates of prospect theoretic functionals on our data, mirroring the approach taken in our meta-analysis (indeed, we included these estimates in the meta-analysis itself) and find similar results.

The left-hand panel (A) plots results for gains, while the right-hand panel (B) plots results for losses. For the choice list treatment, the results nearly perfectly reflect the fourfold pattern: we find risk-seeking at low probability gains and at high probability losses, and risk aversion at high probability gains and (more weakly) low probability losses. In the binary choice tasks, however, the choice proportions look rather different. Here, the fourfold pattern disappears, and tends to be replaced by a twofold pattern. In particular, subjects in the gain domain on average show risk-aversion at both low and high probabilities, whereas subjects in the loss domain are always (mildly) risk-

[13]By 'well-behaved' we mean that subjects either switch only one time per list, or that — in the presence of multiple switching — switches are concentrated around the point of indifference. None of these assumptions are necessary, as we show in several alternative ways of analyzing the data reported below, but they make for a transparent way of representing choice patterns in first approximation.

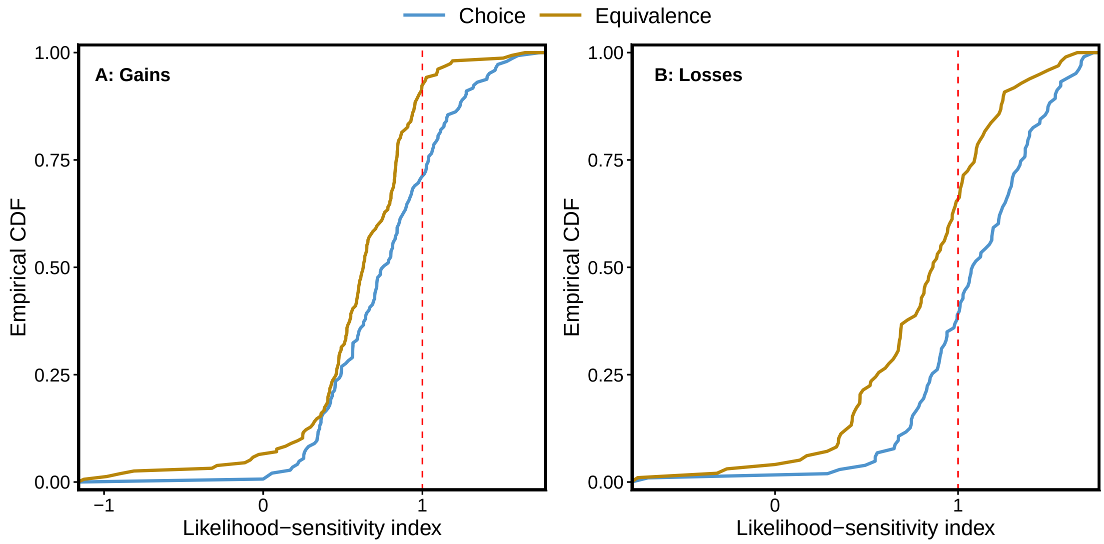


Figure 8: Measured likelihood sensitivity in Equivalence and Choice

Empirical cumulative distribution function of the likelihood-sensitivity index, calculated as the average of the differences in normalized CEs for the pairs $(\pm 24, 0.9; \pm 0)$ and $(\pm 24, 0.1; \pm 0)$, $(\pm 16, 0.8; 0)$ and $(\pm 16, 0.2; 0)$, $(\pm 16, 0.7; 0)$ and $(\pm 16, 0.3; 0)$, and $(\pm 8, 0.8; 0)$ and $(\pm 8, 0.2; 0)$. **Panel A** presents scatter plots of observations for gains. **Panel B** displays observations for the losses. Vertical solid lines indicate the point of perfect likelihood-sensitivity.

seeking.

Thus, a raw analysis of our data suggests that the fourfold pattern is a phenomenon of choice lists but fails to appear in binary choice. In Section 4 below we revisit these data structurally under the lens of PT and show (i) that there is nothing unusual about our data in either case, in the sense that under both binary choice and choice lists we estimate standard PT parameters from these data (e.g., we get standard inverse-S shaped probability weighting) and (ii) that these PT parameter estimates *themselves* imply exactly what our non-parametric results show directly: that the fourfold pattern does not occur in binary choice as it does in choice lists.[14]

Likelihood insensitivity. Next, we show that, as in our meta-analysis, in our experiment likelihood insensitivity is sharply attenuated in binary choice relative to choice list data. As in our meta-analysis, we calculate our likelihood sensitivity index as the change in the rate at which subjects select the risky lottery at probabilities "mirrored" around 0.5, e.g., 0.9 and 0.1, 0.8 and 0.2 etc. and normalize them by true probability

[14] We also find quite typical variations in choice patterns as stakes increase. In particular, we find evidence for increasing relative risk aversion (IRRA) in Gains in both choice lists and binary choice tasks, although the level of risk aversion is more pronounced in binary choice. In Loss tasks, we replicate the typical finding of constant relative risk aversion in choice lists (Fehr-Duda et al., 2011; Bouchouicha and Vieider, 2017*b*), but we again find IRRA in the size of the loss in binary choice. Online Appendix A.4 reports these findings in more detail.

differences for every subject. As before, subjects who weight probabilities linearly will have an index of 1, while subjects with an index below 1 are likelihood-insensitive and above 1 are likelihood-oversensitive.

Figure 8 plots empirical CDFs of this index from the choice list and binary choice conditions (mirroring Figure 5), including separate plots for gains (left panel) and losses (right panel). In Gains, we find that subjects in both treatments tend to be likelihood insensitive, but that subjects' behavior is closer to the expected utility theory benchmark in binary choice than in choice lists. Indeed, sensitivities in binary choice first order stochastically dominate sensitivities in choice lists, and subjects in binary choice are over twice as likely to be *likelihood oversensitive* as subjects in choice lists. This same pattern intensifies further in Losses where binary choice sensitivities strongly first order stochastically dominate choice list sensitivities. The median subject in binary choice is in fact slightly likelihood *oversensitive*, in sharp contrast with the standard PT account. In both cases, Wilcoxon tests suggest ($p < 0.01$) that subjects are significantly more sensitive to likelihoods in binary choice than in choice lists. Our results thus, once again, strongly echo those from our meta-analysis: likelihood insensitivity, a core descriptive element of prospect theory, is significantly weaker in binary choice than in choice list elicitation.

3.2 Robustness

Internal Consistency of Behavior. Given the importance of the fourfold pattern to the literature, it is natural to wonder whether these results are a consequence of lower data quality in binary choice relative to choice lists. After all, choice list tasks visually organize the same underlying tasks studied in binary choice, grouping these tasks by lottery and monotonically by certain payment. Perhaps this leads to more internally coherent behavior when choice lists are used that better reflects true risk preferences (à la fourfold pattern) than binary choice data do.

On net, this is a difficult interpretation to support. First, individual subjects in binary choice tasks reveal choice proportions that suggest more consistent risk postures across different lotteries than subjects do in choice list tasks. In Figure 9, we plot the rate at which subjects decrease their choices of the lottery as its probability of paying out *increases* (decrease their choice of the sure option for losses), a clear rationality violation,

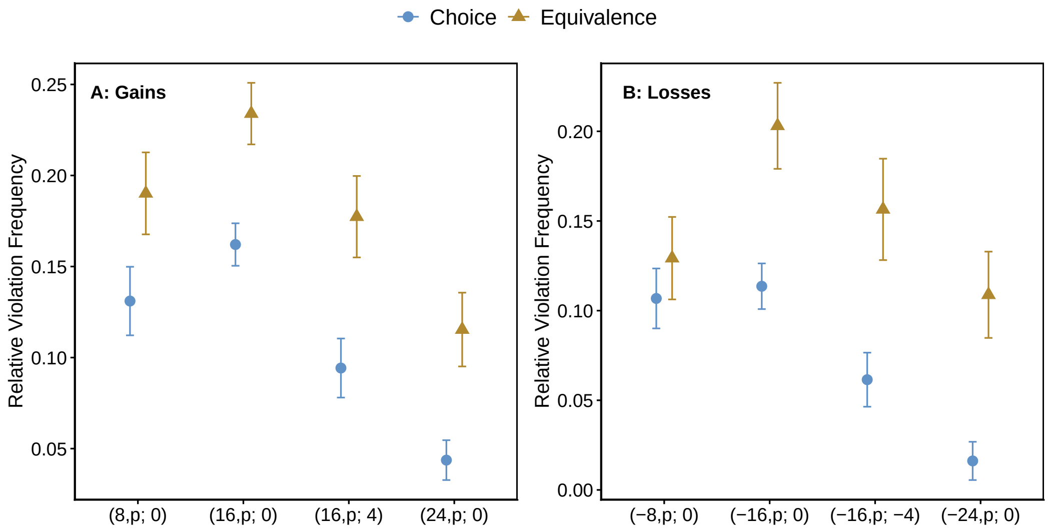


Figure 9: Between-lottery inconsistencies in Equivalence vs Choice

Panels A and B examine consistency violations in levels of risk aversion (risk-taking choices) between 'lists' as the probability increases for identical outcomes. These plots report the relative violation frequency, which is the violation frequency divided by the number of all possible violations.

for choice lists and binary choice. Clearly, subjects in binary choice are less likely to exhibit such inconsistent behavior than subjects in choice lists. By this measure, subjects in the binary choice treatment are thus clearly less noisy than subjects in the choice list treatment.

Second, in addition to making choices that suggest more stable risk postures across lotteries, subjects in binary choice also make individual decisions that are more consistent with one another when facing the same lottery twice, revealing higher test-retest reliability in choice proportions (a common measure of the noisiness of behavior). In the experiment, we assigned subjects two entire 'choice lists', $(8, 0.8; 0)$ and $(16, 0.8; 0)$ for gains and $(-8, 0.8; 0)$ and $(-16, 0.8; 0)$ for losses *twice* (at different points in the experiment) and calculated the proportion of the time the riskier option was selected. The test-retest reliability for choice lists on this metric is 0.590, with a 95% confidence interval of [0.513 , 0.658], a typical rate for choice list experiments. By comparison, the data points obtained using binary choice tasks have a *much higher* test-retest reliability of 0.844 with a 95% confidence interval of [0.808 , 0.874]. Similar observations hold for losses.[15] Thus despite the fact that subjects face individual choices in a random order

[15] The test-retest reliability for Equivalences in losses is 0.568, with a confidence interval of [0.465 , 0.656]. Once again, test-retest reliability is much larger for binary choice at 0.757, with a confidence interval of [0.692 , 0.810]. Reported coefficients and 95% CIs are from Pearson's correlation test. Spearman's ρ yields nearly identical estimates.

in binary choice, they express substantially more stable preferences for risk than they do in the seemingly more orderly choice list tasks.

Subjects are therefore more consistent in repetitions of the same lottery in binary choice than in choice lists. Subjects also show more consistent behavior *across* lotteries in binary choice, being more likely to increase their risk taking as the expected value of the lottery increases. The one sense in which subjects are plausibly *less* consistent is in their rate of "multiple switching" — i.e., in the rate at which subjects choose the risky option at a higher sure payment while also having chosen the sure option at a lower payment. In particular, in binary choice subjects are more likely to switch multiple times "within list" than they are in choice lists. However, in Appendix A.5 we show that these errors are far from random errors, and instead strongly peak near expected value in a region matching subjects' own risk postures. This orderliness and the fact that they occur at points likely near subjects' own certainty equivalents is strongly consistent with documentation of noisiness occurring in "regions of indifference" by Cubitt, Navarro-Martinez and Starmer (2015) and Agranov and Ortoleva (2017). These inconsistencies therefore (unlike instabilities in risk postures and noise in individual choices discussed above) seem at odds with confusion-driven random error and instead resemble standard psychometric trembles widely documented for (near-) indifferent subjects.

Summarizing, the data we report seem inconsistent with an explanation for our results rooted in lower data quality in binary choice relative to choice list tasks. At best, the evidence is mixed. But if anything the results seem more consistent with the opposite interpretation. This interpretation is indeed also supported by our structural estimations of prospect theory functionals (see details in Appendix B.4), which estimate much higher noise variance for choice lists compared to binary choice, suggesting again that behavior in the binary choice treatment is overall more consistent than behavior in choice lists in our data. Such errors are rarely reported and even less often discussed in the PT literature, since they have little to do with the inner workings of the model. Examining errors, however, allows us to 'aggregate' the different types of errors that can occur in Choice and Equivalence. Starting with Gains, the standard deviation of the residuals is 0.118 (SE = 0.003) in Equivalence, but much smaller at 0.037 (SE = 0.001) in Choice. The same holds for Losses, where we find errors of 0.129 (SE = 0.004) for Equivalence and 0.073 (SE = 0.002) in Choice. Binary choice thus induces lower errors than choice lists in our data.

A natural interpretation of these results is that they show that subjects express highly internally consistent preferences towards lotteries in binary choice but tremble near indifference (as documented in the vast literature on psychometrics in psychology). By contrast, choice list tasks induce an artificial internal consistency within-list, causing them to therefore express artificial certainty equivalents that are highly unstable because of their more distant relationship to true preferences. If this interpretation is correct, it is the fourfold pattern that is more likely to be an artifact of elicitation-driven errors in our data.

Endogenous reference points? Finally, in Online Appendix A.3 we show that we cannot use the standard machinery of prospect theory to rationalize these results, ex post. There, we rule out perhaps the most salient way PT might be used to rationalize the surprising difference between choice lists and binary choice in our data: endogenous reference points. If reference points are fixed at 0 in choice list tasks like ours (as is often suggested in the literature, e.g., Hershey and Schoemaker, 1985) but vary across binary choice tasks, this will produce scope for the expression of loss aversion in the latter, driving a wedge between the binary choice and choice lists environments. However in the Appendix we show that this wedge is incapable of fitting our data — no single loss aversion parameter, λ, can organize the data and therefore ex post rationalize the differences we observe across the treatments.

4 Parametric Estimates and Choice Patterns

The results of both our meta-analysis and experimental data paint a clear and unified picture: the fourfold pattern is (and apparently always has been) a phenomenon of choice lists, not a phenomenon of binary choices between lotteries. Given the centrality of the pattern to PT, why has the literature mostly failed to notice its absence from the very setting PT was, arguably, ultimately meant to explain? We think the answer is rooted in the fact that most of the literature on PT has largely been focused not on non-parametric assessment of risk postures, but instead on estimates of PT parameters. This is especially true of binary choice tasks in which it is somewhat more difficult to non-parametrically assess risk postures than in choice lists (where risk postures are readily available by comparing elicited certainty equivalents to expected values), and where data has therefore often been summarized via structural estimates.

Reliance on structural estimates, it turns out, makes it easy to miss failures of the fourfold pattern to appear, because such estimates invite the researcher to heuristically read parametric evidence of standard inverse-S shaped probability weighting as approximate evidence of the fourfold pattern. But this heuristic can badly fail because the probability weighting function is only one of the two elements of PT that contribute to risk attitudes. Sufficient curvature of the utility function (the second element of PT, alongside the probability weighting function) can lead even a very standard inverse S-shaped probability weighting function to coincide with risk attitudes that are starkly inconsistent with the fourfold pattern.[16] Indeed, our meta-analysis suggests that in past binary choice experiments this has *typically* been the case: while 75.7% of the results from prior binary choice studies estimate inverse-S shaped probability weighting functions as described by PT, only 11.4% produce certainty equivalents consistent with the fourfold pattern (none of which are significantly different from risk neutrality)!

Structural estimation of PT parameters. We can illustrate the problem by structurally estimating PT parameters on our own data using the most standard parameterization: (i) a standard power-utility value function and (ii) a 1-parameter weighting function.[17] Focusing on Gains, we estimate a likelihood-sensitivity of 0.542 (SE = 0.007) for our choice list task, yielding a standard inverse S-shaped probability weighting function as pictured in the left hand panel of Figure 10. In binary choice we estimate a sensitivity of 0.696 (SE = 0.009) indicating substantially less insensitive behavior that comes much closer to EU benchmarks (matching our non-parametric findings in Section 3). Nonetheless, as the right hand panel of Figure 10 shows, in binary choice we continue

[16]The probability weighting function can be first concave and then convex and thus inverse-S shaped, but stay entirely below the 45 degree line. Even a probability weighting function that crosses the 45 degree line and shows overweighting of small probabilities may actually occur in the presence of risk aversion for small probability gains (risk seeking for small probability losses) if utility is sufficiently concave (convex for losses).

[17]We use a standard Probit specification with noise variance proportional to the range of the lottery for our estimations (see Online Appendix B.4 for details). Random utility specifications such as the one we use are well known to potentially result in identification problems (Wilcox, 2011; Apesteguia and Ballester, 2018). Nevertheless, closely related models are pervasive in the literature, ensuring comparability of our estimates with previous results. Importantly, Apesteguia and Ballester (2018) show that random utility formulations may lead to an underestimation of utility curvature (risk aversion). Since we use the structural estimates to back out implied certainty equivalents, any such attenuation of curvature works against finding large differences in implied certainty equivalents across settings. In this sense, the structural results we report can be considered conservative.

