Meta-Analysis of Prospect Theory Parameters*

Taisuke Imai[1] Salvatore Nunnari[2]
Jilong Wu[3] Ferdinand M. Vieider[3,†]

[1]Institute of Social and Economic Research, The University of Osaka
[2]Department of Economics, Bocconi University
[3]RISL$\alpha\beta$, Department of Economics, Ghent University
[†]Corresponding author: ferdinand.vieider@ugent.be

June 3, 2025

Abstract

We present a meta-analysis of prospect theory (PT) parameters, summarizing data from 166 papers reporting 812 estimates. These parameters capture risk-taking propensities, thus holding interest beyond PT. We develop an inverse-variance weighted method that accounts for correlations in PT parameters and imputes missing information on standard errors. The mean patterns align with the stylized facts of diminishing sensitivity towards outcomes and probabilities discussed in PT. Beyond this, the analysis yields several new insights: 1) between-study variation in parameters is vast; 2) heterogeneity is difficult to explain with observable study characteristics; and 3) the strongest predictors are experimental and measurement indicators, revealing systematic violations of procedure invariance. These findings highlight the promise of cognitive accounts of behavior in organizing unexplained variation in risk-taking, which we discuss.

JEL Classification Codes: C11, D81, D91

Keywords: Prospect Theory, Probability Weighting Function, Meta-Analysis

*We gratefully acknowledge financial support from the Japan Society for the Promotion of Science (KAKENHI No. JP22K21358 to T.I.), the ERC (Grant No. 852526 to S.N.), and the Research Foundation Flanders (project nr. G008021N, "Causal Determinants of Preferences," to J.W. and F.V.).

1 Introduction

Economists have traditionally explained risk-taking behavior through the lens of diminishing marginal utility of money, an idea going back to the resolution of the St. Petersburg paradox by Daniel Bernoulli (1738/1954). A substantial body of literature, beginning with Preston and Baratta (1948), has further demonstrated systematic likelihood-dependence in risk-taking. Both outcome-dependence and likelihood-dependence in risk-taking have been formally integrated into prospect theory (PT; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wakker, 2010). A large number of studies have quantified the parameters governing the PT functionals. Here, we systematically investigate the consensus emerging from this large set of measurements. The interest of the investigation resides in the fact that PT parameters can be conceived of as summarizing systematic tendencies in risk-taking in general. The results we present here thus hold interest beyond prospect theory itself.

To achieve this, we conduct a systematic review of the universe of estimates of PT functionals by conducting a quantitative meta-analysis of all existing studies estimating PT parameters. This allows us not only to summarize aggregate parameters condensed from the underlying studies, but also to describe heterogeneity between studies, and to determine whether such heterogeneity can be explained using observable study characteristics. Our meta-analysis includes 812 parameter estimates drawn from 166 papers, which summarize the decisions taken by 52,000 subjects across 69 countries. We thereby analyze all PT parameters except loss aversion, which has been recently meta-analyzed by Brown et al. (2024).

Measurement errors in correlated parameter estimates. Meta-analysis consists in treating individual estimates reported in the literature as the object of study. The core model takes the form of a measurement error model: reported parameter means are treated as noisy estimates of the true, but latent, parameters, with the estimation noise assumed to be proportional to the standard errors of the parameters. The latent true parameters are then modeled as being drawn from a common mean, resulting in adjustments of the true study-level parameters towards the aggregate mean in proportion to 1) the standard error surrounding the estimate; and 2) the distance from the aggregate mean. Here, we model the true parameters as being drawn from a distribution of parameters, and further allow the estimates to systematically differ by study characteristics using meta-regression techniques.

Applying meta-analysis to multiple parameters estimated to describe data about risk-taking presents several challenges. One such challenge derives from poor statistical reporting: many papers report parameter estimates, but neglect to report statistical information that is sufficient to infer a standard error surrounding that estimate. Another challenge arises because PT parameters tend to be correlated both for mechanical (cf. Zeisberger, Vrecko and Langer, 2012) and substantive reasons (cf. Vieider, 2024*b*). Finally, parameterizations tend to differ between PT estimations, which results in non-trivial questions on how to jointly analyze the parameters. We tackle these issues by developing a novel Bayesian meta-analysis approach that 1) imputes missing standard errors organically within the model structure; 2) analyzes all PT parameters estimated in a given study *jointly*; and 3) leverages the correlation structures both in the estimation of latent parameters, and in the imputation of standard errors. This allows us to explicitly model the data-generating process underlying the estimates we analyze—a crucial element in the hierarchical modeling of data (see Gelman et al., 2014, ch. 8).

Meta-analytic parameter averages support stylized PT patterns. The meta-analytic average parameters across all 812 estimates we obtain—a "collective best guess" of what the PT parameters may be—provide strong support for the stylized PT patterns of decreasing sensitivity towards changes in wealth and probabilities. The average utility curvature (constant relative risk aversion, CRRA) coefficient for gains is 0.33 (95% credible interval: 0.31–0.36), indicating decreasing sensitivity towards increases in wealth. We also document decreasing sensitivity towards decreases in wealth (convex utility over losses). Sensitivity towards losses, however, decreases significantly more slowly than towards gains (with a mean of 0.29 and a credible interval of 0.25 to 0.32). The average elevation parameter of the probability weighting function is 0.98 (95% credible interval: 0.95–1.02), and it does not differ by domain. Overall, we do not find much support for either probabilistic optimism or pessimism.

We also find clear evidence for likelihood-insensitivity—the observation that relative risk aversion systematically increases in the probability of winning (decreases in the probability of losing). The mean likelihood-sensitivity parameter across outcome domains is 0.68 (95% credible interval: 0.66–0.70). Some individual studies in the literature have reported *over*-sensitivity to probabilities, resulting in an S-shaped probability weighting function. Surprisingly, however, such patterns receive no support in our latent-parameter estimates: the relatively large standard errors of the raw estimates indicating S-shapes combined with their outlying nature relative to the bulk of the evidence results in their meta-analytic "true effects" being invariably corrected to fall below the perfect sensitivity cutoff.[1]

We furthermore document significant correlations between PT parameters. The first set of correlations concerns the same parameters for gains and losses, showing reflection of risk attitudes between gains and losses (e.g., Schoemaker, 1990). Utility curvature parameters show a positive correlation, whereas the parameters governing the elevation of the probability weighting function show no correlation. Taken together, these correlations suggest that, at least at the level of aggregate estimates, risk attitudes are indeed reflected: subject populations that are more risk averse for gains tend to be more risk seeking for losses. Furthermore, populations that exhibit greater sensitivity toward probabilities in the gain domain also show higher likelihood-sensitivity in the loss domain. Beyond these gain-loss relationships, we also observe some evidence of correlations between different PT parameters: utility curvature and the elevation of the probability weighting function are significantly correlated across all studies. Additionally, there is some evidence of a correlation between utility curvature and likelihood-sensitivity.

Large unexplained variation in parameters results in poor predictive ability. The PT literature has mainly—and with a few exceptions on S-shaped probability weighting function mentioned above— emphasized agreement in estimates across studies. A surprising insight from our meta-analysis is thus just how much heterogeneity in parameter estimates there is. This is all the more surprising given that our meta-analytic measurement error model accounts for heterogeneity in estimates arising from sampling error. This is, indeed, one of the main rationales and hence strengths of meta-analysis. This large degree of heterogeneity furthermore persists once we control for a large set of observable study characteristics using meta-regression techniques. This shows that differences between studies are not (only) driven by obvious elements such as the measurement or estimation methods, the composition of the subject pool, the incentivization protocol, or the geographical location of the study.

One way to approach this issue is to consider our best guess of future parameter

[1]Note that we purely focus on given probabilities or *risk* in our meta-analysis. S-shaped probability weighting has, for instance, been documented by many studies when PT parameters are obtained from experience-based choice, where probabilities have to be discovered by sampling. Such studies are not included in our meta-analysis, given that they have been meta-analyzed recently by Wulff, Mergenthaler-Canseco and Hertwig (2018).

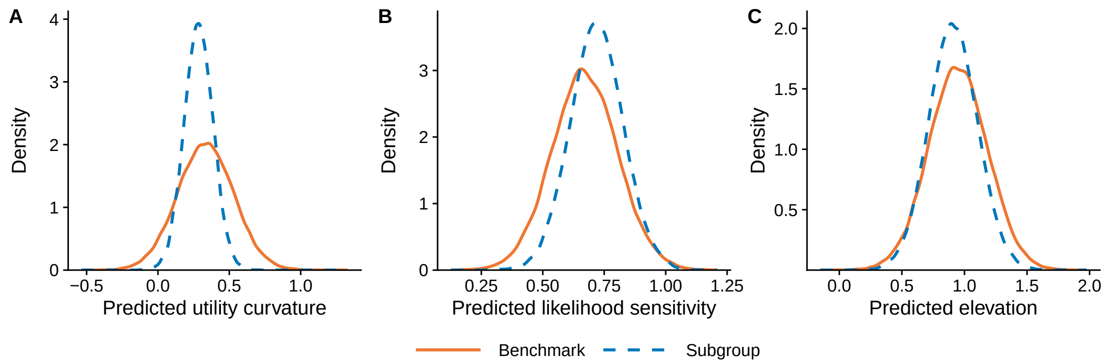

FIGURE 1: Predictive distributions for (A) elicited utility curvature, (B) likelihood sensitivity, and (C) elevation parameters. *Notes*: The solid orange lines represent benchmark predictions, which for the utility parameter controls whether the estimate is based on an exponential function. The dashed blue lines represent subgroup-specific predictions, which control for a large variety of observable study characteristics (the effects shown are for a study employing a linear in log-odds probability weighting function over gains, CRRA utility, measured through binary choice with real incentives, using aggregate data, conducted with university students in a laboratory setting in Europe).

values for a study with specified characteristics. Figure 1 presents the predictive distributions for the model parameters, both before and after controlling for observable study characteristics. Our best guess for a future study measuring parameters for gains and adopting a CRRA utility specification (which we always control for: see Methodological Framework) attributes 95.4% of the probability mass to values greater than zero, indicating concavity. The probability of observing convex utility over gains is 4.6%.[2] Predictions for the elevation parameter are similarly dispersed, encompassing both optimism and pessimism.[3] The only parameter for which predictions are qualitatively unambiguous is likelihood-sensitivity: nearly all of the probability mass (99.2%) indicates likelihood-insensitivity. Nonetheless, the predicted range of quantitative values remains very wide indeed.

Our predictions show little improvement even when controlling for a broad set of study characteristics, though some differences exist across parameters. After controlling for a range of study characteristics, we can explain 49.5% of the variance in the true, latent parameters for utility curvature. However, we only explain 21.5%

[2]Our analysis is conducted within a Bayesian framework, allowing for a direct probabilistic interpretation of the estimated distributions. We describe the probability mass associated with future outcomes accordingly. We refer to conventional cutoff points commonly used in the economics literature for significance testing.

[3]These results rely on posterior inferences about true effect sizes, which account for potential covariation among latent parameters. Thus, the observed variability cannot be attributed solely to econometric challenges in disentangling distinct motivational effects. We elaborate on this when we discuss the method and results.

in the extensive variation of likelihood-sensitivity, and only 20.9% in the variation of elevation (pessimism and optimism) estimates. This means that, after controlling for these study characteristics, our prediction for a future CRRA parameter (for a study employing a linear in log-odds probability weighting function over gains, measured through binary choice with real incentives, using aggregate data, conducted with university students in a laboratory setting in Europe) improves to attributing the entire probability mass to a concave utility function. However, there remains considerable uncertainty about future elevation and sensitivity parameters: the combined force from all existing PT estimates does still not allow us to make precise quantitative predictions about future estimates.

Meta-regression reveals evidence of violations of procedural invariance. The single most important predictor of the estimated PT parameters in our meta-regressions is the experimental method employed to elicit or measure risk attitudes.[4] Utility curvature is significantly less pronounced when a choice list design is used instead of a binary choice design. It is also attenuated in studies that employ bisection procedures to pinpoint indifference points. In contrast, lottery menu methods tend to yield greater utility curvature relative to binary choice. Choice list formats also influence probability weighting: they are associated with more elevated probability weighting functions and reduced likelihood sensitivity. Not all choice lists are, however, created equal. Notably, Holt and Laury (2002) type choice lists, where probabilities vary within the list, produce higher likelihood sensitivity than certainty equivalent lists, with a fair share even indicating likelihood *over*-sensitivity (similar effects have been documented for probability equivalents; see Feldman and Ferraro, 2023). Direct matching methods, typically implemented via willingness-to-pay or willingness-to-accept tasks, result in more depressed probability weighting functions (indicating pessimism for gains, optimism for losses).

These findings point to violations of *procedure invariance*—the principle that measured preference parameters should not depend on the method of elicitation, a tenet implicitly assumed by PT. Such results raise concerns for PT, which makes no predictions about measurement methods impacting behavior. This pinpoints another strength of meta-analysis: even while analyzing parameters estimated within a given theoretical framework, the methods allows us to pinpoint poten-

[4]Dummies for the elicitation method employed in the experiment account for 36.9%, 35.8%, and 9.4% of the overall heterogeneity we can explain for utility curvature, likelihood sensitivity, and elevation, respectively.

tial weaknesses in said framework, given that it makes no predictions about the differences in parameter estimates we observe.

Wider implications for accounts of risk-taking. Both the systematic violation of procedure invariance and the large degree of unexplained heterogeneity create problems for a model like PT, which explains risk-taking as a preference-based reaction to objectively perceived choice primitives. Our findings, in particular, imply that risk-taking behavior may be influenced by more subtle experimental cues, or details of the estimation procedure or data collection that ought to be irrelevant according to the model used to produce the estimates. One possibility is that reactions to these subtle cues reveals underlying cognitive mechanisms that generate PT-like behavior.

Some recent studies have indeed cast doubt on whether observed risk-taking behavior can be ascribed to stable "preferences", and as a consequence, whether it is meaningful to use experimental tasks to make inferences about such quantities. For instance, Khaw, Li and Woodford (2021) present a model in which apparent constant relative risk aversion over small stakes may be the result of noisy number perception. Khaw, Li and Woodford (2023), Vieider (2024*b*) and Frydman and Jin (2023) present theoretical accounts whereby apparent "probability weighting" may emerge from cognitive frictions in the mental representation of probabilities. Enke and Graeber (2023) show how likelihood-sensitivity in choice lists can be predicted by survey measures of 'cognitive uncertainty'. Oprea (2024) shows that the type of likelihood-insensitivity obtained under risk can be reproduced when removing the risk entirely, while maintaining the complexity of the choice situation. Oprea and Vieider (2024) show that providing redundant information under the form of samples from fully described choice options makes likelihood-insensitivity disappear, resulting in broadly neoclassical behavior.

Accounts of risk-taking as arising from cognitive frictions in the representation of choice stimuli, and optimal ways of dealing with such frictions by leveraging information about the distribution of choice stimuli in the environment, hold the promise to provide a generative account of the origin of the PT violations we document in this paper (Vieider, 2025). We thus see them as a prime way of explaining some of the puzzles we document in our meta-analysis—a point to which we will return in the discussion.

Limitations: Heterogeneity between studies vs. between individuals. A common limitation of meta-analysis is that, while it allows for the examination of absolute variation in (predicted) parameters, it remains difficult to benchmark between-study variation against individual-level variation. Brown et al. (2024) attempted to address this by using the distribution of loss-aversion parameters from the 30-country experiment of L'Haridon and Vieider (2019) as a benchmark. However, this is at best a proximate solution, as the uniformity of experimental methods and analyses in that study likely underestimates true individual-level heterogeneity. Properly addressing this issue is challenging. On the one hand, it is unclear what one might learn from such an exercise. Individual-level estimates, based on much sparser data, are inherently more variable than study-level estimates, which aggregate across individuals by design. Moreover, there is no clear benchmark for how much additional variability we should expect at the individual level. The problem is not merely data-driven but also involves substantive conceptual questions. Given the substantial unexplained heterogeneity we document across estimates, an experimental benchmark that simply reproduces key methods would be insufficient.[5] This important methodological question is therefore best addressed by a future large-scale experimental study explicitly designed to examine parameter variability at both individual and study levels.

Paper structure. The remainder of the paper is structured as follows. Section 2 introduces prospect theory (PT) and its most commonly used functional forms. Section 3 describes the assembly of the dataset and presents descriptive statistics for the characteristics of the collected studies. Section 4 outlines the Bayesian hierarchical model employed in this study. Section 5 presents the results, and Section 6 discusses their implications and concludes the paper.

2 Prospect Theory

We start by providing a succinct overview of PT, and by outlining the functional forms of utility and probability weighting used in the literature. For the sake of

[5] A meaningful benchmark—given the significant unexplained heterogeneity—would require re-analyzing the individual-level data from *all* studies included in our meta-analysis using all combinations of functional specifications employed in the literature. This is infeasible. Many datasets, including that of the seminal study by Tversky and Kahneman (1992), are no longer available. Moreover, elicitation formats are often tailored to specific functional forms, meaning not all combinations are identifiable across all datasets. Any benchmark based on a subset of studies would likely underestimate individual-level heterogeneity, rendering the conclusions from such an exercise unreliable.

clarity, we will focus on binary lotteries $(x, p; y)$, where outcome x occurs with probability p and outcome y with a complementary probability $1-p$. Such simple binary lotteries make up the lion's share of tasks used in the studies we meta-analyze (over 80%). For simple binary lotteries like this, original prospect theory, rank-dependent expected utility theory, dual-expected utility theory, and disappointment aversion are all special cases of PT (see Wakker, 2010, Section 7.11). This allows us to cast a wide net and to include studies estimating the functionals of these models in our analysis.