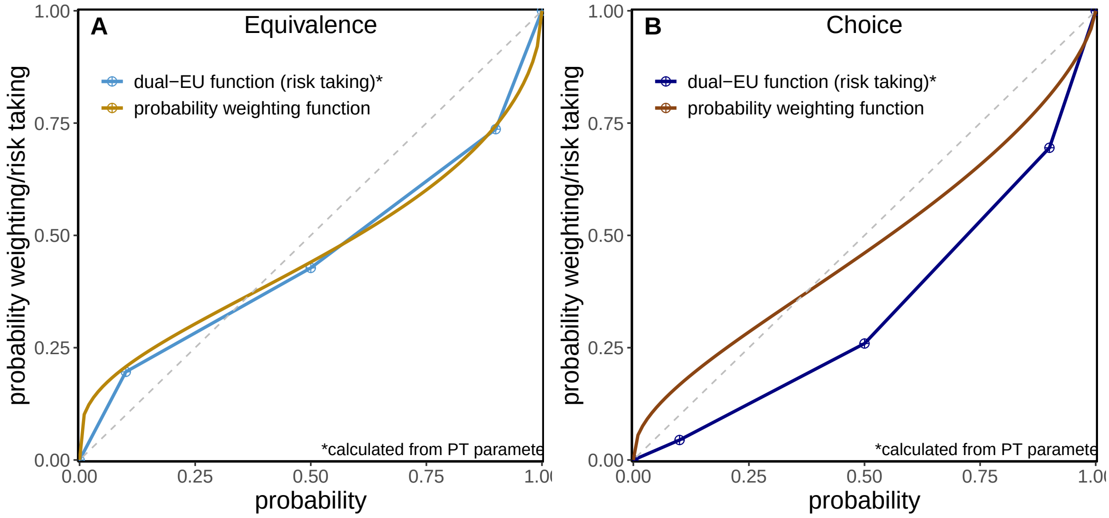


Figure 10: Probability weighting and risk taking

Panel A shows the parametric Prelec 1-parameter function estimated on our Equivalence data, with a likelihood-sensitivity parameter of 0.542 (SE 0.007) (estimates using a Tversky-Kahneman function are similar). The blue points indicate CEs calculated based on our estimated functionals. Given that utility is estimated to be close to linear (with a power utility parameter of 0.965, SE=0.007), implied CEs closely track the probability weighting function. Panel B shows the probability weighting function obtained from our Choice data, with a likelihood-sensitivity parameter of 0.696 (SE=0.009). The function is thus inverse-S shaped, just like for choice lists. The implied CEs, however, paint a very different picture, due to the power utility parameter that now indicates much more risk aversion (0.574, SE=0.005).

to find the expected inverse S-shaped weighting function. Were we to focus primarily on these estimates when assessing the fourfold pattern we would be lead to believe that the pattern arises in binary choice just as it does in choice lists. However, as we've seen from our analysis of the raw data, drawing this conclusion would be a significant mistake.

To see this, alongside the parametric estimates in each panel we also plot (in blue) *certainty equivalents*, calculated from the equation $\widehat{c} = u^{-1}\left[w(p)u(x)\right]$, where $w(p)$ indicates the probability weighting function, $u(x)$ the utility function, and u^{-1} indicates the inverse of the utility function. Unlike probability weights, these certainty equivalents take utility curvature into account and so can be directly compared to the expected value (EV) of the wager, indicating risk aversion if $\widehat{c} \leq EV$, and risk seeking if $\widehat{c} \geq EV$ (or equivalently, if $\widehat{c}/x \leq p$ and $\widehat{c}/x \geq p$, respectively). While these line up nicely with probability weighting estimates in choice lists, they imply fundamentally different risk postures in binary choice. Instead of showing the characteristic "flip" in apparent risk preferences at low vs. high probabilities implied by the parametric estimates, behavioral data in binary choice indicate that subjects are globally risk averse. The reason for this

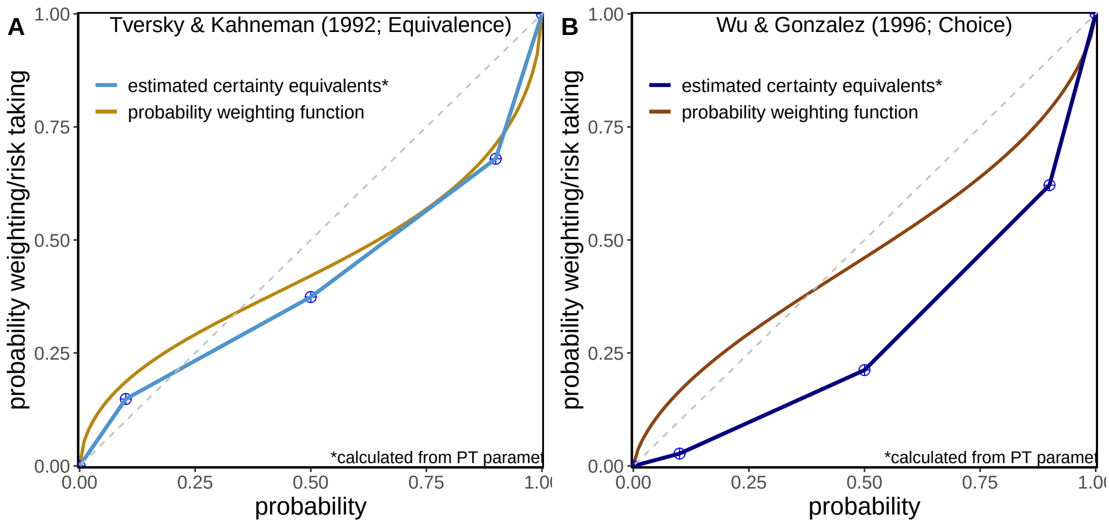


Figure 11: Probability weighting and risk taking

Panel A shows the parametric Tversky-Kahneman function estimated by Wu and Gonzalez (1996), with a curvature parameter $\gamma = 0.71$. The blue points indicate CEs calculated based on their estimated functions. While the parametric function shows overweighting of small probabilities, risk attitudes indicate risk *aversion* for small probability gains. The discrepancy is explained by their power utility coefficient, which at $\rho = 0.5$ indicates substantial risk aversion. Panel B shows the parametric probability weighting function estimated by Tversky and Kahneman (1992) from certainty equivalents. In this case, the overweighting of small probabilities of a gain indeed translates into risk seeking. The much smaller gap between the two functions can be traced back to utility curvature, which is much less pronounced at $\rho = 0.88$.

disconnect is that these identical estimation exercises yield almost no utility curvature for subjects in choice lists but substantial utility curvature in Choice. While power utility estimates for choice lists in our structural estimations are nearly linear (risk neutral) in choice lists with $\rho = 0.965$ (SE = 0.007), utility is extremely concave in binary choice with an estimate of $\rho = 0.574$ (SE = 0.005).[18] To contextualize these results in terms of the literature, in Figures 1-4 we plot certainty equivalents inferred from our structural estimates alongside similarly computed estimates in our meta-analysis, and we recover exactly the same pattern our non-parametric analysis suggests: a fourfold pattern in choice lists and twofold pattern in binary choice.

Similar patterns and opportunities for confusion have arisen throughout the literature's history, stretching back to the earliest efforts to estimate PT functionals. To illustrate, consider two of the landmark studies in the early literature: Tversky and Kahneman (1992) (a choice list study) and Wu and Gonzalez (1996) (a binary choice study). As

[18]Similar findings also obtain for losses, where choice lists produce sensitivity 0.804 (SE = 0.014) and utility 0.945 (SE = 0.01), whereas binary choice results in sensitivity 1.050 (SE = 0.015) and utility curvature $\rho = 0.794$ (SE = 0.009). It has often been remarked in the literature that curvature in losses is less pronounced than in gains (Abdellaoui, 2000; L'Haridon and Vieider, 2019), so that the failure to find insensitivity in the loss domain is hardly remarkable.

Figure 11 shows, estimated likelihood sensitivity is nearly identical in the two datasets, producing virtually identical inverse S-shaped probability weighting functions. However, as in our data, estimates suggest very little utility curvature in the Tversky and Kahneman (1992) choice lists study (a power estimate of $\rho = 0.88$) but pronounced curvature in the Wu and Gonzalez (1996) binary choice study ($\rho = 0.5$). When we calculate certainty equivalents from these estimates as we did with our own data (plotted in blue in Figure 11), we find strikingly similar evidence to ours. Probability weighting estimates are a good approximation of certainty equivalents in the Tversky and Kahneman (1992) choice lists study and therefore of risk attitudes. But, the heuristic value of probability weighting for assessing risk attitudes falls apart in binary choice: as in our data, there is a significant gulf between probability weights and certainty equivalents in the Wu and Gonzalez (1996) binary choice study, where certainty equivalents are starkly inconsistent with the fourfold pattern.

Discussion. The similarity of Figures 10 and 11 strongly reinforces our main observation in this section: in an important sense our experimental findings are not new, but instead have been with us since the beginning. However, they have gone largely unnoticed probably because of the literature's intensive focus on structural parameters rather than their implications for risk attitudes. This is perhaps understandable because, as we've discussed, the distinctive character of the fourfold pattern is ultimately attributable to probability weighting so that it is perhaps natural to treat evidence of the latter as implicit evidence of the former. In order to identify the shortcomings of this heuristic inference, the literature would have had to calculate certainty equivalents by joining evidence from the probability weighting function with evidence from the utility function as we have above. The literature has rarely done this and so has missed the regularity we document. For instance Wu and Gonzalez (1996) emphasize the importance of the fourfold pattern as one of the "critical empirical regularities that any good descriptive model should accommodate [...]: risk aversion for most gains and low probability losses, and risk seeking for most losses and low probability gains" (p. 1676) but (like most papers in the literature) do not seem to have directly assessed whether their estimates imply the pattern. If they had, they would have noticed that prospect theory, estimated on their data, fails this criterion for a good descriptive model.[19]

[19] The failure to detect the absence the fourfold pattern is particularly understandable in the case of binary choice because (unlike with choice lists), certainty equivalents are difficult to directly infer non-

Ultimately, the literature's failure to identify the collapse of the fourfold pattern in binary choice is downstream of a larger point the literature has, perhaps, under-emphasized: the fact that binary choice and choice list behaviors tend to differ significantly, even in parametric estimates of PT parameters. First, even when estimated probability weighting parameters are similar, estimated utility curvature tends to be substantially more severe in binary choice than in choice lists. Indeed, in our meta-analysis we find that power utility parameters are far higher in *choice lists* tasks (mean = 0.897, median = 0.910) than in *binary choice* tasks (mean = 0.614, median = 0.608) for gains, with similar divergences in losses.[20] Second, likelihood insensitivity — the core driver of probability weighting's distinctive shape — tends to be much less pronounced in binary choice than choice lists in the prior data (mean = 0.950 and 0.803 *vs.* 0.560 and 0.580). Both of these patterns were probably under-noticed because very few studies actually conduct head-to-head comparisons of the two settings as we do in our experiment.

5 Cognitive Frictions and the Equivalence-Choice Gap

Our results so far suggest that a core prediction of PT fails to arise in direct binary choice. But a more fundamental problem is that utility-based models like PT are not equipped to *explain* why there are significant differences between risk attitudes in binary choice and choice lists. These differences can be summed up as follows:

1. Likelihood insensitivity, the core characteristic of the *probability weighting function*, tends to be far *less severe* in binary choice than in choice lists.

2. Utility curvature, the core characteristic of the *utility function* tends to be far *more severe* in binary choice than in choice lists.[21]

3. Although less often measured, in datasets like ours, certainty equivalents tend to

parametrically from choice data and often have to be inferred indirectly based on certainty equivalents calculated from parameters. This is doubly true in Wu and Gonzalez (1996) who use a ladder design in which it is especially difficult to infer certainty equivalents based on raw choice patterns.

[20] In losses, power utility parameters are higher in *choice lists* tasks (mean = 1.130, median = 1.110) than in *binary choice* tasks (mean = 0.704, median = 0.700). Note, indeed, the smaller values indicate increased *convexity* for losses, and are thus an indication of risk *seeking*.

[21] Curvature is the core characteristic of the utility function when assessing unmixed lotteries, as we do in our paper. For mixed lotteries, the reference point and parameter of loss aversion are equally important. The evidence for loss aversion is currently also under review, with recent studies documenting stake-dependence of loss attitudes (Ert and Erev, 2013), increasing evidence that the parameter may be close to 1, or even fall below 1 (Chapman et al., 2025), and challenges to the explanatory power of the concept of loss aversion (Chapman et al., 2023*b*).

be *noisier* in choice lists than in binary choice.

It is important to emphasize that a failure to explain these differences is hardly unique to prospect theory — *no* standard preference-based theories of risk-taking can account for patterns like these. This is because standard theories of risk preferences (PT included) root risk postures entirely in the payoffs and probabilities underlying the lotteries being evaluated, which do not vary between the two choice environments. This is cast in particularly sharp relief in our experiment (Section 3), which was designed to feature *identical* menus, information and incentives in the two settings. Under the lens of preference-based theories, these tasks are identical.

Noisy coding. If preference-based theories cannot explain these regularities, what can? We consider the possibility that the widespread gulf we've documented between binary choice and choice lists is an outgrowth of the way cognitive frictions differentially distort behavior in the two settings. Our starting point is a groundswell of recent evidence suggesting that probability weighting itself represents, not an expression of preferences, but instead a severe cognitive distortion in the evaluation of lotteries. In particular, a class of recent *noisy coding* models has proved remarkably successful in organizing and predicting distinctive anomalies surrounding probability weighting (Khaw, Li and Woodford, 2021; 2023; Frydman and Jin, 2023; Enke and Graeber, 2023; Oprea, 2024*b*; Oprea and Vieider, 2024; Vieider, 2024). These models are rooted in the idea that (i) cognitive limitations cause decision makers to perceive or represent the descriptive primitives of lotteries (e.g., probabilities, payoffs) with *noise* and (ii) have prior beliefs about what these lottery primitives are. The key idea of noisy coding is that decision makers minimize the negative effects of their noisy perception by combining those perceptions with their prior belief in a standard, Bayesian manner. As a result, imperfections in the perception or representation of lottery primitives produce not just noise but systematic biases. In particular:

1. Bayesian shrinkage will produce likelihood insensitivity, systematically distorting the mapping between probabilities and values in a manner that exactly matches the standard inverse S-shape of standard probability weighting. The noisier perceptions are, the stronger the likelihood insensitivity.

2. Prior beliefs about the log-odds of the lottery paying an extreme amount will

systematically distort the apparent risk attitudes of decision makers, producing apparent curvature in estimates of utility functions. Such apparent risk aversion can further be distorted by noise, with noise resulting in an uplift of the prior and hence an apparent increase in risk taking (for gains) or risk aversion (for losses).

3. Noisiness in decision-makers' perceptions will generate noise in their *decisions* — the noisier perceptions are, the noisier choices will be (at least over some, empirically plausible, ranges).

Given these three implications (and their correspondence to the three differences between binary choice and choice lists listed above), to whatever degree binary choice and choice lists (i) generate different levels of noise in the perception of lottery primitives and (ii) distort prior beliefs about payoff distributions, we should expect likelihood sensitivity, utility curvature and behavioral noise to also differ across the two settings.

Procedural Differences Between Binary Choice and Choice Lists. Noisy cognition models, like most cognitive models (and unlike standard models of risk preferences like EU and PT), make predictions that depend on the *procedure* the decision-maker uses to make decisions. In particular the way cognitive acts are sequenced and arranged by the DM to make decisions will have substantial impacts on the formation of beliefs and the way cognitive noise evolves. This is important because there are strong reasons to think that the way decision makers are primed to process information is procedurally different in choice lists than in binary choice.

Intuitively, in binary choice the most obvious procedure for choosing between a lottery and a sure payment is for the DM to separately and independently evaluate the relative worth of each, and then choosing whichever seems more valuable. By contrast, when *valuing* a lottery (i.e., in choice lists), the decision maker *first* must assess the value of the lottery and only after doing so search over many possible sure payments for one that is equivalent to that lottery value. These two choice procedures are equivalent for decision-makers who suffer no cognitive frictions, but not for decision makers who suffer from the kinds of frictions described in noisy coding models.