A lottery is deemed non-mixed if both outcomes are either positive or negative ($x > y \geq 0$ or $x < y \leq 0$), and mixed when it includes both a positive and a negative outcome, so that $x > 0 > y$. In PT, the value of a non-mixed lottery $(x, p; y)$ is given by:

$$w^s(p) \cdot v(x) + (1 - w^s(p)) \cdot v(y), \tag{1}$$

where v is the utility function (typically assumed to have a fixed point at 0, $v(0) = 0$), and w^s represents the probability weighting function, which satisfies $w^s(0) = 0$ and $w^s(1) = 1$; $s = +\ (-)$ denotes the gain (loss) domain.[6] For mixed lotteries $(x, p; y), x > 0 > y$, the PT value is represented as

$$w^+(p) \cdot v(x) + w^-(1 - p) \cdot v(y). \tag{2}$$

Mixed lotteries are used to estimate loss aversion—the kink of the utility at the origin. We exclude loss aversion from our meta-analysis since it has recently been analyzed elsewhere (Brown et al., 2024).

Empirically studying PT typically requires assuming a specific functional form (see Abdellaoui, 2000 and Bleichrodt and Pinto, 2000 for nonparametric estimates of PT; see Gonzalez and Wu, 1999 for a semi-parametric approach). These assumptions enable researchers to summarize patterns of choice across diverse tasks—differing in stakes and probabilities—using relatively few parameters. In our meta-analysis, we code up to three parameters per *estimate*: two for probability weighting and one for utility. A *study* may report separate parameters for gains and losses, which we treat as distinct estimates. Since few studies report a complete set of parameters for both domains, we code gains and losses separately and control for the outcome domain in all meta-regressions. While crucial in es-

[6] Some papers have estimated separable versions of probability weighting (e.g., Camerer and Ho, 1994), i.e., $w^s(p) \cdot v(x) + w^s(1 - p) \cdot v(y)$. Our dataset excludes these articles since they cannot be easily mapped into separable estimates.

timating PT parameters, the stochastic choice model has typically not received much attention in the PT literature, and detailed reports of the error terms are too scant to allow us to include an additional noise term in our analysis.

Utility functions. Panel A of Table 1 presents the functional forms commonly used in the literature, along with their frequencies in our dataset of 812 collected estimates. For further details on the data, see Section 3. We use ρ^s throughout to represent the utility curvature parameter, capturing attitudes towards stakes, where $s \in \{+, -\}$ indicates the corresponding payoff signs. Functions from the power utility family constitute the large majority of utility functions, accounting for 81.7% of the total (e.g., Tversky and Kahneman, 1992). For power utility functions, the parameter ρ^s is also known as the relative risk aversion coefficient.[7] The linear function represents 6.2% of all estimates, amounting to not estimating utility curvature at all (which is a special case of power utility with $\rho^s = 0$). The remaining category, with 12.2%, consists of exponential functions (e.g., Köbberling and Wakker, 2005).[8]

Utility functions from the power and exponential families are challenging to compare directly. Mappings from one into the other—while possible in principle—are only valid locally, and are very sensitive to the stake levels assumed for the approximation. We will thus consistently control for functional form in parametric analysis, and describe results for the two functional families separately. Panel A of Figure 2 illustrates three power utility functions obtained from different coefficient values. Specifically, the function is convex for $\rho^+ < 0$, linear for $\rho^s = 0$, and concave for $\rho^+ > 0$ (while not shown, it is concave for $\rho^- < 0$, and convex for $\rho^- > 0$). Assuming a linear probability weighting function, a convex utility function implies risk-seeking, while a concave utility function reflects risk aversion. Under more general, nonlinear probability weighting functions, risk attitudes are jointly captured by probability weighting and utility curvature—one of the reasons why we will analyze the different model parameters *jointly*.

[7]Note that simple power functions (where $u(x) = x^r$) coexist in the literature with CRRA functions. To make the parameters comparable, we consistently code ρ^s as the CRRA parameter for utility functions from the power family, i.e., $\rho^s = 1 - r^s$.

[8]Some PT estimations we encountered did not report which functional forms were used, and we thus felt compelled to exclude these estimates ($N = 46$). A few estimates also used less commonly used utility function specifications, such as the logarithmic function ($N = 5$) and the expo-power function ($N = 2$). Once again, we excluded these estimates from our analysis. Note, however, that papers reporting these estimates typically also reported alternative specifications, which we did include, so that we did not exclude the studies as a whole, but only the specific estimates using these non-standard functional forms.

Table 1: Utility and probability weighting function specifications.

Function	Functional form	Freq.	%
A: Utility			
Power (TK)	$v(x) = \begin{cases} \frac{1}{1-\rho^s} x^{1-\rho^s} & \text{if } x \geq 0 \\ \frac{-\lambda}{1-\rho^s} (-x)^{1-\rho^s} & \text{if } x < 0 \end{cases}$	663	81.7
Exponential	$v(x) = \begin{cases} \frac{1}{\alpha^s} (1 - \exp(-\alpha^s x)) & \text{if } x \geq 0 \\ \frac{-\lambda}{\alpha^s} (1 - \exp(-\alpha^s (-x))) & \text{if } x < 0 \end{cases}$	99	12.2
Linear	$v(x) = \begin{cases} x & \text{if } x \geq 0 \\ -\lambda x & \text{if } x < 0 \end{cases}$	50	6.2
Total number of estimates		812	100
B: Probability weighting			
Tversky-Kahneman (TK)	$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}$	281	34.6
Prelec II	$w(p) = \exp\left(-\frac{1}{\delta}(-\ln p)^\gamma\right)$	219	27.0
LLO	$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$	199	24.5
Prelec I	$w(p) = \exp\left(-(-\ln p)^\gamma\right)$	102	12.6
Power	$w(p) = p^{\frac{1}{\delta}}$	10	1.2
Gul	$w(p) = \frac{p}{1+(1-p)\frac{1}{\delta}}$	1	0.1
Total number of estimates		812	100

Probability weighting functions. We use γ^s to denote *likelihood-sensitivity*, and δ^s to designate the elevation of the function. The elevation has different meanings in the gain and loss domains: δ^+ captures *optimism*, while δ^- captures *pessimism*. This derives from the convention of attaching the decision weight $w(p)$ to the best outcome for gains, but to the worst outcome for losses (i.e., to largest loss). There are six probability weighting functions underlying the estimates in our dataset. Unlike utility functions, however, it is relatively straightforward to map parameters across several functional forms into each other. While some mappings rely on approximations, this approach enables us to analyze the parameters collectively. We will furthermore control for functional forms in meta-regression to determine the degree to which the functional form assumptions impact the parameter estimates.

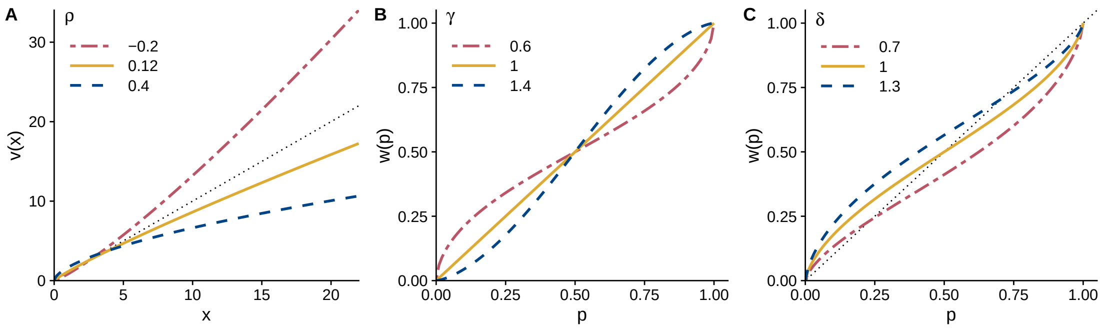

FIGURE 2: Examples of prospect theory functions. (A) The CRRA utility function with three levels of curvature in the gain domain. This is the specification used by Tversky and Kahneman (1992), $v(x) = \frac{x^{1-\rho}}{1-\rho}$ for $x \geq 0$. (B) The LLO function proposed by Goldstein and Einhorn (1987), $w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$, with three levels of likelihood insensitivity (γ) and a neutral elevation coefficient ($\delta = 1$). (C) The LLO function with three levels of elevation (δ) and a neutral likelihood insensitivity level ($\gamma = 1$).