In particular, in binary choice, noise in the independent imprecise evaluations DMs make of lotteries and rewards will tend to cancel one another out when the two are compared to make a choice, limiting (or even eliminating) the effects of cognitive noise and sharply attenuating the three implications of noisy coding sketched above. The result will be weak likelihood dependence, high risk aversion (high apparent utility curvature in estimation) and relatively low-noise decisions. By contrast, the sequential procedure of first evaluating the lottery and then iteratively searching for an equivalent value, as required in choice lists, will tend to produce the opposite. Evaluating potential outcomes by comparison to the initially valued lottery will introduce an additional Bayesian bias to the later evaluation of outcomes that will *intensify* rather than attenuate cognitive noise. Following the three implications of noisy coding models discussed above, this will produce strong likelihood dependence, low apparent risk aversion and relatively noisy decisions. Thus, noisy coding, when applied to the different procedures induced by choice list elicitations and binary choice, produce exactly the three differences between the two we summarized at the beginning of this section.

The technical arguments underlying these implications are subtle and require significant notation, so we defer them to Online Appendix C. There we offer a noisy coding model based on Vieider (2024), which uses technical assumptions suggested by recent neuroscience. However, the implications we draw (as sketched above) do not depend on the idiosyncrasies of our model, but rather seem to be general implications of noisy coding applied to binary choice-like and choice list-like evaluation procedures. For instance Khaw, Li and Woodford (2023), in contemporaneous work, draw similar conclusions about valuation tasks using a noisy coding setup building on the somewhat different modeling choices used in Khaw, Li and Woodford (2021).

Implications. The most important implication of this is that coding models like ours *predict* the three key distinctions between choice lists and binary choice that we have documented in our experiment and in the prior literature, and therefore ultimately explain why the fourfold pattern appears in choice lists but not in binary choice. Such models predict that the *reason* the fourfold pattern does not appear in binary choice is ultimately that the pattern itself is a consequence of cognitive errors that are intensified in choice lists relative to binary choice. In particular, the fourfold pattern is a consequence of the fact that noisy evaluation in choice lists produces *both* intense likelihood insensitivity and artificial apparent risk neutrality (with the corollary of highly inconsistent behavior between tasks), the combination of which are necessary for the fourfold pattern to arise. Thus, in addition to explaining the behavioral gap between binary choice and choice lists, noisy coding offers an interpretation of the key patterns of prospect theory as growing not out of risk preferences, but of the way the brain manages cognitive frictions.

Of course, the measure of an explanation like ours is its *excess explanatory power* — its ability to explain further phenomena that it wasn't designed to account for. One crucial implication of our explanation is that the difference between choice lists and binary choice actually isn't caused by the list format or the orderliness of typical choice list tasks etc., but instead to something subtle in the decision-making procedure choice list tasks tend to induce. In particular, it is the fact that choice lists require information to be processed *sequentially* that our model suggests generates the difference in behavior. To value a lottery, the DM must first form an assessment of the lottery and then subsequently assess certain payments to find one that equalizes with this value. Simply put, this is a more difficult and noisier process than direct binary choice.[22]

This suggests a distinctive test for our explanation of the gap between binary choice and choice lists. If, as our model suggests, it is the order of evaluation of information that generates the gap, we should be able to cause binary choice behavior to converge towards choice list behavior simply by forcing subjects to evaluate binary choice problems in a way similar to choice lists. In particular, if we induce subjects to evaluate a lottery up front and inform the subjects that they will be asked to subsequently contrast this with certain payments in a series of binary choices, then our explanation suggests that we should see some degree of convergence in likelihood dependence, utility curvature and noise and the emergence of distinctive features of the fourfold pattern. We report just such a test next.

5.1 An Experimental Test

To test the hypothesis discussed at the end of the previous subsection, we conducted an experiment on Prolific UK using a similar design to the experiment reported in Section 3,

[22]Notably, this explanation is highly consistent with the fact that the fourfold pattern *does* arise under the BDM mechanism used to elicit valuations, an alternative mechanism to choice lists for eliciting certainty equivalents (see e.g. Chapman et al., 2023*a*, for evidence of very high correlations between valuations elicited in choice lists and valuations obtained from BDM mechanisms). This suggests that the gap between binary choice and choice lists is driven not by details of the elicitation mechanism, but instead by the cognitive act of valuation itself.

but with a simpler set of lotteries.[23] In the experiment we repeated our choice list versus binary choice designs, and introduced a new treatment we call Sequential-binary choice. Approximately 50 subjects participated in each condition, resulting in 150 participants for the entire experiment.

In our novel Sequential-binary choice experiment, we attempt to induce subjects to reason about binary choice in a choice list-like, sequential way. For each probability, subjects are *first* shown the lottery and asked to evaluate it. They are then given a series of binary choices between the lottery they have just evaluated and certain payments in a random order. Thus the only real differences relative to our binary choice treatment is (i) binary choices are temporally grouped by lottery and (ii) subjects are asked to consider the lottery prior to making those choices.

Results. Figure 12 shows the results. Panel A shows the choice proportions for each treatment with the 45 degree line plotting the risk neutral benchmark: choice proportions above (below) this line reveal evidence of apparent risk seeking (aversion). The choice list and binary choice conditions replicate the main results from the Gains treatment of the experiment in Section 3 above. In particular, subjects in choice lists are risk seeking at low and risk averse at high probabilities, replicating the gains component of the fourfold pattern. However in binary choice, we find uniform risk aversion, consistent with the twofold pattern we've documented in our experiment and the prior literature. What's more, we see non-parametric evidence of the three key differences we opened this section with and which our model predicts. Panel A shows that choice proportions change much more gradually in choice lists than in binary choice (evidence of greater likelihood insensitivity), and the lower risk choice at $p = 0.5$ in binary choice than in choice lists is consistent with far greater utility curvature. In Panel B, we show that as in earlier work, behavior is considerably more noisy in choice lists than in binary choice (as measured by between-task inconsistencies in observed choice proportions).

Our main finding, however, is that these gaps in behavior almost entirely disappear in our Sequential-binary choice treatment. Simply by forcing subjects to evaluate the lottery intensively prior to making binary choices, we cause the fourfold pattern to re-appear and indeed for choice proportions to become *virtually identical* to our choice

[23] In particular, subjects evaluated lotteries that paid £24 with probabilities of 0.1, 0.3, 0.5, 0.7, or 0.9, and £0 otherwise.

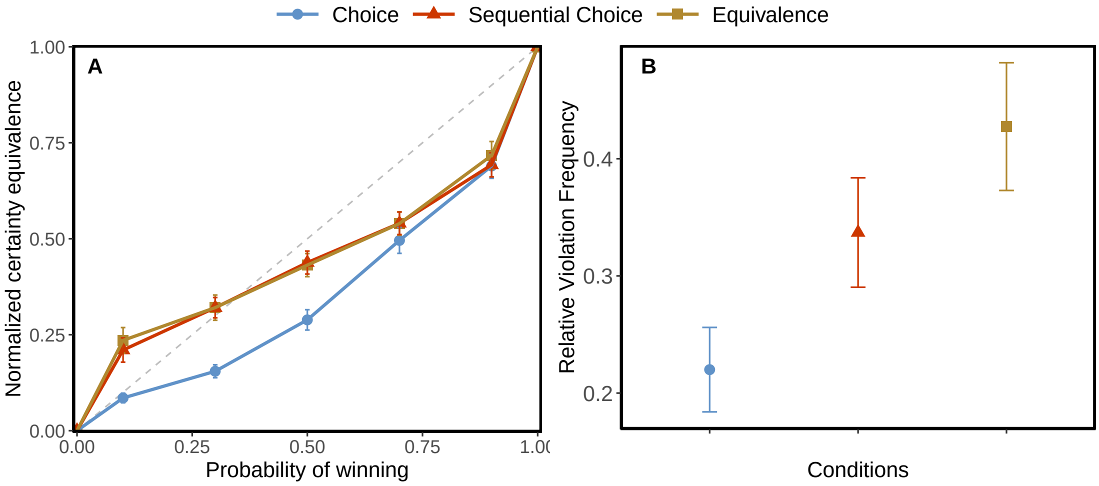


Figure 12: binary choice patterns in Equivalence, binary choice, and Sequential-binary choice. Panel A plots risky choice proportions from each of the three treatments. See corresponding statistics in Table A.3. Panel B plots inconsistency in choice proportions across lotteries, defined as the choice proportion of the lottery declining as its expected value increases (i.e., as the probability of winning increases, since all else is kept constant).

list treatment. We also find that choice inconsistency rises in this treatment relative to binary choice, evidence for a distinctive claim of our model — that sequential evaluation produces noisier cognition and behavior.

Summarizing then, our experimental results provide strong evidence in favor of, perhaps, the most distinctive implication of our explanation. The highly likelihood-insensitive probability weighting and near-linearity of utility that drives the fourfold pattern in choice lists are not actually expressions of subjects' risk preferences. Rather they are outgrowths of the way decision makers procedurally approach the particularly difficult task of valuation in choice lists, which requires a sequential mode of evaluation that tends to amplify perceptual and evaluative biases. Simply inducing this style of evaluation in binary choice is sufficient to induce these same amplifications and thereby causes binary choices to obey the predictions of prospect theory.

6 Discussion

In this paper, we provide evidence that one of the core empirical patterns in behavioral economics — the fourfold pattern of risk — is an apparent outgrowth of the choice environments typically used to directly elicit preferences (e.g., choice lists). By contrast, it typically does not arise in direct choices between lotteries, where it tends to be replaced by a very different pattern of risk-taking. Meta-analyzing decades worth of previous studies, we find that this has always been true in the literature, but has gone largely unnoticed. Using new experiments in which binary choice and choice lists are *identical* under the lens of any theory of preferences (including prospect theory), we find the same disjunction between choice and direct elicitation. These results suggest that both the gap between binary choice and choice lists and the fourfold pattern itself are a consequence of cognitive limitations on the part of decision makers rather than the maximization of stable preferences. We develop a model of noisy cognition that rationalizes both the gap and the fourfold pattern and provide novel experimental evidence in favor of this model's distinctive implications.

We believe these findings have two primary implications.

First, because the fourfold pattern is "the most distinctive implication of prospect theory," (Tversky and Kahneman, 1992) our results help to inform long-running debates over how prospect theory should be interpreted. Prospect theory is and always has been a *descriptive theory* and the literature has long been ambivalent over what it is exactly the theory describes: (i) stable, welfare-relevant preferences or (ii) a suite of heuristic, cognitive adaptations. We believe that our finding that core patterns predicted by the theory change sharply (i.e., likelihood sensitivity) or disappear altogether (i.e., the fourfold pattern) in response to superficial changes to the choice environment lends significant weight to intepretation (ii). We therefore view our findings as contributing to a growing body of evidence suggesting that prospect theory describes a body of boundedly-rational cognitive adaptations rather than rationally expressed tastes for risk.

Second and more generally, our results highlight the value of complementing and fortifying tractable, descriptive theories like prospect theory with explicit models of the cognitive processes underlying the decisions these theories are built to describe. Such models, in an important sense, endogenize the parameters of descriptive theories like prospect theory (Vieider, 2025), improving our ability to interpret them and helping us to understand and predict when they are likely to change in response to payoff-irrelevant factors. Indeed, we show that a popular class of cognitive models ("noisy coding" models) can explain both the fourfold pattern and its disappearance in binary choice as an outgrowth of the imperfect way humans internally represent lottery primitives. Nonetheless, there are legitimate questions concerning whether the added insight cognitive models provide are worth their often higher complexity and sometimes lower tractability. We believe that our finding that the most important prediction of, arguably, the most important model in behavioral economics is unstable in the face of superficial aspects of the choice environment (and disappears in a setting as important as binary choice) makes a strong case in favor of the complementary development and use of such models.

References

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

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

Apesteguia, Jose, and Miguel A Ballester. 2018. "Monotone stochastic choice models: The case of risk and time preferences." *Journal of Political Economy*, 126(1): 74–106.

Barretto-García, Miguel, Gilles de Hollander, Marcus Grueschow, Rafael Polania, Michael Woodford, and Christian C Ruff. 2023. "Individual risk attitudes arise from noise in neurocognitive magnitude representations." *Nature Human Behaviour*, 7(9): 1551–1567.

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

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

Brown, Alexander L, and Paul J Healy. 2018. "Separated decisions." *European Economic Review*, 101: 20–34.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer. 2024. "Meta-Analysis of Empirical Estimates of Loss Aversion." *Journal of Economic Literature*, 63(3): 485–616.

Chapman, Jonathan, Erik Snowberg, Stephanie W Wang, and Colin Camerer. 2025. "Looming large or seeming small? Attitudes towards losses in a representative sample." *Review of Economic Studies*, 92(5): 2893–2922.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2023*a*. "Econographics." *Journal of Political Economy Microeconomics*, 1(1): 115–161.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2023*b*. "Willingness to Accept, Willingness to Pay, and Loss Aversion." Working paper.

Chark, Robin, Soo Hong Chew, and Songfa Zhong. 2020. "Individual preference for longshots." *Journal of the European Economic Association*, 18(2): 1009–1039.

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

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

Enke, Benjamin. 2024. "The Cognitive Turn in Behavioral Economics." Working paper.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *The Quarterly Journal of Economics*, 138: 2021–2067.

Enke, Benjamin, Thomas Graeber, Ryan Oprea, and Jeffrey Yang. 2024. "Behavioral attenuation." Working paper.

Ert, Eyal, and Ido Erev. 2013. "On the descriptive value of loss aversion in decisions under risk: Six clarifications." *Judgment and Decision making*, 8(3): 214–235.

Fehr-Duda, Helga, Thomas Epper, Adrian Bruhin, and Renate Schubert. 2011. "Risk and rationality: The effects of mood and decision rules on probability weighting." *Journal of Economic Behavior & Organization*, 78(1): 14–24.

Feldman, Paul J, and Paul J Ferraro. 2023. "A Certainty Effect for Preference Reversals Under Risk: Experiment and Theory." Working paper.

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, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder. 2017. *Risky curves: on the empirical failures of expected utility.* Routledge.

Friedman, Daniel, Sameh Habib, Duncan James, and Brett Williams. 2022. "Varieties of risk preference elicitation." *Games and Economic Behavior*, forthcoming.

Frydman, Cary, and Lawrence J. Jin. 2023. "On the Source and instability of probability weighting." Working paper.

Furukawa, Toshi A, Corrado Barbui, Andrea Cipriani, Paolo Brambilla, and Norio Watanabe. 2006. "Imputing missing standard deviations in meta-analyses can provide accurate results." *Journal of clinical epidemiology*, 59(1): 7–10.

Glimcher, Paul W, and Agnieszka A Tymula. 2023. "Expected subjective value theory (ESVT): A representation of decision under risk and certainty." *Journal of Economic Behavior & Organization*, 207: 110–128.

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

Grether, David M, and Charles R Plott. 1979. "Economic theory of choice and the preference reversal phenomenon." *American Economic Review*, 69(4): 623–638.

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." *Review of Economic Studies*, 88(4): 1979–2013.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2023. "Cognitive imprecision and stake-dependent risk attitudes." Working paper.

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.

Lichtenstein, Sarah, and Paul Slovic. 1971. "Reversals of preference between bids and choices in gambling decisions." *Journal of experimental psychology*, 89(1): 46.

Li, Yuchi, and Ferdinand M. Vieider. 2025. "Meta-analysis of incentive effects on risk-taking and delay-discounting." Working Paper.

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.

McGranaghan, Christina, Kirby Nielsen, Ted O'Donoghue, Jason Somerville, and Charles D Sprenger. 2024. "Distinguishing common ratio preferences from common ratio effects using paired valuation tasks." *American Economic Review*, 114(2): 307–347.

Natenzon, Paulo. 2019. "Random choice and learning." *Journal of Political Economy*, 127(1): 419–457.

Netzer, Nick. 2009. "Evolution of time preferences and attitudes toward risk." *American Economic Review*, 99(3): 937–55.

Netzer, Nick, Arthur Robson, Jakub Steiner, and Pavel Kocourek. 2024. "Risk perception: measurement and aggregation." *Journal of the European Economic Association*, jvae053.

Oprea, Ryan. 2024*a*. "Complexity and its measurement." Working paper.

Oprea, Ryan. 2024*b*. "Decisions under risk are decisions under complexity." *American Economic Review*, 114(12): 3789–3811.

Oprea, Ryan, and Ferdinand M. Vieider. 2024. "Minding the Gap: On the Origins of Probability Weighting and the Description-Experience Gap." Working paper.

Prat-Carrabin, Arthur, and Michael Woodford. 2022. "Efficient coding of numbers explains decision bias and noise." *Nature Human Behaviour*, 6(8): 1142–1152.

Robson, Arthur J. 2001. "Why would nature give individuals utility functions?" *Journal of Political Economy*, 109(4): 900–914.

Ruggeri, Kai, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, Emir Demić, Celia Esteban-Serna, Maja Friedemann, et al. 2020. "Replicating patterns of prospect theory for decision under risk." *Nature human behaviour*, 4(6): 622–633.

Schmidt, Ulrich, Chris Starmer, and Robert Sugden. 2008. "Third-generation prospect theory." *Journal of Risk and Uncertainty*, 36(3): 203–223.

Shubatt, Cassidy, and Jeffrey Yang. 2024. "Tradeoffs and Comparison Complexity." Working paper.

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

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

Vieider, Ferdinand M. 2024. "Decisions under uncertainty as bayesian inference on choice options." *Management Science*, 70(12): 9014–9030.

Vieider, Ferdinand M. 2025. "Neuro-biological origins of prospect theory functionals." In *Handbook of Prospect Theory.* , ed. Mohammed Abdellaoui and Han Bleichrodt. Cheltenham, UK:Edward Elgar Publishing.

Wakker, Peter, and Daniel Deneffe. 1996. "Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown." *Management Science*, 42(8): 1131–1150.

Wakker, Peter P. 2010. *Prospect Theory for Risk and Ambiguity.* Cambridge:Cambridge University Press.

Wilcox, Nathaniel T. 2011. "'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk." *Journal of Econometrics*, 162(1): 89–104.

Wu, George, and Richard Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science*, 42(12): 1676–1690.

Zhou, Wenting, and John Hey. 2018. "Context matters." *Experimental economics*, 21: 723–756.

ONLINE APPENDIX

Cognitive Frictions and Canonical Patterns in Risk-Taking

Ranoua Bouchouicha Ryan Oprea Ferdinand M. Vieider Jilong Wu

Table of Contents

A Experiments: Additional results

A.1 Stimuli and results Experiment I

Table A.1 presents the choice proportions of the lottery for gains by treatment.Table A.2 presents the choice proportions of the sure amount for losses by treatment. Both tables provide a nonparametric test for differences in choice proportions by treatment (called

Table A.1: Choice proportions and significance tests — Gains

Lottery	Cond.	Prop.	SE	Perm.	p_{between}
(8, 0.2; 0)	Choice	0.121	0.012	0.000***	0.000***
	Equiv.	0.319	0.019	0.000***	
(8, 0.5; 0)	Choice	0.316	0.017	0.000***	0.000***
	Equiv.	0.495	0.015	0.749	
(15, 0.5; 4)	Choice	0.348	0.017	0.000***	0.000***
	Equiv.	0.450	0.016	0.002***	
(16, 0.2; 0)	Choice	0.105	0.009	0.000***	0.000***
	Equiv.	0.266	0.015	0.000***	
(16, 0.3; 0)	Choice	0.131	0.011	0.000***	0.000***
	Equiv.	0.317	0.014	0.214	
(16, 0.5; 0)	Choice	0.270	0.017	0.000***	0.000***
	Equiv.	0.441	0.015	0.000***	
(16, 0.7; 0)	Choice	0.465	0.021	0.000***	0.003***
	Equiv.	0.549	0.016	0.000***	
(16, 0.1; 4)	Choice	0.114	0.012	0.232	0.004***
	Equiv.	0.200	0.019	0.000***	
(16, 0.5; 4)	Choice	0.330	0.018	0.000***	0.000***
	Equiv.	0.430	0.017	0.000***	
(16, 0.9; 4)	Choice	0.682	0.019	0.000***	0.212
	Equiv.	0.705	0.020	0.000***	
(16, 0.5; 5)	Choice	0.352	0.018	0.000***	0.001***
	Equiv.	0.417	0.016	0.000***	
(17, 0.5; 4)	Choice	0.344	0.017	0.000***	0.000***
	Equiv.	0.426	0.014	0.000***	
(24, 0.1; 0)	Choice	0.055	0.006	0.000***	0.000***
	Equiv.	0.193	0.018	0.000***	
(24, 0.5; 0)	Choice	0.262	0.016	0.000***	0.000***
	Equiv.	0.424	0.015	0.000***	
(24, 0.9; 0)	Choice	0.658	0.020	0.000***	0.040**
	Equiv.	0.712	0.019	0.000***	
(8, 0.8; 0)	Choice	0.571	0.021	0.000***	0.002***
	Equiv.	0.668	0.018	0.000***	
(8, 0.8; 0)	Choice	0.586	0.022	0.000***	0.000***
	Equiv.	0.694	0.018	0.000***	
(16, 0.8; 0)	Choice	0.557	0.022	0.000***	0.063*
	Equiv.	0.621	0.018	0.000***	
(16, 0.8; 0)	Choice	0.575	0.022	0.000***	0.299
	Equiv.	0.616	0.018	0.000***	
(24, 0.4; 12)	Choice	0.314	0.018	0.000***	0.000***
	Equiv.	0.438	0.021	0.075*	
(24, 0.6; 12)	Choice	0.425	0.020	0.000***	0.293
	Equiv.	0.467	0.022	0.000***	

List of choice tasks with proportion of risky choices per treatment condition. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Perm. is a one-sample permutation test against risk-neutral proportion $\mu = p$. p_{between}: Wilcoxon rank-sum test across conditions.

Table A.2: Choice proportions and significance tests — Losses

Lottery	Cond.	Prop.	SE	Perm.	p_{between}
$(-8, 0.2; 0)$	Choice	0.132	0.014	0.000***	0.295
	Equiv.	0.169	0.020	0.121	
$(-8, 0.5; 0)$	Choice	0.370	0.022	0.000***	0.001***
	Equiv.	0.474	0.022	0.231	
$(-15, 0.5; -4)$	Choice	0.337	0.020	0.000***	0.000***
	Equiv.	0.442	0.020	0.004***	
$(-16, 0.2; 0)$	Choice	0.139	0.014	0.000***	0.001***
	Equiv.	0.229	0.021	0.166	
$(-16, 0.3; 0)$	Choice	0.201	0.017	0.000***	0.000***
	Equiv.	0.320	0.021	0.352	
$(-16, 0.5; 0)$	Choice	0.409	0.020	0.000***	0.006***
	Equiv.	0.489	0.019	0.585	
$(-16, 0.7; 0)$	Choice	0.717	0.018	0.336	0.008***
	Equiv.	0.644	0.021	0.009***	
$(-16, 0.1; -4)$	Choice	0.094	0.014	0.694	0.207
	Equiv.	0.109	0.020	0.671	
$(-16, 0.5; -4)$	Choice	0.339	0.019	0.000***	0.005***
	Equiv.	0.421	0.022	0.000***	
$(-16, 0.9; -4)$	Choice	0.805	0.017	0.000***	0.843
	Equiv.	0.760	0.028	0.000***	
$(-16, 0.5; -5)$	Choice	0.340	0.020	0.000***	0.035**
	Equiv.	0.399	0.020	0.000***	
$(-17, 0.5; -4)$	Choice	0.358	0.020	0.000***	0.005***
	Equiv.	0.425	0.019	0.000***	
$(-24, 0.1; 0)$	Choice	0.084	0.013	0.245	0.131
	Equiv.	0.130	0.018	0.105	
$(-24, 0.5; 0)$	Choice	0.462	0.020	0.070*	0.596
	Equiv.	0.497	0.018	0.890	
$(-24, 0.9; 0)$	Choice	0.858	0.012	0.000***	0.809
	Equiv.	0.795	0.025	0.000***	
$(-8, 0.8; 0)$	Choice	0.750	0.020	0.016**	0.163
	Equiv.	0.703	0.024	0.000***	
$(-8, 0.8; 0)$	Choice	0.753	0.018	0.011**	0.111
	Equiv.	0.679	0.027	0.000***	
$(-16, 0.8; 0)$	Choice	0.783	0.016	0.285	0.005***
	Equiv.	0.692	0.025	0.000***	
$(-16, 0.8; 0)$	Choice	0.782	0.017	0.267	0.002***
	Equiv.	0.694	0.023	0.000***	
$(-24, 0.4; -12)$	Choice	0.263	0.018	0.000***	0.001***
	Equiv.	0.361	0.023	0.087*	
$(-24, 0.6; -12)$	Choice	0.445	0.022	0.000***	0.271
	Equiv.	0.486	0.025	0.000***	

List of choice tasks with proportion of sure amount choices per treatment condition. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$. Perm. is a one-sample permutation test against risk-neutral proportion $\mu = p$. p_{between}: Wilcoxon rank-sum test across conditions.

$p_{between}$) and a permutation tests against the probability benchmark indicating risk-neutrality. For gains, 17 our of 21 comparisons between treatments show comparisons significant at conventional levels. The exceptions for this rule all occur for moderately large probabilities. Similar observations hold for losses, where 12 our of 21 comparisons are significant at conventional levels (and no comparison is significant in the opposite direction). Exception to the rule again mainly (but not exclusively) include large probability lotteries.

Tests against risk-neutrality benchmarks are most relevant for small probabilities. Here, tests tend to be significant in opposite directions for the two treatments: In Equivalence tasks, choice proportions tend to exceed the risk neutrality benchmark, indicating risk seeking for small probability gains and risk aversion for small probability losses. For binary choice, most tests indicate risk aversion for small probability gains and large probability losses. There are a few exceptions from this rule, in which risk neutrality cannot be rejected. This notably includes lotteries with nonzero lower outcomes. Weakened patterns are expected for such lotteries based on decreasing absolute risk aversion (Bouchouicha and Vieider, 2017*a*). We also note that risk neutrality cannot be statistically rejected for the lottery yielding a loss of £24 with a probability $p = .1$ based on two out of the four tests we report, and that the difference between treatments is not significant for this particular lottery.

A.2 Stimuli and results Experiment II

Table A.3 presents the choice proportions for the three treatment conditions by task (24, p; 0). This analysis employs nonparametric tests to assess the differences in calculated choice proportions. Notably, we observe that subjects' choices in Sequential-binary choice and choice lists do not differ significantly (refer to Column *p-value (3)*), as illustrated in Figure 12. In contrast, it is evident that for small and moderate probabilities ($p = 0.1, 0.3, 0.5$), subjects answering the binary choice task opted for fewer risky options compared to their counterparts in both the Sequential-binary choice and choice lists conditions.

Table A.3: Choice proportions by treatment over (24, p; 0)

				vs. risk-neutral				between conditions		
p	Condition	Prop.	SE	Wilcoxon	Sign	t-test	Perm.	Ch–Seq	Ch–Eq	Seq–Eq
0.1	Choice	0.085	0.013	0.105	0.015**	0.247	0.248	0.002***	0.000***	0.368
	Equivalence	0.235	0.033	0.000***	0.024**	0.000***	0.000***			
	Seq. Choice	0.210	0.032	0.004***	0.220	0.001***	0.000***			
0.3	Choice	0.155	0.017	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.799
	Equivalence	0.321	0.033	0.922	1.000	0.535	0.535			
	Seq. Choice	0.320	0.027	0.564	0.341	0.444	0.454			
0.5	Choice	0.289	0.027	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.780
	Equivalence	0.431	0.030	0.009***	0.011**	0.027**	0.026**			
	Seq. Choice	0.438	0.030	0.007***	0.001***	0.043**	0.041**			
0.7	Choice	0.496	0.034	0.000***	0.000***	0.000***	0.000***	0.434	0.648	0.688
	Equivalence	0.540	0.031	0.000***	0.000***	0.000***	0.000***			
	Seq. Choice	0.540	0.029	0.000***	0.000***	0.000***	0.000***			
0.9	Choice	0.690	0.032	0.000***	0.000***	0.000***	0.000***	0.906	0.311	0.308
	Equivalence	0.717	0.036	0.000***	0.001***	0.000***	0.000***			
	Seq. Choice	0.692	0.031	0.000***	0.000***	0.000***	0.000***			

Proportion of risky choices per treatment condition. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Wilcoxon–Sign–t–Perm. are one-sample tests against the risk-neutral proportion $\mu = p$. Ch–Seq, Ch–Eq, Seq–Eq: Wilcoxon rank-sum tests between conditions.

A.3 Endogenous reference points

A key reason why choice list tasks are the tool of choice to elicit PT parameters is that they allow to exogenously fix the reference point to 0. This is essential if one wants to separately identify all the different components of PT, since in the presence of endogenous reference points all wagers would become mixed. Hershey and Schoemaker (1985) claimed 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 wagers involve both gains and losses, PT can no longer be fully identified. 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. To test this, we rescale the PT equation following Hershey and Schoemaker (1985):

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

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. We have all the elements to identify utility function curvature and probability weighting from choice lists tasks for both gains and losses. This implies that we can infer the loss aversion coefficient that would explain the discrepancy between choice lists and binary choice tasks from the following equation:

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

The exact parameters will of course depend on assumptions made about functional forms and errors. We use a 'standard' PT implementation. That is, we estimate a simple aggregate PT model from the choice lists data, by letting $u(x) = x^\rho$, with different parameters for gains and losses, entered in terms of absolute amounts. The probability weighting function adopt the one-parameter Prelec specification, again with different parameters for gains and losses. We estimate the model using Bayesian techniques in Stan (see section B.4 for an example of the code used). The priors used for the parameters are mildly regularizing, i.e. they are uninformative in the sense of being centred on neutral values ($\rho = 0$, $\gamma = 1$), 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, 95% of the probability mass is allocated to the interval between 0 and 7).

The estimations are executed based on a standard discrete choice Probit model, with a noise term defined on the value scale, and errors that are heteroscedastic depending on the length of a choice list. The choice probability is modeled as follows:

$$Pr[(x,p;y) \succ s] = \Phi\left[\frac{\pi u(x) + \tilde{\pi} u(y) - u(s)}{\sigma|x-y|}\right],$$

where Φ is the standard normal cumulative distribution function providing the 'link function', π indicates a decision weight, and $\tilde{\pi}$ is the decision weight associated to the complementary event. For gains and losses, $\pi = w(p)$ and $\tilde{\pi} = 1 - \pi$, whereas for mixed prospects $\pi = w^+(p)$ and $\tilde{\pi} = w^-(1-p)$, where $+$ and $-$ indicate the weighting function for gains and losses respectively. The likelihood function is then constructed by mapping the choice patterns for the risky option into choice probabilities via a Bernoulli density.

Once we have obtained the PT parameters from the choice lists data for gains and losses, we calculate the choice objects p, $x - s$ and s for each of the three tasks in figure 7, panel A, in the main text. We then inject the PT parameters estimated from choice lists. 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. Finally, we use the equations thus obtained to calculate the loss aversion parameter, λ, that would be needed to bridge the gap between choice lists and binary choice for each of the three data points displayed in the figure (i.e., for $p = \{0.1, 0.5, 0.9\}$).

For wagers offering £24 with probability $p = \{0.1, 0.5, 0.9\}$ or else 0, we find that the loss aversion coefficient needed to explain the gap between choice lists and binary choice for $p = 0.1$ is $\lambda_{0.1} = 2.833$ [2.680 ; 2.991]. The loss aversion coefficient for intermediate probabilities is $\lambda_{0.5} = 2.489$ [2.360 ; 2.616], and is thus significantly smaller. The loss aversion coefficient for large probabilities is $\lambda_{0.9} = 3.038$ [2.775 ; 3.321], which is larger than for both previous ones. A different coefficient is thus needed to close the gap for each probability, implying that loss aversion is not a viable explanation for the discrepancy between choice lists and binary choice we observe. Feldman and Ferraro (2023) have furthermore shown that even the gap between certainty equivalents and probability equivalents cannot actually be organized by loss aversion.

A.4 IRRA over stakes

Panel A in Figure A.1 shows a measure of relative risk aversion in the Gain tasks, given by the choice proportion of the wager subtracted from the probability of winning ($p - \frac{\widehat{c}-y}{x-y}$). 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 thus use tasks varying x, while keeping y fixed at 0 and p fixed at 0.5.