We take the linear in log-odds (LLO; Gonzalez and Wu, 1999) specification as our baseline. In probability space, the function takes the form $w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$. This specification accounts for 24.5% of all estimates. The parameter $\gamma > 0$ captures likelihood-sensitivity—the phenomenon whereby risk attitudes systematically vary across probabilities. A value of $\gamma = 1$ captures perfect sensitivity (the EUT case), $\gamma < 1$ likelihood-insensitivity, and $\gamma > 1$ likelihood-oversensitivity. The parameter $\delta > 0$ captures the elevation of the function, with $\delta > 1$ capturing optimism for gains (pessimism for losses), and $\delta < 1$ capturing pessimism for gains (optimism for losses).[9] Panel B and C of Figure 2 respectively illustrate how likelihood-sensitivity and elevation affect the shape of a weighting function. Gul's disappointment aversion function (Gul, 1991) is a special case of the LLO function with the sensitivity parameter fixed to 1, and its pessimism parameter derived as the inverse of the LLO optimism parameter, δ^{-1}.

The Prelec (1998) 2-parameter function (Prelec II) does not have a direct mapping into the LLO function. To obtain a mapping, we approximate its anti-elevation parameter by the inverse of the LLO elevation parameter to have a common interpretation. We also equate the sensitivity parameters across the two functional forms. While this constitutes an approximation, the two functions are only dis-

[9]This interpretation of the parameters is particularly intuitive in the log-odds version of this weighting function, which takes the form $\log\left(\frac{w(p)}{1-w(p)}\right) = \gamma \log\left(\frac{p}{1-p}\right) + \log(\delta)$. Values of $\gamma < 1$—which takes the form of a power to the odds—will "compress" the odds towards 1, since odds smaller than 1 (i.e., a probability smaller than 0.5) will be uplifted towards 1, and odds larger than 1 will be reduced (with the opposite effect for $\gamma > 1$). Values $\gamma < 1$ thus produce a sort of regression to the mean of 1. This mean, however, is further affected by the value of δ, which acts as the intercept of the function in log-odds space.

tinguishable for extreme probabilities, which are rarely included in experiments for practical reasons. The Prelec I function is a special case of the Prelec II with $\delta = 1$. Another special case of the Prelec II function is the power function, with $w(p) = p^{\frac{1}{\delta}}$, where $\gamma = 1$. This is the case because $\exp(-\frac{1}{\delta}(-\ln(p))) = p^{\frac{1}{\delta}}$. This functional family makes up 40.6% of all estimates in our data. The most important function still missing is the Tversky and Kahneman (1992) (TK) function, used in 34.6% of all estimates, which cannot be reduced to any of the others. Nevertheless, the sensitivity parameter will typically be similar to those of other functions, and we will thus analyze them jointly.[10]

3 Data

3.1 Identification and Selection of Relevant Studies

We identified and selected papers estimating PT parameters based on clearly specified inclusion criteria. The primary criterion was to include "all empirical papers that estimate PT parameters." Under this criterion, we included papers utilizing choice data from both laboratory or field/online experiments and surveys conducted by letter or telephone call. Papers only estimating utility but not probability weighting were not included (see Online Appendix A.1 for precise search terms). Our search for relevant papers was conducted on the scientific citation indexing database Web of Science.

The initial search, performed in the summer of 2022, yielded a total of 2,034 papers. In the initial phase of paper identification, we scrutinized titles and abstracts, setting aside 1,453 papers that were evidently irrelevant to our study. Subsequently, we thoroughly examined the remaining papers, applying our inclusion criteria based on content, and proceeded to code the relevant information. Additionally, we employed IDEAS/RePEc and Google Scholar to explore unpublished working papers. The comprehensive search and selection procedure are outlined by Figure A.2 in the Online Appendix. By the end of this process, we identified 166 papers, 12 of which remained unpublished as of the initial data compilation in winter 2023.[11]

[10]Apart from the estimates reported in Table 1, three estimates adopted the Karmarkar specification (Karmarkar, 1978), $w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\delta}}$, which is a generalization of the TK. Given how few occurrences there are ($N = 3$) and the difficulty of linking it to other forms, we exclude them from our analysis.

[11]Once a first version of this paper is completed, we will share it with the community to determine whether we missed any papers. We are also committed to updating our database and

3.2 Data Coding

Our meta-analysis dataset is assembled by encoding estimates of PT parameters and their associated standard errors (SEs), as well as details about the experiment, subject pool, and estimation procedures underlying the parameters. For meta-analyzes, SEs play a critical role in computing weighted averages. In situations where SEs are not explicitly reported, we reconstructed them utilizing alternative available information, such as standard deviations (SD), p-values, or approximated them using the inter-quartile range (IQR). A detailed overview of our calculation methodology can be found in the Online Appendix A.2. We also coded variables detailing the location of the experiment (e.g., lab, field, classroom, online), types of rewards (e.g., real money or hypothetical money, and other consequences), subject population (e.g., university population, general population, and other population like farmers and athletes), functional forms of probability weighting (e.g., Prelec I, Prelec II, TK, and LLO) and utility (Power, Exponential, and Linear) which will be described in Section 2, among other characteristics. Online Appendix A.3 contains a comprehensive list of all the variables coded in the study. A random subset of 10% of the data was coded independently by at least two co-authors, to ensure data quality and coding consistency.

3.3 Descriptive Statistics

We identified a total of 166 papers for inclusion in our meta-analysis (see the Online Appendix A.4 for a list). Out of these, 154 articles were published across 61 journal outlets, and the rest remained unpublished. The dataset encompasses papers from diverse disciplines, such as economics, management, psychology, neuroscience, medicine, psychiatry, agriculture, environment, transportation, and operations research. Moving forward, we shift our focus to the primary variable of interest—the estimated parameters of the utility function and probability weighting function. The dataset comprises a total of 812 estimates (refer to Table 2). Of these estimates, 54.7% pertain to aggregate or 'representative agent' estimations, pooling data from all subjects. Some 24.4% report means of individual-level estimates, and 20.9% report medians. There are 61 cases where both the mean and median of the distribution of the PWF parameters estimated at the individual level are available. 113 of the 812 estimations do not report SEs or any statistical information that can be used to calculate SEs, including the seminal estimation of Tversky and Kahneman (1992).

results close to the publication date, to make sure the results are as up-to-date as possible.

Table 2: Types of PT estimates.

	All estimates		With SE	
	Freq.	%	Freq.	%
Aggregate-level	444	54.7	390	55.8
Individual-level mean	198	24.4	180	25.8
Individual-level median	170	20.9	129	18.5
Total	812	100.0	699	100.0

We also determined the country where the data were collected. The majority of papers report estimates from data collected in a single country. Four collected data from multiple countries/regions, and two of these four (Rieger, Wang and Hens, 2017; L'Haridon and Vieider, 2019) conducted large-scale cross-country studies, gathering data from 30 countries or more. In total, the estimates included in our dataset originate from 69 countries (see Figure A.3 for a global map). Some of the main characteristics of interest to our analysis include the type of reward, the type of subject pool, the data type (e.g., lab and field), the type of elicitation method, and the domain of stimuli. Table 3 reports the descriptive statistics of these characteristics.

Our dataset primarily comprises studies conducted in Europe, which represent 59.6% of the total. The second largest proportion, accounting for 16.4%, originates from North America. Additionally, 30 studies (12.8%) were carried out in Asia, while a few others were conducted across various other continents. Binary choices emerged as the most popular data collection method, utilized in 256 estimates (31.5%). The choice list format was the preferred method for 31.3% of the estimates. Among these choice lists, certainty equivalents were the most widely used tool, accounting for 217 out of 254 estimates. The third most common approach was the matching method (20.9%), which involves directly asking participants to provide a matching value, usually in the form of a willingness-to-pay or willingness-to-accept measure. The bisection method constitutes a smaller share, while the lottery menu category has 2.7% of estimates.

The majority of our data comes from laboratory experiments (72.2%), augmented by studies conducted in field settings or online.[12] Monetary rewards, whether real or hypothetical, were overwhelmingly the most common form of incentive, mak-

[12]Online experiments primarily refer to those conducted on online platforms, such as Prolific and Amazon MTurk. This category also includes a small proportion of studies conducted via mail or phone call.

Table 3: Characteristics of studies estimating PT parameters.

	Freq.	%		Freq.	%
Total number of studies	812	100			
Continent type			*Data type*		
Europe	484	59.6	Lab	586	72.2
North America	133	16.4	Class	120	14.8
Asia	104	12.8	Online	58	7.1
Central/South-America	44	5.4	Field	48	5.9
Africa	26	3.2	*Reward type*		
Oceania	21	2.6	Real money	464	57.1
Elicitation type			Hypo money	307	37.8
Binary	256	31.5	Other	41	5.0
List	254	31.3	*Subject type*		
Matching	170	20.9	University	652	80.3
Bisection	110	13.5	Other	85	10.5
Lottery Menu	22	2.7	General	75	9.2

Notes: Regarding *Data type*, the category of "online" includes experiments conducted online as well as other special cases, such as survey data collected via phone calls or mail. For *Subject type*, the "other" category mainly includes farmers ($N = 20$), athletics ($N = 20$), health professionals ($N = 10$), businessman ($N = 11$). Last, the other reward types contain health and time.

ing up over 90% of all rewards offered. Non-monetary rewards, such as health and time, were less common. A significant portion of the subjects were university students (80.3%) or the general public (10.5%), and the remaining studies focused on niche demographics, including businessmen, health practitioners, and farmers.