Figure A.1 shows the results. We see important level effects indicating generally higher levels of relative risk aversion in binary choice than in choice lists. Patterns for choice lists tasks indicate the typical increasing relative risk aversion (IRRA) documented in the literature (Holt and Laury, 2002; Bouchouicha and Vieider, 2017*b*; Di Falco and Vieider, 2022). A pattern of IRRA of similar magnitude is also observed in binary choice tasks, thus pointing to the robustness of the phenomenon. Panel B shows changes in relative risk aversion with stake size in Loss tasks. Most measures are negative, indicating

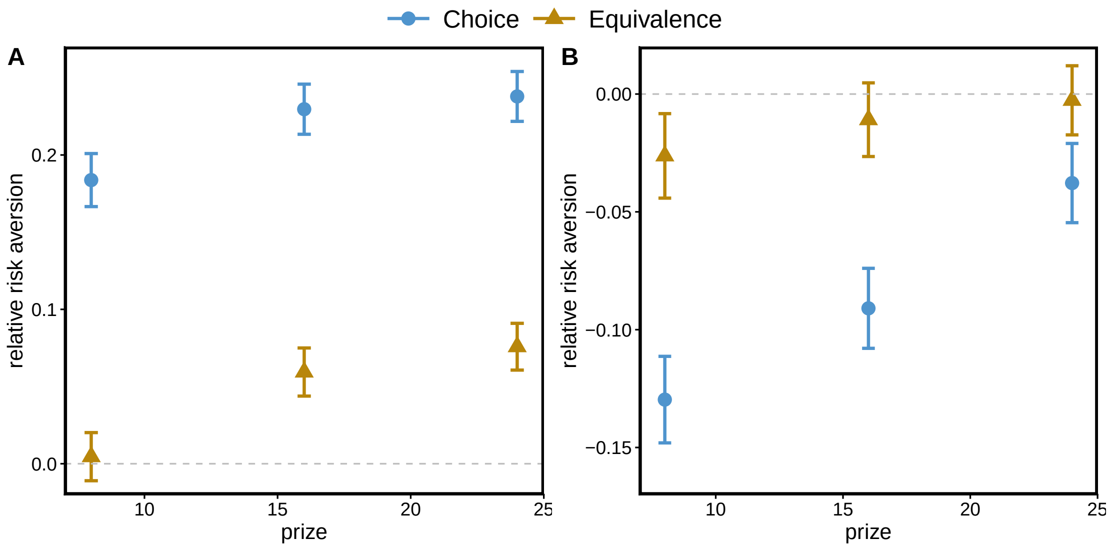


Figure A.1: Nonparametric measure of relative risk aversion by stake size

The figure plots a nonparametric measure of relative risk aversion, defined as $p - \frac{\hat{s}}{x}$, where $\hat{s}$ is stochastic switching point approximated by the choice proportion. The figure shows the evolution of the measure as the prize is increased from £8 to £24. Panel A shows the patterns for gains, and Panel B for losses.

a tendency towards risk seeking. In choice lists 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*b*). In binary choice, however, we observe clear evidence for increasing relative risk aversion. In other words, whereas in choice lists tasks the patterns for losses tend to differ from the ones for gains, as also documented in the previous literature, in binary choice we find convergent evidence for increasing relative risk aversion. Overall, increasing relative risk aversion is thus more robust in binary choice than in choice lists.

A.5 Binary choice consistency and errors

In the main text, we have shown how *violation* or *choice inconsistencies* across tasks are more severe in choice lists than binary choice. Here, we show that within-list inconsistencies – something akin to 'multiple switching' – is more severe in binary choice, but is nevertheless highly regular and concentrated around plausible points of indifference as shown in Figure A.2. As pointed out by Cubitt, Navarro-Martinez and Starmer (2015) and Agranov and Ortoleva (2017), such a pattern indicates those multiple switching is driven by indifference between two options.

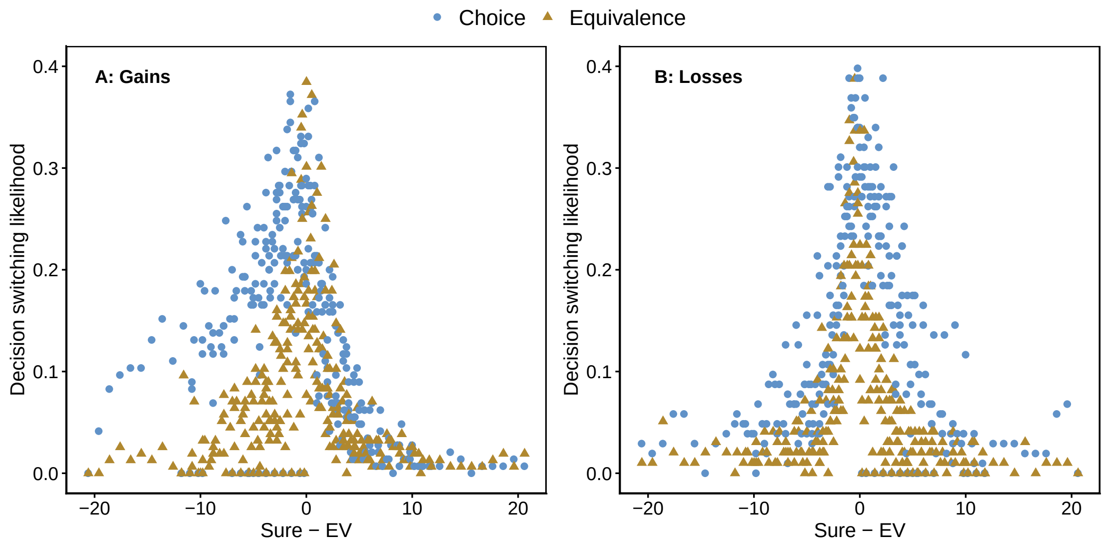


Figure A.2: Multiple Switching Likelihood for choice lists and binary choice.

The figure compares 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 expected value of the prospect, and the y-axis shows switching frequencies with each increase in the sure outcome.

B Meta-Analysis

B.1 Method

Paper selection. To substantiate our assertion that the absence of the four-fold pattern is not infrequent in the context of employing binary choices for the elicitation of risk preferences, we undertook a comprehensive meta-analysis. To ensure its unbiasedness, we rigorously selected relevant papers based on well-defined inclusion criteria. These criteria centered on the inclusion of experimental papers that estimated parameters of the probability weighting function under monetary risk, encompassing laboratory, lab-in-field, or online studies. Notably, this encompassed papers using pure binary choices or certainty equivalent choice lists as risk preference elicitation method in experiments. Regarding choice lists, specifically, we keep iterative certainty equivalent choice lists (e.g., Tversky and Kahneman, 1992) but drop those following bisection procedures (e.g., Abdellaoui, 2000). Our search for these papers primarily involved the scientific citation indexing database Web of Science. We initially screened titles and abstracts, and subsequently evaluated the remaining papers against our inclusion criteria, coding the relevant information. Additionally, we explored IDEAS/RePEc and Google Scholar for unpublished working papers to ensure a comprehensive review of the literature.

It is worth highlighting that certain studies encompass multiple estimations of prospect theory (or rank-dependent utility) parameters, owing to the adoption of various model specifications, the introduction of multiple treatment arms, or the examination of sub-sample variations. To ensure uniformity and facilitate comparative analysis across studies, we have employed the following filtering criteria:

- In the case of studies implementing diverse model specifications, characterized by varying utility functions and probability weighting functions, our initial step involves excluding those employing non-Constant Relative Risk Aversion (non-CRRA) utility functions. This step is imperative as our meta-analysis relies on the imputation of the certainty equivalent of a specific lottery $(100, 0.1; 0)$, and non-CRRA utility functions (e.g., exponential functions) can exert quantitative and even qualitative impacts on the calculated results. Subsequently, to align with the objectives of our meta-analysis, we prioritize the selection of estimations that either directly report the standard error or provide information that enables the approximate calculation of the standard error. Lastly, we opt for the estimation that is predominantly discussed or initially presented in the main body of the text.
- In instances where studies involve multiple treatments or explore specific population subsets, our focus is on selecting estimations associated with the control conditions or groups; Otherwise we are unable to know whether the presence or absence of the fourfold pattern is caused by the treatment or by population heterogeneity. For example, we do not incorporate observations stemming from treatments such as the sampling treatment in (Glöckner et al., 2016) or the outcome feedback treatment (Haffke and Hübner, 2014) in our analysis.

In terms of the availability of standard errors, a crucial element for our meta-analysis, it's noteworthy that 30.5% of estimations (N=36) within our dataset lack information that would be sufficient to calculate standard errors, including the seminal work by Tversky and Kahneman (1992). Among the remaining observations, a substantial majority, accounting for 76.8% (N=63), explicitly provide standard errors or other statistical metrics that allow for the precise calculation of standard errors, such as the inclusion of 95% confidence intervals. Additionally, there are 14 observations that present statistics such as 95% credible intervals, which, without certain assumptions, cannot be utilized directly in our analysis. In these cases, we adopt a conservative approach to ensure data retention while managing the potential inaccuracies. For example, when confronted with a study that exclusively reports the maximum and minimum values of estimated parameters, we calculate a conservative standard deviation as $(Max - Min)/4$. This method enables us to include the data while mitigating the impacts of imprecision. Apart from the estimates of our Experiment I, our final dataset includes 76 papers and 141 PT estimates, as listed in Subsection B.5.

Predictive certainty equivalents and associated standard error calculation With the collected data, we first calculate the predicted certainty equivalent for each observation, $\widehat{c} = u^{-1}[w(p)u(x)]$, where u designates the utility function, w the probability weighting function, and u^{-1} is the inverse of the utility function. We use $x = 100$ and $p = 0.1$, but the qualitative results do not change much for different monetary outcomes or even smaller probabilities. To obtain the standard errors of such predicted certainty equivalents, we then apply a bootstrap procedure. Specifically, we assume that each PT parameter is normally distributed around their mean estimate with variance equal to the squared standard error of the parameter encoded from the papers. We then draw 4000 samples from these parameter distributions to obtain a vector of certainty equivalents of $(100, 0.1)$ for each study. We then use this vector to obtain the standard error associated with the predicted certainty equivalent.

Some studies that do not report any standard errors for their PT parameter estimates. nor any information from which tsuch errors could be reasonably approximated. As a result, we could not apply the bootstrap procedure to these studies. Given this situation, one option might be to drop these observations from the meta-analysis. This would, however,result in the loss of a substantial number of observations, including the seminal study of Tversky and Kahneman (1992). Also, it is possible that these studies which do not report standard errors are different than those indeed reporting. This implies that dropping these incomplete observations could lead to biased conclusions. Due to these reasons, we choose to impute the standard errors of the predicted certainty equivalent for those incomplete observations.

The approach we take involves estimating the parameters characterizing their distribution in the data from the equation $\log(se_o) \sim \mathcal{N}\left(\mu_{se}, \sigma_{se}^2\right)$, where using the log ensures that we only impute positive values. Utilizing these distributional parameters, we then imputed the missing values in SE by modeling $\log(se_m) \sim \mathcal{N}\left(\widehat{\mu}_{se}, \widehat{\sigma}_{se}^2\right)$, where the subscripts o and m denote *observed* and *missing*, respectively. The parameters $\left(\widehat{\mu}_{se}, \widehat{\sigma}_{se}^2\right)$

represent the estimated quantities. In implementing this estimation, we will initially obtain values for the missing observations in standard errors (SE) that maintain the same mean and variance. However, our approach can be significantly improved by identifying variables within our dataset that are strongly associated with SEs. The most effective predictors of SEs in our data include the predicted certainty equivalent, the dummy indicator whether the experiment is conducted in a lab, the dummy indicator whether the experiment is conducted in the field, the number of subjects, and the dummy indicator of whether the adopted PWF has two parameters. We thus conduct the imputation by defining $\mu_{se} = \alpha_{se} + \beta_{se} \times \mathbf{Z}$, where $\mathbf{Z}$ represents the vector of these optimal predictors.

Meta-analysis: Bayesian hierarchical estimation. Consider the dataset $(\widehat{c}_i, se_i)$, where $\widehat{c}_i$ is the imputed certainty equivalent of the ith observation in the dataset and the associated se_i quantifies the uncertainty around it. We assume that the $\widehat{c}_i$ is normally distributed around the parameter $\widetilde{c}_i$:

$$\widehat{c}_i \sim \mathcal{N}\left(\widetilde{c}_i, se_i^2\right), \tag{2}$$

where the variability of $\widehat{c}_i$ around the true but latent mean $\widetilde{c}_i^p$ is supposed to stem from sampling variation in small studies, as captured by the known standard error se_i. This is indeed a central feature of meta-analysis or indeed of measurement error models in general – see Vieider (2024*a*) for a discussion and a tutorial.

Sampling variation contributes to the observed variability in $\widetilde{c}_i$, but it's not the only source; there may also be "genuine" heterogeneity across measurements, perhaps due to differing experimental settings, different subject pools, etc. To account for this, we assume that the study-level $\widetilde{c}_i$ follow a normal distributionacross all studies:

$$\widetilde{c}_i \sim \mathcal{N}\left(\mu, \tau^2\right), \tag{3}$$

where μ represents the meta-analytic mean of the imputed certainty equivalents, and τ represents the standard deviation of the true, latent certainty equivalents across studies. Incorporating variation across estimates due to observable characteristics, commonly known as meta-regression, can be achieved simply by defining $\mu = \boldsymbol{X\beta}$, where $\boldsymbol{X}$ is a matrix of study characteristics, including a column vector of 1s, and $\boldsymbol{\beta}$ is a vector of regression coefficients.

We estimate our models in Stan (Carpenter et al., 2017), executed from `R` (R Core Team, 2023) through `CmdStanR` (Gabry et al., 2024). Population-level parameter priors are selected to be mildly regularizing, providing informative yet broad ranges significantly larger than the expected estimates from data analysis. Lower-level parameter priors are derived from these estimated population-level parameters, ensuring a cohesive modeling framework.

B.2 Results

Meta-analysis Table B.1 presents the meta-analysis results for four groups. The absolute values of the average predictive certainty equivalent (CE) for the wager (100, 0.1) in the choice lists studies are above 0.1 for both gains and losses, with both results being significant within the 95% CrI. Conversely, the absolute values for the binary choice studies are significantly below 0.1 for both gains and losses.

Table B.1: Meta-analysis results of normalized predictive CEs

Group	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
	(100, 0.1)				(100, 0.9)			
Equivalence - Gains	0.181	0.010	0.161	0.201	0.721	0.014	0.693	0.748
Choice - Gains	0.040	0.005	0.029	0.052	0.746	0.021	0.703	0.787
Equivalence - Losses	0.213	0.022	0.171	0.256	0.764	0.014	0.737	0.791
Choice - Losses	0.075	0.011	0.054	0.097	0.784	0.032	0.713	0.842

Figure B.1 and Figure B.2 present the funnel plots of the calculated predictive CEs for gains and losses, respectively. Our primary focus here is on Figure B.1, which indicates that there is no significant publication bias favoring the four-fold pattern for both groups of studies.

Meta-regression Our meta-regression aggregates all predictive certainty equivalents (CEs) and incorporates a range of covariates. These include indicators for whether the method used was Equivalence or Choice (*choice lists*), whether the payoff domain was positive or negative (*Loss*), whether the incentive was real or hypothetical (*Hypothetical*), whether each choice included a sure amount option (*Sure*), and whether the experimental environment was a field experiment (*Field*). Additionally, the regression includes two interaction terms: one between the method dummy and payoff domain, and an-

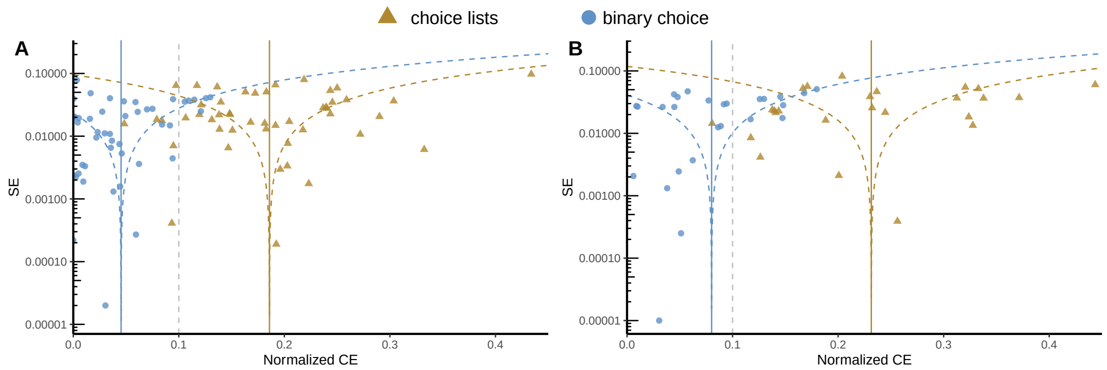


Figure B.1: Funnel plots of inferred CEs for a wager $(100, 0.1)$ and $(-100, 0.1)$
Funnel plots of calculated normalized CEs and associated standard errors. Panel A scatters CEs-SEs observations for the gain wager $(100, 0.1)$ inferred from studies using certainty equivalents or binary choices to measure risk attitudes, whereas panel B shows CEs-SEs observations for the loss wager $(-100, 0.1)$. The vertical solid lines mark the mean of normalized CEs from each condition, while the dashed gray line indicates the neutrality status (normalized CE $= p = 0.1$). Two dashed curves delineate the boundaries for a statistically significant deviation from the "true" CEs value. The y-axis is presented in log scale for improved visualization.