4 Methodological Framework

We begin our analysis within a conventional meta-analytic framework, treating the parameter estimates reported in the literature as noisy measurements of underlying true effect sizes (parameter values). However, applying this standard approach to Prospect Theory (PT) parameters presents several challenges. Most important, perhaps, is the interdependence of the parameters: because PT parameters *jointly* characterize risk attitudes and are econometrically identified from choice data, assuming independence among them can lead to biased inferences. To address this issue, we extend standard meta-analytic methods to a joint modeling framework that explicitly accounts for the correlation structure among parameters. Moreover, the presence of missing standard errors and the variability in which subsets of parameters are reported across studies require further adaptations. These modifications allow us to incorporate as many estimates as possible in the meta-analysis while maintaining methodological rigor.

4.1 Joint Meta-Analysis Model

Our core model closely follows the standard meta-analytic framework but extends it to accommodate multiple, potentially correlated parameters. We begin by outlining this multivariate generalization of conventional meta-analysis methods.

Measurement error model. Let $\boldsymbol{\theta}_i = (\rho_i, \gamma_i, \delta_i)$ be a vector of parameters for estimate i (which we will, for now, assume to be complete). Let $\boldsymbol{se}_i = (se(\rho)_i, se(\gamma)_i, se(\delta)_i)$ be a vector containing the standard errors for estimate i corresponding to the parameters in $\boldsymbol{\theta}_i$. We assume that the encoded parameters $\boldsymbol{\theta}_i$ constitute a noisy measure of the true but unobserved underlying effect sizes, designated by $\widehat{\boldsymbol{\theta}}_i$. The noise may arise from sampling error, weak parameter identification, or limitations in econometric procedures. We capture this relationship using the following measurement error model:

$$\boldsymbol{\theta}_i \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_i \,, \, \mathrm{diag}\left(\boldsymbol{se}_i^2\right)\right),$$

where the diag operator transforms the vector of squared standard errors into a variance-covariance matrix, placing the squared standard errors on the main diagonal and zeros on the off-diagonal. Each parameter in $\boldsymbol{\theta}_i$ is thus modeled as providing a measure of the true latent parameter in $\widehat{\boldsymbol{\theta}}_i$, with potential measurement error proportional to $\boldsymbol{se}_i^2$.

Meta-analytic aggregation. Assuming that the parameter estimates are drawn from an underlying distribution of parameters, one source of variation is sampling error. Additional noise may arise from weak parameter identification, which can stem from limitations in the data, the econometric methods employed, or both—each contributing to large standard errors around individual estimates. The impact of such errors can be mitigated by aggregating across studies. This approach effectively pools individual parameter estimates toward the meta-analytic mean, with greater weight assigned to estimates with lower uncertainty:

$$\widehat{\boldsymbol{\theta}}_i \sim \mathcal{N}\left(\boldsymbol{\mu}\,,\, \boldsymbol{\Sigma}\right), \tag{3}$$

where $\boldsymbol{\mu}$ denotes a vector of parameter-specific meta-analytic means, and $\boldsymbol{\Sigma}$ is a covariance matrix containing the variances $\boldsymbol{\sigma}^2$ along its main diagonal and the covariances between parameters in the off-diagonal elements. This covariance structure explicitly allows for correlations among the true effect sizes, capturing the extent to which the different parameters co-vary. By modeling the covariance structure of the true (rather than observed) effect sizes, we can account for substantive reasons underlying potential parameter correlations (see, for example, Vieider, 2024*b*, for model-based predictions of such dependencies). More importantly, this approach enables us to assess these correlations while controlling for observable study characteristics, which we describe below.

Meta-regression. Given the wide variation in measurement methods, estimation techniques, and study locations in our dataset, it is highly unlikely that measurement error alone accounts for the observed differences in parameter estimates $\boldsymbol{\theta}_i$. This highlights the importance of controlling for study-level characteristics, both to better explain between-study variance and to obtain more accurate estimates of the aggregate effect sizes. More critically, when study-level differences that may influence effect sizes are known *ex ante*, the assumption of exchangeability—that is, the notion that effects are drawn from a common distribution—is violated. Exchangeability is a foundational assumption in meta-analysis (Gelman et al., 2014), and it is clearly violated in our data. For example, exponential utility parameters are expressed on a different scale than CRRA parameters, and our dataset includes separate entries for gains and losses. To address these issues, all meta-regressions include controls for the functional form of utility and for the outcome domain.[13]

This is achieved by replacing $\boldsymbol{\mu}$ with $\boldsymbol{X}_i\boldsymbol{\beta}$ in (3), where $\boldsymbol{X}_i$ is a $1 \times K$ vector of study characteristics for estimate i (including a column of ones to capture the intercept), and $\boldsymbol{\beta}$ is a $K \times 3$ matrix of regression coefficients. This specification allows us to examine, inter alia, the extent to which the variance in each parameter

[13] An alternative modeling approach would be to include parameters for gains and losses within a single parameter vector. While this would allow for the exploitation of additional correlations, it would preclude a straightforward meta-analytic assessment of *differences* between gain and loss parameters. We therefore chose to treat losses as separate estimates, especially given the relatively small number of studies that jointly estimate the full set of gain and loss parameters.

can be explained by the observable characteristics captured in $\boldsymbol{X}$. Let θ_i denote a generic (scalar) parameter, and let τ^2 represent its variance. We then obtain a measure of the explained variance as $R^2 = 1 - \tau_1^2/\tau_0^2$, where τ_0^2 is the variance from a model without covariates (i.e., the baseline model), and τ_1^2 is the variance from the regression model whose explanatory power is being assessed.

4.2 Model Refinements

The model described above follows a conventional approach as typically used for scalar parameters, with its primary innovation being the joint analysis structure, which facilitates the assessment of parameter covariances. We now introduce refinements that enable the simultaneous incorporation of all coded parameters into the meta-analysis, ensuring that each parameter both contributes to and is informed by the endogenously estimated meta-analytic averages.

Error imputation. Several studies estimating PT functionals report parameter estimates but lack sufficient statistical information to recover standard errors. This includes the seminal work by Tversky and Kahneman (1992). In some cases, studies provide statistical information for certain parameters but not for others. Rather than discarding such studies and losing valuable data, we treat the missing standard errors as missing variables and impute them within our meta-analytic model. Note that this approach assumes that missing statistical information is "missing at random," i.e., there are no systematic differences between parameters from studies that report standard errors and those that do not. Comparisons between parameters with and without reported standard errors support this assumption in our dataset. As a robustness check, we also conduct the meta-analysis using only the subset of data with complete observations. The results are consistent with those presented in the main text (see Online Appendix B).

To impute standard errors, we construct an error-prediction model. Let $\log(\boldsymbol{se}_i)$ denote the element-wise natural logarithm of the standard errors, which ensures non-negativity. We model this transformed vector as:

$$\log(\boldsymbol{se}_i) \sim \mathcal{N}\left(\boldsymbol{Z}_i \boldsymbol{\xi}\,,\, \boldsymbol{\Omega}\right), \tag{4}$$

where $\boldsymbol{Z}_i$ is a $1 \times M$ vector of characteristics of study i predictive of its standard errors, $\boldsymbol{\xi}$ is a $M \times 3$ matrix of coefficients, and $\boldsymbol{\Omega}$ is a covariance matrix with variances on the diagonal and covariances on the off-diagonal. The predictors in

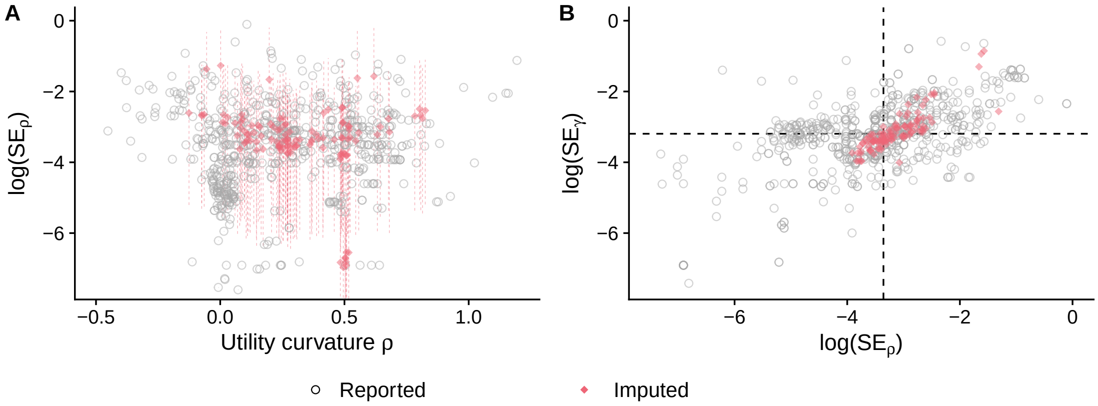

FIGURE 3: Missing standard error imputation and correlation of standard errors for utility curvature and likelihood sensitivity. (A) Scatter plot showing the relationship between power utility curvature (ρ) and the logarithm of its standard errors (SEs). (B) Scatter plot showing the correlation between the logarithms of the SEs for utility curvature (SE_ρ) and likelihood sensitivity (SE_γ). *Notes*: In both panels, observations with reported standard errors (SEs) are displayed as gray circles, while those with imputed SEs are highlighted with red diamonds. The x-axis is truncated for improved visualization. In Panel A, dashed red lines represent posterior uncertainty intervals for the imputed values. In Panel B, dashed gray lines indicate the medians of the respective variables.

$\boldsymbol{Z}_i$ include the characteristics such as elicitation methods and the square root of the number of subjects—which, as expected, is strongly negatively correlated with the standard errors—and a constant term to capture the intercept.

The vector $\log(\boldsymbol{se}_i)$ has a distinctive structure: it includes the logarithm of observed standard errors for studies where these are available, and the logarithm of *predicted* standard errors for cases where the standard errors are missing and must be inferred. The model thus fulfills two roles: (1) it estimates the regression coefficients $\boldsymbol{\xi}$ from the observed study characteristics in $\boldsymbol{Z}$; and (2) it uses these characteristics, along with the estimated coefficients $\boldsymbol{\xi}$, to *predict* (or impute) standard errors where they are not observed. The correlation structure encoded in $\boldsymbol{\Omega}$ ensures that the dependencies among errors are properly accounted for during the imputation process.

Figure 3 illustrates the impact of error imputation on the utility curvature parameter, ρ. The effects on other parameters are qualitatively similar and are presented in Online Appendix B. Panel A shows that the imputed standard errors, depicted as red diamonds, fall within the range of the observed standard errors. The vertical dashed lines highlight a key feature of imputed errors: as variables imputed within the model, they are treated as uncertain quantities themselves. This modeling approach has the advantage that the second-order uncertainty is incorporated directly into the estimation process. Panel B demonstrates the correlation between the standard errors of the utility parameter and the likelihood-sensitivity parameter, again highlighting imputed values in red diamonds. The figure shows two things. First, the standard errors are strongly correlated: the correlations between parameters estimated from $\mathbf{\Omega}$ are 0.56 between ρ and γ, 0.51 between ρ and δ, and 0.75 between γ and δ. Second, this correlation is clearly mirrored in the imputed standard errors. In fact, the correlation structure has a stronger influence than sample-size effects alone, as evidenced by the clustering of imputed standard errors around the 45-degree line.

Missing parameters. Of the 812 estimates in our dataset, only 372 include a complete set of three parameters. An additional 389 estimates contain both a utility and a sensitivity parameter, while 47 include two probability weighting parameters but no utility parameter, implicitly assuming linear utility. The remaining 14 estimates either include a utility and elevation parameter or only estimate sensitivity (three cases). To enable a unified joint estimation across all available data, we construct sub-models tailored to each parameter combination. Importantly, the overall model is structured so that estimates with missing parameters still contribute to the estimation of relevant regression coefficients, as well as to the associated variance and covariance components.

Consider a general case involving a two-parameter model, with parameters indexed by j and k. The imputation model is specified as:

$$\log(\boldsymbol{se}\{j,k\}_i) \sim \mathcal{N}\left(\boldsymbol{Z}_i\boldsymbol{\xi}_{\{j,k\}}\,,\, \mathbf{\Omega}_{\{j,k\}}\right),$$

where $\log(\boldsymbol{se}_i\{j,k\})$ denotes a vector comprising the log standard errors for parameters j and k, and the subscript $\{j,k\}$ indicates subsetting of vectors to elements $\{j,k\}$ (and of matrices to the corresponding rows and columns. The modeling of measurement error and the hierarchical aggregation of latent parameters proceed analogously:

$$\begin{aligned}
\boldsymbol{\theta}\{j,k\}_i &\sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}\{j,k\}_i\,,\, \mathrm{diag}\left(\boldsymbol{se}\{j,k\}_i^2\right)\right),\\
\widehat{\boldsymbol{\theta}}\{j,k\}_i &\sim \mathcal{N}\left(\boldsymbol{X}_i\boldsymbol{\beta}_{\{j,k\}}\,,\, \mathbf{\Sigma}_{\{j,k\}}\right),
\end{aligned}$$

where $\widehat{\boldsymbol{\theta}}\{j,k\}_i$ represents the latent effects corresponding to the pair of parameters j and k, and the remaining notation follows the same subsetting convention described above.

Performing this iteratively across all parameter combinations ensures that: (1) all parameter estimates can be analyzed jointly, even when the number of parameters differs; and (2) all parameters contribute to the estimation of *the same* regression and covariance parameters, conditional on study characteristics in $\boldsymbol{X}$.

4.3 Implementation

We estimate our model using Bayesian hierarchical techniques in Stan (Carpenter et al., 2017), launched from `R` (R Core Team, 2023) through `CmdStanR` (Gabry et al., 2023). Meta-analysis is inherently Bayesian, since the endogenously estimated parameters in $\boldsymbol{\mu}$ serve as priors for the latent parameters in $\boldsymbol{\theta}$. The key distinction from frequentist approaches lies in whether explicit priors are specified for the aggregate-level parameters—when they are not, such methods are often labeled "empirical Bayes," regardless of the estimation technique used. In our model, we specify diffuse hyperpriors so as not to affect our conclusion in any way, serving purely as an aid to make the exploration of the posterior parameter space more efficient.

The coding in Stan framework provides several advantages. It allows us to hand-code a model precisely tailored to our data and directly quantify the probability mass in favor of a given hypothesis or parameter range, consistent with Bayesian interpretability. Importantly, the Bayesian framework allows us to directly assess the plausibility that a parameter lies within a region close to a neutral benchmark (e.g., perfect likelihood-sensitivity), by quantifying the posterior probability mass within that region. This enables us to both support and challenge such benchmark-based hypotheses based on the posterior distribution. For a practical guide to estimating (hierarchical) decision models in Stan, including sections on measurement error models and meta-analysis, see Vieider (2024*a*).

5 Results

We structure the results section by presenting findings for each parameter individually, while emphasizing that all estimates are derived from a joint analysis model. We conclude with a discussion of parameter correlations and an examination of potential publication bias. Our meta-analytic estimates control for the use of exponential utility, due to its distinct properties relative to power utility. However, when reporting results for the two weighting function parameters, we do not control for the utility specification, as it showed no significant effect. All reported effects of study characteristics, including outcome domains, are based on meta-regressions that simultaneously include all relevant study-level covariates.

5.1 Utility Curvature ρ

Descriptive statistics. The coded CRRA utility coefficients have a mean of 0.30 across domains, a median of 0.27, and an interquartile range (IQR) of $[0.08, 0.51]$. This indicates that most studies report a ρ value between 0 and 1, although a minority (16.7%) of estimates are zero or negative—predominantly in the loss domain (see below). For exponential utility, the IQR is $[0.00, 0.05]$, with a median of 0.03 and a mean of 0.08. However, 24.2% of these estimates are negative. These observations align with the stylized patterns discussed in the PT literature, where the utility function is typically concave for gains and convex for losses, exhibiting decreasing sensitivity towards changes in wealth.

Parameter distributions. Panel A of Figure 4 shows the density of the estimated latent parameter values ($\widehat{\rho}_i$), represented by the blue dashed line. This density is notably narrower than that of the raw data values, ρ_i, reflecting the effect of meta-analytic pooling: individual estimates that diverge substantially from the overall mean are adjusted toward more credible values. The degree of this adjustment, often referred to as shrinkage, is inversely related to the precision of the estimates; that is, only outliers with large standard errors are substantially pulled toward the mean. We estimate the overall mean of the CRRA functions at 0.32, with a Bayesian 95% credible interval (CrI) of $[0.29, 0.34]$. In contrast, the mean for the exponential function is much lower at 0.05, with a CrI of $[0.00, 0.10]$, as reflected by the peak around 0 in the density plot. These results confirm the expected pattern: individuals generally display a concave utility function for gains and a convex utility function for losses.