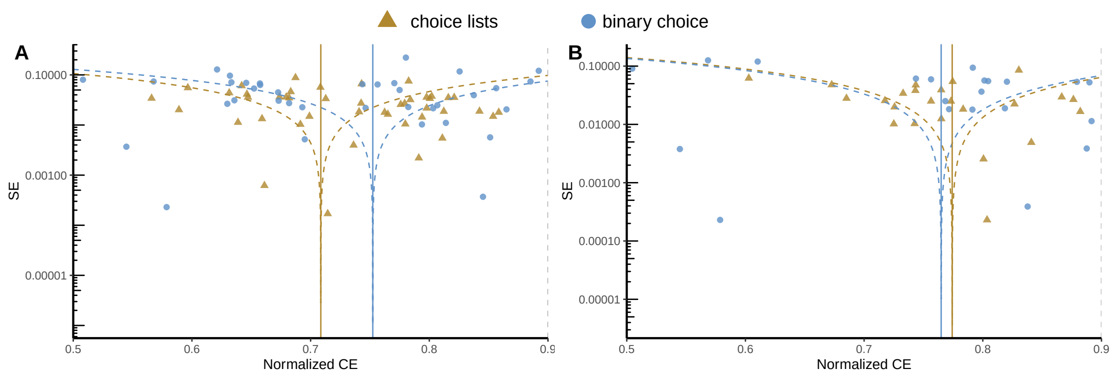


Figure B.2: Funnel plots of inferred CEs for a wager $(100, 0.9)$ and $(-100, 0.9)$
Funnel plots of calculated normalized CEs and associated standard errors. Panel A scatters CEs-SEs observations for the gain wager $(100, 0.9)$ inferred from studies using certainty equivalents or binary choices to measure risk attitudes, whereas panel B shows CEs-SEs observations for the loss wager $(-100, 0.9)$. The vertical solid lines mark the mean of normalized CEs from each condition, while the dashed gray line indicates the neutrality status (normalized CE $= p = 0.9$). Two dashed curves delineate the boundaries for a statistically significant deviation from the "true" CEs value. The y-axis is presented in log scale for improved visualization.

other between the payoff domain and incentive type. Finally, the regression incorporates other characteristics related to the stimuli, such as the largest amount and the highest probability in the lottery outcomes. For the wager (100, .1), Table B.2 indicates that the coefficient for the *Valuation* dummy variable is 10.762 and for the dummy variable *Loss* is 3.794, both showing significant positive impacts. No other variables significantly influenced the predictive CEs. The table also provides the estimation results for the predictive CE of (100, 0.9). For this predictive CE, we only see a marginally significant difference of -6.787 (90% CI: [-12.968, -0.606]) between the two elicitation methods, namely Equivalence and Choice.

Table B.2: Meta Regression of Predictive CE of (100, 0.1) and (100, 0.9)

Variable	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
	(100, 0.1)				(100, 0.9)			
Equivalence	10.762	1.882	7.075	14.448	-6.787	3.757	-14.113	0.482
Loss	3.794	1.442	0.919	6.648	3.569	2.958	-2.299	9.305
Hypothetical	-2.359	1.761	-5.788	1.062	-3.537	3.202	-9.849	2.665
Sure_Presence	3.003	1.834	-0.673	6.540	6.150	3.665	-1.037	13.277
Log(Stiml_high_Amt)	0.066	0.325	-0.565	0.708	-2.288	0.621	-3.525	-1.088
Log(Stiml_high_P)	7.750	4.522	-0.921	16.857	-1.919	8.642	-17.904	15.712
Stiml_Num_outcome	-0.839	0.729	-2.273	0.586	-1.173	1.516	-4.286	1.752
Field	1.800	2.325	-2.704	6.414	-4.354	4.758	-14.219	4.754
Equivalence*Loss	0.872	2.042	-3.084	4.885	0.618	3.536	-6.280	7.704
Loss*Hypothetical	-3.743	2.446	-8.524	1.133	1.696	4.042	-6.387	9.820
Cons	6.800	2.223	2.441	11.132	86.679	4.672	77.784	96.182

Note: This table presents a meta-regression of predictive certainty equivalents (CEs), incorporating the following variables: *Equivalence*: Indicator variable (1 = valuation method, 0 = choice method) used to distinguish whether certainty equivalents or binary choices were applied. *Loss*: Indicator variable (1 = negative payoff domain, 0 = positive payoff domain), representing whether the study dealt with losses or gains. *Hypothetical*: Indicator variable (1 = hypothetical incentives, 0 = real incentives), denoting whether monetary incentives were hypothetical or real. *Sure*: Indicator variable (1 = sure option included, 0 = no sure option), capturing whether a sure amount option was provided in the choice set. *Log(Stimul_high_Amt)*: The natural logarithm of the largest monetary amount presented in the stimuli (lottery). *Log(Stimul_high_P)*: The natural logarithm of the highest probability used in the lottery outcomes. *Stimul_Num_outcome*: The number of outcomes in the stimuli (lottery). *Field*: Indicator variable (1 = field experiment, 0 = laboratory experiment), distinguishing between field and laboratory settings. *choice lists*Loss*: Interaction term between valuation method and the payoff domain (loss or gain). *Loss*Hypothetical*: Interaction term between the payoff domain (loss or gain) and the incentive type (hypothetical vs. real).

B.3 Robustness

B.3.1 Different standard error imputation methods

To verify the robustness of our meta-analysis results concerning the calculation of predictive certainty equivalents' standard error (SE), this subsection introduces two alternative imputation methods to get standard errors for those incomplete observations. Specifically, the first method uses Bruhin, Fehr-Duda and Epper (2010) as the benchmark study and calculate the SD for other studies as follows:

$$SE_i = SE^* \cdot \frac{\sqrt{N^*}}{\sqrt{N_i}}, \tag{4}$$

where SE^* and N^* represent the bootstrapped standard error and the number of subjects from the "*Zurich-03*" study in Bruhin, Fehr-Duda and Epper (2010), and N_i is the number of subjects in study i. Table B.3 reports the resulting estimates with using the newly derived standard errors for the wager (100, 0.1) and (100, 0.9), which are very close to those we report in the main text.

Table B.3: Meta-analysis results with the first different standard error calculation method

Group	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
	(100, 0.1)				(100, 0.9)			
Equivalence - Gains	0.182	0.010	0.162	0.201	0.720	0.013	0.694	0.746
Choice - Gains	0.045	0.006	0.034	0.056	0.751	0.023	0.706	0.796
Equivalence - Losses	0.227	0.020	0.189	0.266	0.772	0.015	0.742	0.801
Choice - Losses	0.078	0.010	0.059	0.098	0.768	0.035	0.699	0.835

Table B.4: Meta-analysis results with the second different standard error calculation method

Group	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
	(100, 0.1)				(100, 0.9)			
Equivalence - Gains	0.183	0.010	0.162	0.203	0.721	0.014	0.693	0.747
Choice - Gains	0.046	0.006	0.034	0.058	0.752	0.023	0.707	0.797
Equivalence - Losses	0.228	0.019	0.190	0.266	0.774	0.014	0.746	0.802
Choice - Losses	0.079	0.011	0.058	0.099	0.769	0.035	0.700	0.837

The second SE calculation method instead takes the average standard deviation of all studies from a same group (binary choice or choice lists), denoted as $SD_{average}$. Then, for every single incomplete observation, we can derive its predictive CE's standard error as below:

$$SE_i = \frac{SD_{average}}{\sqrt{N_i}}, \tag{5}$$

where N_i is the number of subjects for study i. Table B.4 reports the results after adopting these differently derived standard errors for the wager (100, 0.1) and (100, 0.9). Again, these results are essentially similar to those reported in the main text, and thus showing that our findings of the meta analysis are robust to different standard error calculation approaches.

B.3.2 Different Wagers

To demonstrate the robustness of our conclusion with respect to the wager employed for the predictive CE calculation, we have re-estimated using an alternative wager (200, 0.05) and (20, 0.05) respectively. Tables B.5 presents these estimates, which are largely consistent with our primary conclusions detailed in the main text, namely the absence of the four-fold pattern in the binary choice condition.

Table B.5: Meta-analysis Results for the inferred CE of different wager

Group	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
	(200, 0.05)				(20, 0.05)			
Equivalence - Gains	0.131	0.009	0.113	0.149	0.134	0.009	0.116	0.151
Choice - Gains	0.014	0.008	0.000	0.025	0.019	0.003	0.013	0.026
Equivalence - Losses	0.158	0.021	0.118	0.198	0.161	0.019	0.124	0.198
Choice - Losses	0.038	0.007	0.025	0.053	0.039	0.007	0.026	0.053

B.4 Structural Estimation of PT parameters

Methods and code We recover PT parameters from aggregate Bayesian estimations using Stan. In our preferred specification reported in the main text, we use a power utility function, $u(x) = x^{\rho}$. Following the majority of the papers in our meta-analysis, we estimate a 1-parameter specification of the probability weighting function proposed by Prelec (1998), which takes the form $w(p) = exp(-(-ln(p))^{\gamma})$, where values of $\gamma < 1$ capture likelihood insensitivity. Estimation results using alternative 1-parameter functions, such as the one of Tversky and Kahneman, produce very similar results. Estimations obtained using 2-parameter functions are discussed farther below.

We estimate the functionals using purpose-coded models in Stan, which we launch from R. Vieider (2024*a*) provides a detailed tutorial on estimations of structural models in Stan. We use the following R code:

```
sr <- cmdstan_model("stancode/agg_PT_pp1.stan")

fit_save_agg_pp1 <- function(file, cl_value, suffix, is_loss = FALSE) {
  d <- read.csv(file) %>%
    filter(CL == cl_value)

  stanD <- list(
    N = nrow(d),
    high = d$high,
    p = d$p,
    low = d$low,
    sure = d$sure,
    choice_risky = if (is_loss) 1 - d$choice_risky else d$choice_risky
  )

  nm <- sr$sample(
```

```
    data = stanD,
    seed = 123,
    chains = 4,
    parallel_chains = 4,
    refresh = 200,
    init = 0,
    show_messages = TRUE,
    diagnostics = c("divergences", "treedepth", "ebfmi")
  )

  nm$save_object(file = paste0("stanoutput/agg_e23_pp1_", suffix, ".Rds"))

  invisible(nm)
}

fits <- list(
  cg = fit_save_agg_pp1("data/d23_gains.csv",  cl_value = 1, suffix = "cg"),
  bg = fit_save_agg_pp1("data/d23_gains.csv",  cl_value = 0, suffix = "bg"),
  cl = fit_save_agg_pp1("data/d23_losses.csv", cl_value = 1, suffix = "cl", is_loss = TRUE
  bl = fit_save_agg_pp1("data/d23_losses.csv", cl_value = 0, suffix = "bl", is_loss = TRUE
)
```

and Stan code (the baisc Prelec I specification):

```
data {
  int<lower=1> N;
  array[N] real high;
  array[N] real low;
  array[N] real sure;
  array[N] real p;
  array[N] int choice_risky;
}
parameters {
  real rho;
  real<lower=0> gamma;
  real<lower=0> sigma;
}
model {
  vector[N] pw;
  vector[N] pv;
  vector[N] udiff;
```

```
  rho ~ normal(1 , 0.5);
  gamma ~ normal(1 , 0.5);
  sigma ~ normal(0 , 0.5);

  for (i in 1:N) {
    pw[i] = exp(-(-log(p[i]))^gamma);
    pv[i] = pw[i] * pow(high[i], rho) + (1 - pw[i]) * pow(low[i], rho);
    udiff[i] = (pv[i] - pow(sure[i], rho)) / (sigma * (high[i] - low[i]));
  }

  choice_risky ~ bernoulli_logit(udiff);
}
```

Estimation results Table B.6 reports PT parameters' estimation results for Experiment I, and the four groups respectively: binary choice/choice lists * Gains/Losses.

Table B.6: Prospect theory parameters estimation results - Experiment I

Group	ρ	SD_ρ	γ	SD_γ	δ	SD_δ	σ	SD_σ
			$Power - Prelec\,I$					
Equivalence - Gains	0.965	0.007	0.542	0.007			0.118	0.003
Choice - Gains	0.574	0.005	0.696	0.009			0.037	0.001
Equivalence - Losses	0.945	0.011	0.804	0.014			0.129	0.004
Choice - Losses	0.794	0.009	1.050	0.015			0.073	0.002

B.5 Collected Studies for Meta-analysis

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2018. "Multiattribute Utility Theory, Intertemporal Utility, And Correlation Aversion." *International Economic Review*, 59: 537–555.

Andersen, Steffen, John Fountain, Glenn W. Harrison, and E. Elisabet Rutström. 2014. "Estimating subjective probabilities." *Journal of Risk and Uncertainty*, 48: 207–229.

Andreoni, James, and Charles Sprenger. 2011. "Uncertainty Equivalents: Testing the Limits of the Independence Axiom." Working paper.

Baillon, Aurélien, and Lætitia Placido. 2019. "Testing constant absolute and relative ambiguity aversion." *Journal of Economic Theory*, 181: 309–332.

Baillon, Aurélien, Han Bleichrodt, and Vitalie Spinu. 2020. "Searching for the reference point." *Management Science*, 66: 93–112.

Baláž, Vladimír, Viera Bacova, Eva Drobna, Katarina Dudekova, and Kamil Adamik. 2013. "Testing Prospect Theory Parameters." *Ekonomický časopis*, 61: 655–671.

Beauchamp, Jonathan P., Daniel J. Benjamin, David I. Laibson, and Christopher F. Chabris. 2020. "Measuring and controlling for the compromise effect when estimating risk preference parameters." *Experimental Economics*, 23: 1069–1099.

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica*, 88: 1363–1409.

Birnbaum, Michael H, and Alfredo Chavez. 1997. "Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence." *Organizational Behavior and Human Decision Processes*, 71: 161–194.

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

Bougherara, Douadia, Lana Friesen, and Céline Nauges. 2021. "Risk Taking with Left- and Right-Skewed Lotteries*." *Journal of Risk and Uncertainty*, 62: 89–112.

Brandstätter, Eduard, Anton Kühberger, and Friedrich Schneider. 2002. "A Cognitive-Emotional Account of the Shape of the Probability Weighting Function." *Journal of Behavioral Decision Making*, 15: 79–100.

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

Bruhin, Adrian, Luís Santos-Pinto, and David Staubli. 2018. "How do beliefs about skill affect risky decisions?" *Journal of Economic Behavior and Organization*, 150: 350–371.

Carpio, María Bernedo Del, Francisco Alpizar, and Paul J. Ferraro. 2022. "Time and risk preferences of individuals, married couples and unrelated pairs." *Journal of Behavioral and Experimental Economics*, 97: 101794.

Charupat, Narat, Richard Deaves, Travis Derouin, Marcelo Klotzle, and Peter Miu. 2013. "Emotional balance and probability weighting." *Theory and Decision*, 75: 17–41.

Choi, Syngjoo, Jeongbin Kim, Eungik Lee, and Jungmin Lee. 2022. "Probability Weighting and Cognitive Ability." *Management Science*, 68: 5201–5215.

Conte, Anna, John D. Hey, and Peter G. Moffatt. 2011. "Mixture models of choice under risk." *Journal of Econometrics*, 162: 79–88.

Conte, Anna, M. Vittoria Levati, and Chiara Nardi. 2018. "Risk Preferences and the Role of Emotions." *Economica*, 85: 305–328.

Coricelli, Giorgio, Enrico Diecidue, and Francesco D. Zaffuto. 2018. "Evidence for multiple strategies in choice under risk." *Journal of Risk and Uncertainty*, 56: 193–210.

Enke, Benjamin, and Cassidy Shubatt. 2023. "Quantifying Lottery Choice Complexity." Working paper.

Epper, Thomas, Helga Fehr-Duda, and Adrian Bruhin. 2011. "Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting." *Journal of Risk and Uncertainty*, 43: 169–203.

Fan, Yuyu, David V. Budescu, and Enrico Diecidue. 2019. "Decisions with compound lotteries." *Decision*, 6: 109–133.

Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert. 2010. "Rationality on the rise: Why relative risk aversion increases with stake size." *Journal of Risk and Uncertainty*, 40: 147–180.

Fehr-Duda, Helga, Manuele De Gennaro, and Renate Schubert. 2006. "Gender, financial risk, and probability weights." *Theory and Decision*, 60: 283–313.

Fehr-Duda, Helga, Thomas Epper, Adrian Bruhin, and Renate Schubert. 2011. "Risk and rationality: The effects of mood and decision rules on probability weighting." *Journal of Economic Behavior and Organization*, 78: 14–24.

Gao, Xiaoxue Sherry, Glenn W. Harrison, and Rusty Tchernis. 2023. "Behavioral welfare economics and risk preferences: a Bayesian approach." *Experimental Economics*, 26: 273–303.

Glatt, Markus, Roy Brouwer, and Ivana Logar. 2019. "Combining Risk Attitudes in a Lottery Game and Flood Risk Protection Decisions in a Discrete Choice Experiment." *Environmental and Resource Economics*, 74: 1533–1562.

Glöckner, Andreas, and Thorsten Pachur. 2012. "Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory." *Cognition*, 123: 21–32.

Glöckner, Andreas, Baiba Renerte, and Ulrich Schmidt. 2020. "Violations of coalescing in parametric utility measurement." *Theory and Decision*, 89: 471–501.

Glöckner, Andreas, Benjamin E. Hilbig, Felix Henninger, and Susann Fiedler. 2016. "The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice." *Journal of Experimental Psychology: General*, 145: 486–508.

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

Gonzalez, Richard, and George Wu. 2022. "Composition rules in original and cumulative prospect theory." *Theory and Decision*, 92: 647–675.

Haffke, Peter, and Ronald Hübner. 2014. "Effects of different feedback types on information integration in repeated monetary gambles." *Frontiers in Psychology*, 5.

Hajimoladarvish, Narges. 2018. "How do people reduce compound lotteries?" *Journal of Behavioral and Experimental Economics*, 75: 126–133.

Hajimoladarvish, Narges. 2021. "Explaining Heterogeneity in Risk Preferences Using a Finite Mixture Model." *Journal of Money and Economy*, 16: 533–554.

Harbaugh, William T, Kate Krause, and Lise Vesterlund. 2002. "Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses." *Experimental Economics*, 5: 53–84.

Harrison, Glenn W., and J. Todd Swarthout. 2019. "Eye-tracking and economic theories of choice under risk." *Journal of the Economic Science Association*, 5: 26–37.

Harrison, Glenn W., Andre Hofmeyr, Don Ross, and J. Todd Swarthout. 2018. "Risk Preferences, Time Preferences, and Smoking Behavior." *Southern Economic Journal*, 85: 313–348.

Harrison, Glenn W., Morten I. Lau, and Hong Il Yoo. 2020. "Risk attitudes, sample selection, and attrition in a longitudinal field experiment." *Review of Economics and Statistics*, 102: 552–568.

Hey, John D., Andrea Morone, and Ulrich Schmidt. 2009. "Noise and bias in eliciting preferences." *Journal of Risk and Uncertainty*, 39: 213–235.

Hsu, Ming, Ian Krajbich, Chen Zhao, and Colin F. Camerer. 2009. "Neural response to reward anticipation under risk is nonlinear in probabilities." *Journal of Neuroscience*, 29: 2231–2237.

Kellen, David, Markus D. Steiner, Clintin P. Davis-Stober, and Nicholas R. Pappas. 2020. "Modeling choice paradoxes under risk: From prospect theories to sampling-based accounts." *Cognitive Psychology*, 118.

Kellen, David, Thorsten Pachur, and Ralph Hertwig. 2016. "How (in)variant are subjective representations of described and experienced risk and rewards?" *Cognition*, 157: 126–138.

Kusev, Petko, Paul van Schaik, Peter Ayton, John Dent, and Nick Chater. 2009. "Exaggerated Risk: Prospect Theory and Probability Weighting in Risky Choice." *Journal of Experimental Psychology: Learning Memory and Cognition*, 35: 1487–1505.

Lejarraga, Tomás, Thorsten Pachur, Renato Frey, and Ralph Hertwig. 2016. "Decisions from Experience: From Monetary to Medical Gambles." *Journal of Behavioral Decision Making*, 29: 67–77.

l'Haridon, Olivier, and Ferdinand M. Vieider. 2019. "All over the map: A world-wide comparison of risk preferences." *Quantitative Economics*, 10: 185–215.

Maafi, Hela. 2011. "Preference reversals under ambiguity." *Management Science*, 57: 2054–2066.

Murad, Zahra, Martin Sefton, and Chris Starmer. 2016. "How do risk attitudes affect measured confidence?" *Journal of Risk and Uncertainty*, 52: 21–46.

Murphy, Ryan O., and Robert H.W.Ten Brincke. 2018. "Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates." *Management Science*, 64: 308–326.

Newell, Anthony. 2020. "Is your heart weighing down your prospects? Interoception, risk literacy and prospect theory." Working paper.

O'Brien, Megan K., and Alaa A. Ahmed. 2014. "Take a stand on your decisions, or take a sit: Posture does not affect risk preferences in an economic task." *PeerJ*, 2014.

Pachur, Thorsten, Michael Schulte-Mecklenbeck, Ryan O. Murphy, and Ralph Hertwig. 2018. "Prospect theory reflects selective allocation of attention." *Journal of Experimental Psychology: General*, 147: 147–169.

Pachur, Thorsten, Yaniv Hanoch, and Michaela Gummerum. 2010. "Prospects behind bars: Analyzing decisions under risk in a prison population." *Psychonomic Bulletin and Review*, 17: 630–636.

Panidi, Ksenia, Alicia Nunez Vorobiova, Matteo Feurra, and Vasily Klucharev. 2022. "Dorsolateral prefrontal cortex plays causal role in probability weighting during risky choice." *Scientific Reports*, 12: 16115.

Patalano, Andrea L., Alexandra Zax, Katherine Williams, Liana Mathias, Sara Cordes, and Hilary Barth. 2020. "Intuitive symbolic magnitude judgments and decision making under risk in adults." *Cognitive Psychology*, 118.

Patalano, Andrea L., Jason R. Saltiel, Laura Machlin, and Hilary Barth. 2015. "The role of numeracy and approximate number system acuity in predicting value and probability distortion." *Psychonomic Bulletin and Review*, 22: 1820–1829.

Ring, Patrick, Catharina C. Probst, Levent Neyse, Stephan Wolff, Christian Kaernbach, Thilo van Eimeren, Colin F. Camerer, and Ulrich Schmidt. 2018. "It's all about gains: Risk preferences in problem gambling." *Journal of Experimental Psychology: General*, 147: 1241–1255.

Scheibehenne, Benjamin, and Thorsten Pachur. 2015. "Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice." *Psychonomic Bulletin and Review*, 22: 391–407.

Schulreich, Stefan, Yana G. Heussen, Holger Gerhardt, Peter N.C. Mohr, Ferdinand C. Binkofski, Stefan Koelsch, and Hauke R. Heekeren. 2014. "Music-evoked incidental happiness modulates probability weighting during risky lottery choices." *Frontiers in Psychology*, 4.

Spitmaan, Mehran, Emily Chu, and Alireza Soltani. 2019. "Salience-driven value construction for adaptive choice under risk." *Journal of Neuroscience*, 39: 5195–5209.

Stott, Henry P. 2006. "Cumulative prospect theory's functional menagerie." *Journal of Risk and Uncertainty*, 32: 101–130.

Suter, Renata S., Thorsten Pachur, and Ralph Hertwig. 2016. "How Affect Shapes Risky Choice: Distorted Probability Weighting Versus Probability Neglect." *Journal of Behavioral Decision Making*, 29: 437–449.

Suter, Renata S., Thorsten Pachur, Ralph Hertwig, Tor Endestad, and Guido Biele. 2015. "The neural basis of risky choice with affective outcomes." *PLoS ONE*, 10.

Takemura, Kazuhisa, and Hajime Murakami. 2016. "Probability Weighting Functions Derived from Hyperbolic Time Discounting: Psychophysical Models and Their Individual Level Testing." *Frontiers in Psychology*, 7.

Tsang, Ming. 2020. "Estimating uncertainty aversion using the source method in stylized tasks with varying degrees of uncertainty." *Journal of Behavioral and Experimental Economics*, 84.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty*, 5(4): 297–323.

Vieider, Ferdinand M., Abebe Beyene, Randall Bluffstone, Sahan Dissanayake, Zenebe Gebreegziabher, Peter Martinsson, and Alemu Mekonnen. 2018. “Measuring Risk Preferences in Rural Ethiopia.” *Economic Development and Cultural Change*, 66: 417–446.

Vieider, Ferdinand M., Clara Villegas-Palacio, Peter Martinsson, and Milagros Mejía. 2016. “Risk Taking For Oneself And Others: A Structural Model Approach.” *Economic Inquiry*, 54: 879–894.

Vieider, Ferdinand M., Peter Martinsson, Pham Khanh Nam, and Nghi Truong. 2019. “Risk preferences and development revisited.” *Theory and Decision*, 86: 1–21.

Vieider, Ferdinand M., Thorsten Chmura, Tyler Fisher, Takao Kusakawa, Peter Martinsson, Frauke Mattison Thompson, and Adewara Sunday. 2015. “Within- versus between-country differences in risk attitudes: implications for cultural comparisons.” *Theory and Decision*, 78: 209–218.

Wu, George, and Richard Gonzalez. 1996. “Curvature of the Probability Weighting Function.” *Management Science*, 42: 1676–1690.

Wu, J. George, and Alex B. Markle. 2008. “An empirical test of gain-loss separability in prospect theory.” *Management Science*, 54: 1322–1335.

Wölbert, Eva, and Arno Riedl. 2013. “Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity.” Working paper.

Zhou, Wenting, and John Hey. 2018. “Context matters.” *Experimental Economics*, 21: 723–756.

C A noisy coding model of binary choice versus choice lists

Here, we formalize the intuition presented in the main text by sketching a stylized noisy coding model. However, we stress that the intuition described in the main text is considerably more general than many of the specific modelling choices and assumptions we make here. Indeed, any noisy coding model in which sequential evaluation results in an accumulation of coding noise will deliver key insights like those we derive here. For instance, Khaw, Li and Woodford (2023) present a closely related model of valuation built on somewhat different detail-level modeling choices, but which shares some of the key predictions we derive here.

Lottery evaluation

We discuss choices between a lottery $(x, p; y)$ and a sure outcome c, following the setup in all our experiments. We start by describing the evaluation of the lottery, i.e. of the log-odds in favour of winning a prize. To introduce noisy cognition, we assume that the log-odds are not directly accessible to the DM but are noisily coded, i.e. they are mentally represented by a signal r_p. On average, this signal will be unbiased, i.e. it will reflect the true log-odds shown to the decision maker. In any specific instance, however, the signal may be affected by some noise, which we assume to be normally distributed. We thus obtain the following likelihood function encoding a given probability p:

$$r_p \sim \mathcal{N}\left(ln\left(\frac{p}{1-p}\right), \nu_p^2\right),$$

where ν_p^2 is the variance of the coding errors.

To decode the signal – to rein in the errors incurred in the coding process – the signal is decoded by combination with a learned prior, capturing the statistics of the environment. Note that this prior does not necessarily correctly reflect those statistics, either because few stimuli have been encountered as of yet (as is the case at the beginning of an experiment, where the full stimulus distribution will only be known once the experiment is over), or because of the influence of a hyperprior summarizing experiences across different environments. Another possibility is that positive versus negative experiences may have asymmetric impacts on learning (see Vieider, 2023, for a formal model of learning in the present context, and a discussion of why learning will necessarily be affected by noise). In particular, if the prior mean is "too pessimistic" relative to the stimuli presented, then this will result in risk averse choices.

Decoding the log-odds signal by a conjugate normal prior, which takes the form $\mathcal{N}(ln(\eta), \sigma_p^2)$, will yield the following posterior distribution:

$$ln\frac{p}{1-p}\,|\,r_p \sim \mathcal{N}\left(\gamma\, r_p + (1-\gamma)\, ln(\eta)\,,\, \frac{\nu^2\sigma_p^2}{\nu^2+\sigma_p^2}\right). \tag{6}$$

For simplicity, we will henceforth normalize the prior SD to 1 by dividing all parameters by σ_p^2. The upshot is that we reduce the system by one parameter without losing any information (see Natenzon, 2019, for an analogous simplification). The coding noise becomes $\widehat{\nu}_p = \frac{\nu_p}{\sigma_p}$, thus capturing the coding noise relative to the prior variance of the log-odds. The Bayesian evidence weight is then defined as $\gamma \triangleq \frac{1}{\widehat{\nu}_p^2+1} = \frac{1}{\nu_p^2/\sigma_p^2+1}$.

To make this expression observable to the econometrician, we further condition the posterior mean on many repetitions of the same stimulus p. The variation of responses across repeated presentations of the same stimulus yields the following observable expression, referred to as the *response distribution* (Ma, Kording and Goldreich, 2023):

$$\mathbb{E}\left[\gamma\, r_p + (1-\gamma)\, ln(\eta)\,\middle|\, p\right] \;\sim\; \mathcal{N}\left(\gamma\, ln\left(\frac{p}{1-p}\right) + (1-\gamma)\, ln(\eta)\,,\, \gamma^2\nu_p^2\right). \tag{7}$$

Proof. Let $z \sim \mathcal{N}(\widehat{z}\,,\, \tau^2)$. From the well-known properties of the normal distributions it follows that $bz + a \sim \mathcal{N}(b\widehat{z} + a\,,\, b^2\tau^2)$. The result above obtains by letting $b = \gamma$, $z = r_p$, $a = (1-\gamma)\, ln(\eta)$, $\widehat{z} = ln\left(\frac{p}{1-p}\right)$, and $\tau = \nu_p$. □

The average inference on the log-odds displayed above is now systematically shaded or shrunk towards the mean of the prior. Intuitively, this happens because the signal is only taken into account in proportion to the 'confidence' the DM has in the signal, as captured by its precision (the inverse of the coding noise variance, $\widehat{\nu}^{-2}$, since we can alternatively define $\gamma = \frac{\widehat{\nu}^{-2}}{1+\widehat{\nu}^{-2}}$). This results in *systematic bias* in the evaluation of the log-odds, which as we will see shortly is at the origin of probability distortions (see Oprea and Vieider, 2024, for an experimental test of this mechanism). Indeed, substituting $\delta \triangleq \eta^{1-\gamma}$ into the expectation of the normal distribution above yields a linear in log-odds probability weighting function that is frequently used in the prospect theory literature (Gonzalez and Wu, 1999; Bruhin, Fehr-Duda and Epper, 2010).

C.1 Binary binary choice

In binary binary choice, the posterior for the log-odds derived above is simply compared to the posterior for the comparative outcomes, given by the log cost-benefits, $ln\left(\frac{c-y}{x-c}\right)$ (this derives from a choice rule that maximizes expected value – see Vieider, 2024*b*, for a detailed discussion). Importantly, we assume that the log-odds of the lottery and the log-cost benefits are evaluated simultaneously and independently. This seems indeed natural, inasmuch as there is no reason in binary binary choice why the evaluation of one dimension should be conditioned on the other. Given this independence in evaluations, the signal for the log cost-benefits will once again be unbiased, having as its mean the true log cost-benefits, but potentially deviating from them in any single draw from the following distribution:

$$r_o \sim \mathcal{N}\left(ln\left(\frac{c-y}{x-c}\right), \nu_o^2\right),$$

where ν_o is the standard deviation of outcome coding noise.

The evaluation of the log cost-benefits then proceeds much like for the log-odds above (see Vieider, 2024, for details). For simplicity, we assume that costs and benefits are expected to be equal in the prior, so that the mean of the log cost-benefits is 0. This seems a natural assumption, and it is made without loss of generality, as the results presented below will generalize to setups with a non-degenerate prior for outcomes. The response distribution will now take the following form:

$$\mathbb{E}\left[\alpha\, r_o \,|\, c, y, x\right] \sim \mathcal{N}\left(\alpha\, ln\left(\frac{c-y}{x-c}\right), \alpha^2 \widehat{\nu}_o^2\right), \tag{8}$$

where similarly to above $\widehat{\nu}_o = \frac{\nu_o}{\sigma_o^2}$ is the normalized outcome coding noise, and $\alpha \triangleq \frac{1}{\widehat{\nu}_o^2+1}$ is the outcome discriminability parameter.

Following an EV maximization rule that is often employed in signal detection theory (Green, Swets et al., 1966; Gold and Shadlen, 2001), we assume that the log-odds are traded off against the log cost-benefits to reach a decision. Given that the objective quantities are not available to the decision maker, she decides instead based on her posterior inferences on those quantities. On average, this will result in the following stochastic choice rule observable to the econometrician:

$$Pr[(x,p;y) \succ c] = \Phi \left[\frac{\alpha^{-1} \left[\gamma\, ln\left(\frac{p}{1-p}\right) + (1-\gamma)\, ln(\eta) \right] - ln\left(\frac{c-y}{x-c}\right)}{\sqrt{\alpha^{-2}\gamma^2 \widehat{\nu}_p^2 + \widehat{\nu}_o^2}} \right]. \tag{9}$$

Proof. The proof proceeds as in Vieider (2024). Re-arrange the threshold equation in (10) by multiplying both sides by α^{-1}. The rest of the proof proceeds as usual. □

Given that $\alpha^{-1} > 1$ for any $\nu_o > 0$, noise in outcome assessments counteracts noise in probability assessments. Likelihood insensitivity is driven by $\gamma/\alpha < 1$, and will thus appear whenever probability coding noise exceeds outcome coding noise, $\widehat{\nu}_p > \widehat{\nu}_o$, resulting in $\gamma < \alpha$. A second implication is that — as long as $\eta < 1$, indicating a "risk-averse prior" as typically found in experiments (Vieider, 2024; Oprea and Vieider, 2024) — any $\alpha < 1$ (any noise in outcome perceptions) will reduce the upweighting of $\eta < 1$ towards 1 (towards 0 for $ln(\eta)$) produced by the power $1-\gamma$. An interesting special case occurs when coding noise is exactly equal for probabilities and outcomes: In this case, there will be no probability distortions, and decisions will be an expression of the 'true' prior mean $ln(\eta)$. Interestingly, this case can occur even in the presence of considerable coding noise, as long as the level of that noise is the same for probabilities and outcomes.