Gain-loss comparison. An important question is whether PT parameter estimates differ by domain—specifically, whether individual utility curvature varies between gains and losses. Panel B of Figure 4 illustrates the empirical cumulative distribution functions (ECDFs) for encoded ρ^+ (gains) and ρ^- (losses). The plot reveals that, on average, individuals exhibit a concave utility function for gains (mean, 0.30; median, 0.25) and a convex utility function for losses (mean, 0.21;

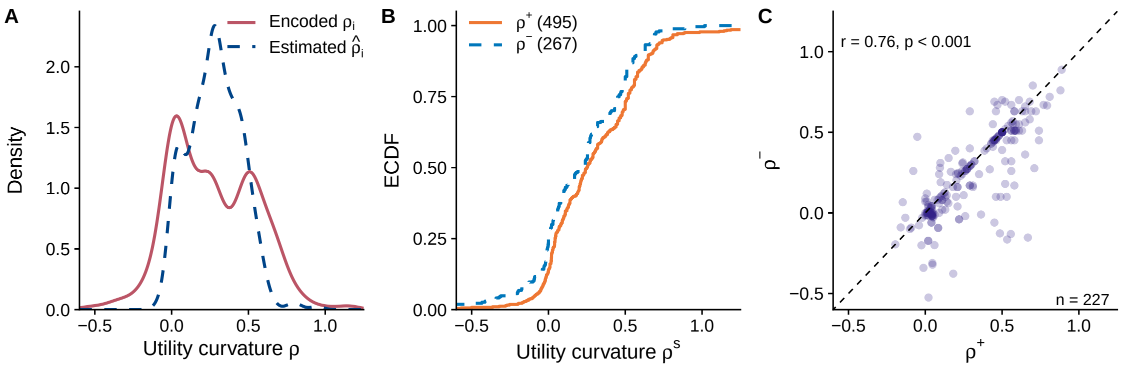

FIGURE 4: (A) Density plot comparing encoded ρ_i values to their corresponding estimates $\widehat{\rho}_i$. (B) Empirical CDFs of the power utility curvature, separated by gain and loss domains. (C) Scatter plot of parameter estimates for studies reporting both gain and loss values. *Notes*: The x-axis is truncated for improved visualization. However, density estimations include all observations, including those beyond the displayed range.

median, 0.23). Moreover, the ECDF for ρ^+ lies almost entirely below that for ρ^-, suggesting that utility functions tend to be more linear for losses than for gains. Additionally, negative values of ρ are more prevalent for losses (24.3%) than for gains (14.1%), suggesting that convex utility for losses coexists with a relatively larger share of concave cases.

It is important to emphasize that differences across outcome domains should not be interpreted causally, as they may partially reflect variation across studies. To more precisely assess the causal role of outcome sign, we focus on the 227 observations reporting estimates of ρ for both gains and losses. The scatter plot in Panel C shows that observations lie below the diagonal line more frequently than above it. We observe a strong correlation in ρ across domains ($r = 0.76$, $p < 0.01$), and a statistically significant difference between gains and losses based on a nonparametric test ($p < 0.01$), in line with patterns in the full dataset. This finding is further supported by a meta-regression including a loss-domain dummy together with other study characteristics: the estimated coefficient is -0.04, which is significantly different from zero (see Table 4 for full results).

Procedure invariance. An important question concerns whether elicited choice patterns vary across different measurement methods. In our dataset, four commonly used elicitation methods are binary choice, bisection procedures, choice lists, and direct matching. As shown in Panel A of Figure 5 and reported in Table 4, utility curvature tends to be more pronounced when measured using binary choices than with the bisection method, direct matching, or certainty equivalent

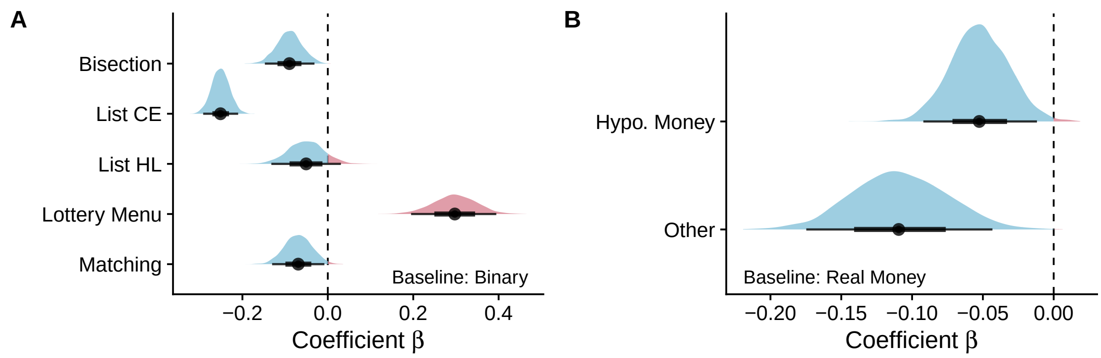

FIGURE 5: Posterior distributions of meta-regression coefficients. (A) Different elicitation methods compared to binary choices. (B) Two alternative reward types compared to real monetary rewards. *Notes*: The posterior distributions of the Bayesian random-effects meta-regression coefficient(s) β, along with the posterior medians (represented by a black dot), 66% credible intervals (indicated by thick solid lines), and 95% credible intervals (shown as thin solid lines), are displayed.

(CE) choice lists—the latter showing the strongest difference. However, curvature estimates from binary choices are still smaller than those obtained from methods that primarily rely on various types of lottery menus.

The most pronounced effects are associated with CE choice lists, which tend to yield power utility curvature estimates that are 0.25 lower than those obtained using the baseline method of binary choices (see Bouchouicha et al., 2024 for a discussion of implications for risk aversion). Bisection and direct matching methods also yield significantly lower estimates, with reductions of 0.09 and 0.07, respectively. In contrast, lottery menus produce substantially higher estimates—about 0.30 above those from binary choices and notably higher than all other methods. Closer examination shows that this outlier pattern is primarily driven by Cheung and Johnstone (2017), which accounts for 12 of the 22 lottery menu estimates and employs a distinct experimental design and estimation approach.[14] These findings highlight the considerable impact that methodological choices can have on the estimation of utility curvature.

Explained versus unexplained heterogeneity. Additional observable study characteristics that may help explain between-study variance are summarized in Table 4. Panel B of the figure further illustrates that utility curvature tends to be less pronounced for hypothetical outcomes and for outcomes other than

[14] Cheung and Johnstone (2017) used an investment task in which earnings depend on both risk and relative skill placement. The authors estimated PT parameters while incorporating an additional subjective belief parameter, q, reflecting perceived skill in the task.

Table 4: Meta-regression analysis of utility curvature.

Category	Variable	Median	2.5%	97.5%
Sign	Loss	-0.04	-0.07	-0.02
U	Expo	-0.22	-0.27	-0.18
PWF	LLO	Baseline		
	Gul/Power	0.59	0.46	0.72
	Prelec I	0.13	0.08	0.19
	Prelec II	0.09	0.04	0.15
	TK	0.03	-0.01	0.08
Elicitation	Binary choice	Baseline		
	Bisection	-0.09	-0.15	-0.03
	List CE	-0.25	-0.29	-0.21
	List HL	-0.05	-0.13	0.04
	Matching	-0.07	-0.13	0.00
	Lottery Menu	0.30	0.19	0.39
Estimate	Aggregate	Baseline		
	Ind. Mean	-0.01	-0.05	0.03
	Ind. Median	-0.02	-0.07	0.02

Category	Variable	Median	2.5%	97.5%
Subject	Univ. pop.	Baseline		
	General	0.03	-0.03	0.10
	Other	-0.02	-0.08	0.04
Reward	Real Money	Baseline		
	Hypo. Money	-0.05	-0.09	-0.01
	Other	-0.11	-0.17	-0.04
Data	Lab	Baseline		
	Class	0.30	0.23	0.38
	Field	0.07	-0.01	0.14
	Online	-0.1	-0.18	-0.02
Continent	Europe	Baseline		
	Africa	0.18	0.10	0.26
	Asia	0.02	-0.02	0.07
	C/S-America	0.13	0.07	0.19
	North America	0.13	0.09	0.18
	Oceania	0.16	0.03	0.29
Frame	Visual aids	0.07	0.02	0.11

money. However, we caution against interpreting these associations as causal. These categories often involve substantially higher stakes than incentivized experiments, raising the possibility that the observed effects may reflect underlying stake effects—specifically, utility functions exhibiting *increasing* relative risk aversion over broader stake ranges (Holt and Laury, 2002; Bouchouicha and Vieider, 2017).

Even after accounting for these study-level characteristics, approximately 50% of the variation in parameter estimates remains unexplained. Notably, a substantial portion of the explained variation is attributable to the elicitation method, highlighting its central role in the analysis. To assess its relative importance, we compare the full meta-regression model with an alternative specification that excludes the elicitation method. This comparison indicates that the elicitation method alone accounts for 36.7% of the explained variation.