C.2 Choice lists

In choice list tasks the decision situation is quite different. DMs are confronted with a list of varying sure outcomes, which is compared to an unchanging lottery, which is prominently displayed. This makes it natural to assume that the lottery is evaluated first, and that the point of indifference is found in a second stage conditional on the evaluation of the lottery. The first stage will then result in an evaluation of the log-odds just as described above. It is in the second-stage evaluation — which now consists in finding an indifference value — that the differences with binary choice situations will emerge.

The second stage evaluation now consists in finding the value of the sure amount that equalizes the log cost-benefits with the posterior inference on the log-odds, call it c^*. This fundamentally changes the decision problem, since the DM now no longer tries to infer the true log-cost benefits from an unbiased signal as in binary choice, but rather tries to identify the (noisy signal for) the sure amount that produces indifference between the posterior log-odds and the signal for the log-cost benefits. As we will see shortly, this implies that the outcome signal is now no longer unbiased as in binary choice, but is itself affected by systematic bias, since it centered on the posterior for the log-odds, which is by definition a biased quantity in the presence of coding noise.

Technically, the choice lists task now takes the form of a search for the noisy outcome signal (out of the vector of signals $\boldsymbol{r_o}$ in the list) that minimizes the absolute difference between the log cost-benefit signal and the posterior of the log odds, i.e.

$$r_o^* \,|\, c^* = \arg\min_{\boldsymbol{r_o}|\boldsymbol{c}} \left| \boldsymbol{r_o} - [\gamma\, r_p + (1-\gamma)\, ln(\eta)\,]\, \right|. \tag{10}$$

This minimization problem will result in the selection of r_o^* such that the average difference with the posterior mean of the lottery evaluation is 0. At the point of indifference, we will thus observe the following relation on average:

$$r_o^* - [\gamma\, r_p + (1-\gamma)\, ln(\eta)] \sim \mathcal{N}\left(0\,,\, \nu_o^2 + \frac{\widehat{\nu}_p^2}{\widehat{\nu}_p^2 + 1}\right), \tag{11}$$

where the variance is the sum of the coding noise variance of the outcome signal and the variance of the posterior distribution of the log-odds in (6). It follows that

$$r_o^* \sim \mathcal{N}\left(\gamma\, r_p + (1-\gamma)\, ln(\eta)\,,\, \nu_o^2 + \frac{\widehat{\nu}_p^2}{\widehat{\nu}_p^2 + 1}\right). \tag{12}$$

An important insight results from this equation. Instead of an outcome signal that is an unbiased estimator of the true log cost-benefits, as in choice, the DM now chooses a signal that is centered on the posterior inference of the log-odds, which is a systematically biased quantity. This, in turn, will result in the accumulation of probability coding noise with outcome coding noise in the process of finding an indifference value.

To see this technically, we start again by deriving the posterior distribution. To simplify notation, we now define $\widetilde{\nu}_o^2 \triangleq \nu_o^2 + \frac{\nu_p^2}{\nu_p^2+1}$. Conditional on the noisy signal r_o^*, the posterior distribution takes the following form:

$$ln\left(\frac{c^* - y}{x - c^*}\right) \,|\, r_o^* \sim \mathcal{N}\left(\beta\, r_o^*\,,\, \frac{\widetilde{\nu}_o^2}{\widetilde{\nu}_o^2 + 1}\right), \tag{13}$$

where $\beta \triangleq \frac{1}{\widetilde{\nu}_o^2+1}$. This equation assumes again implicitly (and without loss of generality) that costs and benefits are equal in the prior on average, so that the prior mean drops out of the equation (or equivalently, that the cost-benefit prior has mean 0).

To make the equation above observable to the experimenter, we can reformulate it in terms of the *response distribution*, i.e. the distribution conditional on repeated presentations of the same choice stimulus (given that r_p and r_o^* are stochastic due to coding noise). To achieve this, we condition on two quantities: 1) the expectation of the probability response distribution in (7), which we omit from the notation below to avoid clutter; and 2) on the vector of sure amounts:

$$\mathbb{E}\left[\mathbb{E}\left(ln\left(\frac{c^*-y}{x-c^*}\right)\,|\,r_o^*\right)|\,\boldsymbol{c},x,y\right] \sim \mathcal{N}\left(\beta\left[\gamma\, ln\left(\frac{p}{1-p}\right)+(1-\gamma)ln\,(\eta)\right],\,\widetilde{\nu}_o^2+\beta^2\gamma^2\widehat{\nu}_p^2\right). \tag{14}$$

This results in an empirically estimable equation describing valuations for lotteries (here described as the density around the indifference value, as typically modelled when estimating valuation data, see e.g. Gonzalez and Wu, 1999; Bruhin, Fehr-Duda and Epper, 2010; L'Haridon and Vieider, 2019).[24]

Proof. From (12), $\mathbb{E}\left[r_o^*\,|\,\mathbb{E}\left[ln\left(\frac{p}{1-p}\right)\,|\,r_p\right]\right] = \gamma\, r_p + (1-\gamma)\,ln(\eta)$. We next make this observable step by step. We first condition on the probability response distribution in (7) to obtain $\mathbb{E}\left[r_o^*\,|\,\mathbb{E}\left[ln\left(\frac{p}{1-p}\right)\,|\,p\right]\right] = \gamma\, ln\left(\frac{p}{1-p}\right)+(1-\gamma)\,ln(\eta)$, with notation in (14) simplified by directly conditioning on p. This brings the variance to $\widetilde{\nu}_o^2+\gamma^2\widehat{\nu}_p^2$. Finally, we exploit the posterior distribution of the log cost-benefits in (13) and condition on the vector of sure amounts $\boldsymbol{c}$ to obtain the expectation in (14). The proof for the variance proceeds as above and is not repeated here. □

Equation (14) produces the key elements distinguishing choice lists from binary choice. The outcome discriminability or 'shrinkage' weight β is now applied to the posterior of the log-odds (instead of to the true log cost-benefits, as in binary choice), implying that

[24]Given the binary choice nature of our choice list tasks, the valuation in (14) may further be noisily implemented — a step that we do not explicitly model here. If choices are noisily implementation row-by-row, this will have two key consequences: 1) To the extent that the assessment of the costs and benefits offered by each row of the choice list is itself noisy, the effect of the multiplier β to the average probabilistic inference may partially cancel out; 2) the additional noise arising in the row-wise cost-benefit assessments will result in a further increase of response noise. The key characteristics of the process, however, remain those modeled in equation 14.

1. the discriminability weight attributed to the true log-odds, γ, is now attenuated in the presence of outcome coding noise, since the log-odds now receive a weight γ, which in the presence of outcome coding noise is smaller than the weight γ/α in binary choice – i.e., likelihood insensitivity will the more pronounced in choice lists than in binary choice;

2. The weight attributed to the prior is now also attenuated, since it is given by $(1 - \gamma)$, which is smaller than $(1-\gamma)/\alpha$ in binary choice in the presence of outcome coding noise (i.e. for $\alpha < 1$). Any value η will thus be "compressed" towards 1 (any prior mean $ln(\eta)$ will be compressed towards 0), which in the presence of risk aversion in the prior (captured by $\eta < 1$) implies an increase in risk taking under choice lists relative to binary choice (a *decrease* in risk taking in the presence of an optimistic prior for losses);

3. Given that $\widetilde{\nu}_o > \widehat{\nu}_o$ and that outcome noise enters the expression twice, we can see that the coding noise at the indifference point will be larger than the coding noise for log cost-benefits in binary binary choice. This, jointly with errors arising mainly at the list level, implies that between-task consistency will be lower in choice lists compared to binary choice.

D Experimental materials

For each specific experiment survey, please refer:

- Experiment I
  - choice lists and binary choice - Gains
  - choice lists and binary choice - Losses
- Experiment II
  - Benchmark-choice lists and Benchmark-binary choice
  - Treatment-Sequential

Below we show the experiment materials of Experiment I with the condition - BCs in gains - as an example.

[Page 1: Attention Pledge]

Attention Pledge

We ask that you give us your full attention throughout the study. Please refrain from all other activities, including using your phone and browsing the internet. If we find that you are not paying attention or are violating any rules you will be dismissed without being paid.

Specifically, by continuing with the study you declare:

- I will be available for the full-time of the study
- I will devote my full attention to the experiment and will not engage in other activities, such as browsing the internet
- I will put my mobile devices in silent mode and I will not use them during the study
- I will not communicate with any other participants during the session

I accept these requirements and wish to participate this study.

I reject these requirements and wish to leave this study.

[Page break here]

Welcome to our study. This study will take around 45 minutes. You will be asked to complete **several different tasks and a short questionnaire**. The answer to the questionnaire and all your decisions on the tasks will be private, and cannot be traced back to you personally.

If you complete the study you will be paid a fixed participation fee of **£7**. On top of that, you may earn additional money based on the decisions that you take during the study. Your participation in this study is voluntary. You have the right to withdraw at any point during the study. However, **if you do not complete the study, you will not be paid.**

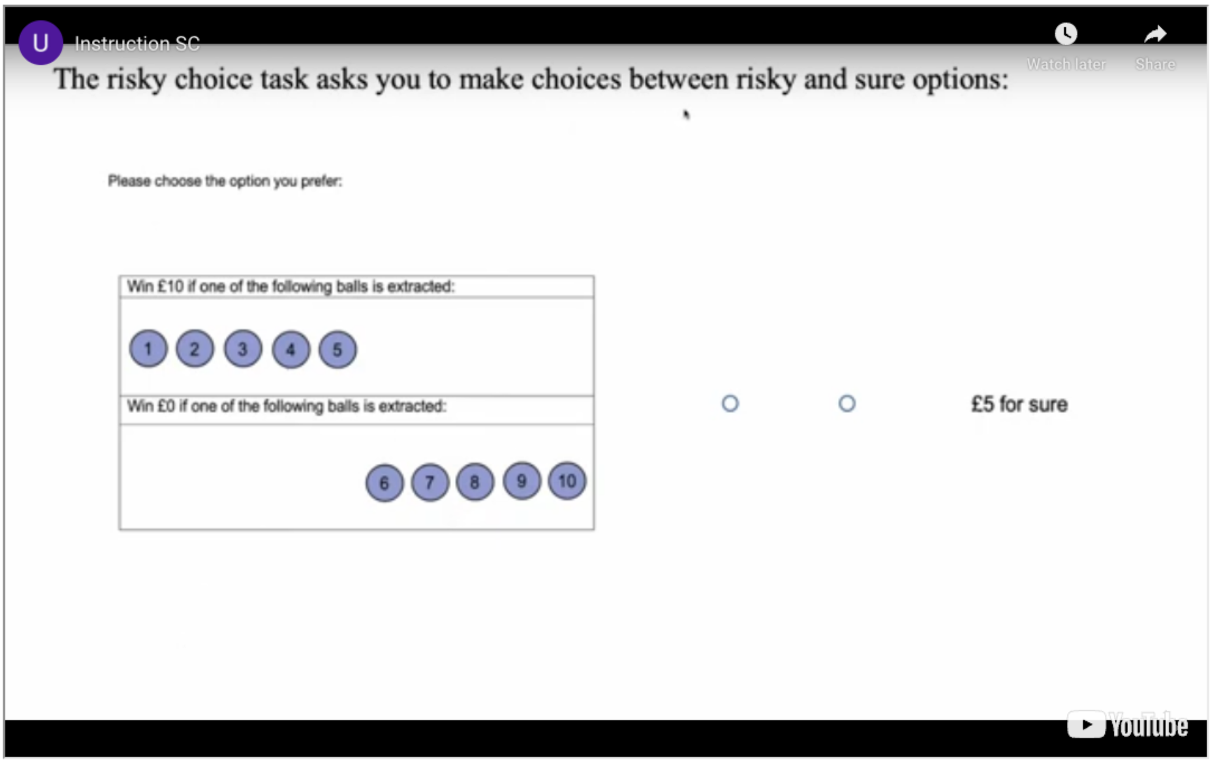
Please consider each decision carefully. Remember that your final payoffs from this experiment depend on the decisions you make.

By clicking the next button below, the instruction video of the risky choice task will be played.

Next

[Page 2: Video Playing]

Please watch the instruction video:



If you have any uncertainty about the content of the video, please feel free to watch it again.

[Page 3: Comprehension Check]

Comprehension Test

Before starting the experimental tasks, please answer the three questions to confirm your comprehension of the study. Please note that if you fail to answer the three questions correctly, the study will be terminated, and you will have to leave the study with being only paid £0.5.

Q1: Please choose the correct option:

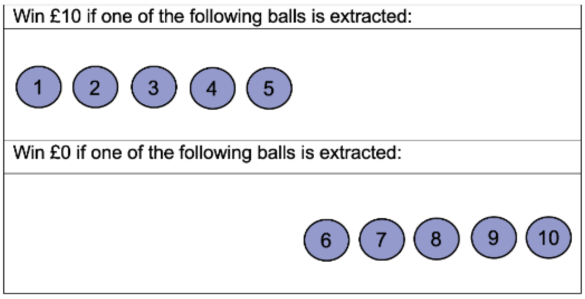
- You will only get paid with a fixed participation fee.
- Even if you withdraw your participation halfway, you would still get the participation fee and additional payouts from two tasks respectively.
- After completing the study, you will be paid an immediate participation fee. In addition, you might get additional payout(s) if you are selected by the system.

Q2: Please choose the correct option:

For the additional payout of the risky task, if you are selected by the system:

- The amount is randomly determined, irrespective of your responses.
- One of your choices is randomly picked by the system to play for real money, so it is important to carefully make every single choice.
- The amount is determined by a specific choice which you know in advance, so you only need to answer that question seriously.

Q3: For the below example choice,



£9 for sure

what outcomes would you have if you selected the lottery option?

Please choose the correct option:

- You have 5/10 chance of winning £10, and 5/10 chance of winning £0
- You have 9/10 chance of winning £9, and 1/10 chance of winning £1
- You will receive £9 for sure.

By clicking the NEXT button, you confirm and submit your answer. If all questions are correctly answered, you will proceed to answer the risky tasks. Good luck!

Appendix Reference

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

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

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

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A probabilistic programming language." *Journal of Statistical Software*, 76(1): 1–32.

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*, 132(648): 2737–2766.

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.

Feldman, Paul J, and Paul J Ferraro. 2023. "A Certainty Effect for Preference Reversals Under Risk: Experiment and Theory." Working paper.

Gabry, Jonah, Rok Češnovar, Andrew Johnson, and Steve Bronder. 2024. "cmdstanr: R Interface to 'CmdStan'." R package version 0.8.1, https://discourse.mc-stan.org.

Glöckner, Andreas, Benjamin E Hilbig, Felix Henninger, and Susann Fiedler. 2016. "The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice." *Journal of Experimental Psychology: General*, 145(4): 486.

Gold, Joshua I, and Michael N Shadlen. 2001. "Neural computations that underlie decisions about sensory stimuli." *Trends in cognitive sciences*, 5(1): 10–16.

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

Green, David Marvin, John A Swets, et al. 1966. *Signal detection theory and psychophysics.* Vol. 1, Wiley New York.

Haffke, Peter, and Ronald Hübner. 2014. "Effects of different feedback types on information integration in repeated monetary gambles." *Frontiers in Psychology*, 5.

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.

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

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2023. "Cognitive imprecision and stake-dependent risk attitudes." NBER Working Paper 30417.

L'Haridon, Olivier, and Ferdinand M. Vieider. 2019. "All over the map: A Worldwide Comparison of Risk Preferences." *Quantitative Economics*, 10: 185–215.

Ma, Wei Ji, Konrad Paul Kording, and Daniel Goldreich. 2023. *Bayesian Models of Perception and Action: An Introduction.* MIT press.

Natenzon, Paulo. 2019. "Random choice and learning." *Journal of Political Economy*, 127(1): 419–457.

Oprea, Ryan, and Ferdinand M. Vieider. 2024. "Minding the Gap: On the Origins of Probability Weighting and the Description-Experience Gap." Working paper.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica*, 66: 497–527.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." Vienna, Austria, R Foundation for Statistical Computing.

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

Vieider, Ferdinand M. 2023. "Cognitive Foundations of Delay-Discounting." Working Paper.

Vieider, Ferdinand M. 2024*a*. "Bayesian Estimation of Decision Models." RISLab.

Vieider, Ferdinand M. 2024*b*. "Decisions under uncertainty as bayesian inference on choice options." *Management Science*, 70(12): 9014–9030.