5.2 Likelihood Sensitivity γ

For the likelihood-sensitivity parameter, we exclude 11 observations corresponding to the Gul and Power functions, as these specifications fix the parameter γ at 1. This leaves 801 parameter estimates that inform our analysis of likelihood sensitivity.

Descriptive statistics. We begin by examining the overall distribution of the likelihood sensitivity parameter, γ^s, depicted in Panel A of Figure 6 (red solid

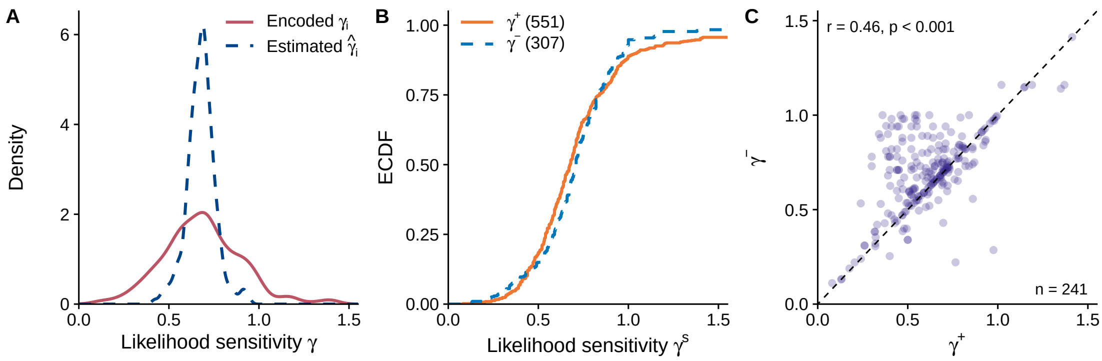

FIGURE 6: (A) Density plot comparing encoded γ_i values to their corresponding estimates $\widehat{\gamma}_i$. (B) Empirical CDFs of the likelihood sensitivity parameter, separated by gain and loss domains. (C) Scatter plot of parameter estimates for studies reporting both gain and loss values. *Notes*: The x-axis is truncated for improved visualization. However, density estimations include all observations, including those beyond the displayed range.

curve). The mean and median are 0.72 and 0.68, respectively, with an IQR of $[0.55, 0.83]$. These values are consistent with an inverse-S shaped probability weighting function, a well-established pattern in the literature. Notably, approximately 7.7% of the raw estimates exceed 1, indicating the presence of S-shaped weighting in a subset of cases.

Parameter distributions. Panel A illustrates how the meta-analysis adjusts raw parameter estimates to infer latent, true effect sizes. The density of the estimated values, $\widehat{\gamma}_i$, shows a clear shift relative to the reported values, with increased concentration between 0.5 and 1.0 and diminished density outside this range. As before, this pattern reflects the meta-analytic pooling, which systematically downweights outliers. For likelihood-sensitivity in particular, pooling is notably strong, largely because extreme estimates tend to be measured with lower precision. As a result, the meta-analytic mean is 0.68, with a 95% CrI of $[0.66, 0.70]$, substantially lower than a simple average of the raw estimates would suggest. The narrow credible interval highlights the high precision of our meta-analytic estimate, driven both by the large number of observations for likelihood-sensitivity and by the fact that outliers tend to be especially noisy.

Gain-loss comparison. Panel B of Figure 6 compares the encoded γ estimates across outcome domains but reveals no consistent pattern. The mean and median values of γ^+ are 0.73 and 0.67, respectively, while for γ^-, both values are 0.71. To control for potential confounds, we turn to the meta-regression results in Table 5, which suggest weaker probability distortions in the loss domain than in the gain domain, as indicated by a positive coefficient. However, this effect is only marginally supported, with a 90% CrI and a posterior probability of 94.2% that the effect is positive. Finally, we analyze studies that report likelihood sensitivity parameter estimates for both outcome domains. First, we observe a significant correlation between likelihood sensitivity across domains ($r = 0.46$, $p < 0.01$). Panel C of Figure 6 shows that γ^- exceeds γ^+ in 136 cases, compared to just 50 cases where the reverse holds. A nonparametric test confirms that likelihood sensitivity is significantly higher for losses than for gains ($p < 0.01$). This highlights the value of isolating subsets of data where causal interpretations are more defensible: in this instance, the causally interpretable subset reveals an effect direction opposite to that observed in the aggregate analysis.

Functional forms. It is informative to explore whether the choice of probability weighting function specification influences the estimated likelihood sensitivity parameter. The meta-regression results, presented in Panel A of Figure 7, indicate that, relative to the baseline LLO form, the Prelec I specification is associated with significantly lower γ values, by approximately 0.08 ($[-0.13, -0.02]$). In contrast, Prelec II and TK yield slightly higher γ estimates, though these differences are not statistically significant. We emphasize that these associations should not be interpreted causally: the selection of functional form may be endogenous to the elicitation method, or to other experimental features, implying that causal claims would require conditioning on the data set used to produce the estimates.

Procedure invariance. As demonstrated earlier, utility curvature tends to be less pronounced when responses are elicited using CE choice lists and bisection methods compared to experiments with binary choices. Panel B of Figure 7 shows that estimates of likelihood sensitivity are also influenced by the measurement method. Specifically, likelihood sensitivity is significantly lower when elicited through CE choice lists and bisection methods compared to binary choices, but it is similar to binary choice when obtained through methods like direct matching. Overall, these findings indicate that methodological choices in preference elicitation can substantially impact the reported values of likelihood sensitivity. While our inference cannot be interpreted causally, Bouchouicha et al. (2024) report an experiment that provides evidence supporting a causal interpretation of the difference between binary choices and certainty equivalents.

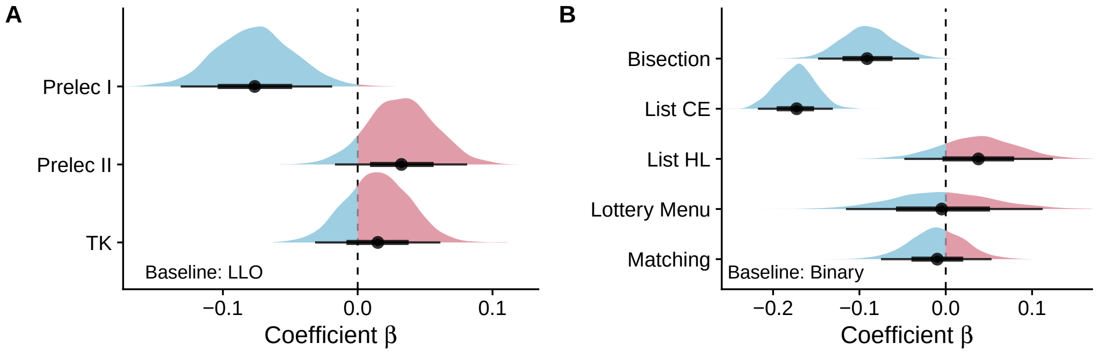

FIGURE 7: Posterior distributions of meta-regression coefficients. (A) Different PWF forms against LLO. (B) Different elicitation methods compared to binary choices. *Notes*: The posterior distributions of the Bayesian random-effects meta-regression coefficient(s) β, along with the posterior medians (represented by a black dot), 66% credible intervals (indicated by thick solid lines), and 95% credible intervals (shown as thin solid lines), are displayed.

TABLE 5: Meta-regression analysis of likelihood sensitivity.

Category	Variable	Median	2.5%	97.5%
Sign	Loss	0.03	-0.01	0.06
Utility	Expo	0.09	0.02	0.15
PWF	LLO		Baseline	
	Gul/Power			
	Prelec I	-0.08	-0.13	-0.02
	Prelec II	0.03	-0.02	0.08
	TK	0.02	-0.03	0.06
Elicitation	Binary Choice		Baseline	
	Bisection	-0.09	-0.15	-0.03
	List CE	-0.17	-0.22	-0.13
	List HL	0.04	-0.05	0.12
	Matching	-0.01	-0.08	0.05
	Lottery Menu	0.00	-0.11	0.11
Estimate	Aggregate		Baseline	
	Ind. Mean	0.04	-0.01	0.08
	Ind. Median	0.02	-0.03	0.06

Category	Variable	Median	2.5%	97.5%
Subject	Univ. pop.		Baseline	
	General	-0.04	-0.12	0.04
	Other	-0.04	-0.11	0.03
Reward	Real Money		Baseline	
	Hypo. Money	-0.11	-0.16	-0.07
	Other	-0.15	-0.23	-0.08
Data	Lab		Baseline	
	Class	0.00	-0.07	0.07
	Field	0.15	0.06	0.24
	Online	0.07	-0.02	0.15
Continent	Europe		Baseline	
	Africa	-0.09	-0.19	0.02
	Asia	-0.03	-0.08	0.02
	C/S-America	0.01	-0.07	0.09
	North America	0.09	0.04	0.14
	Oceania	0.10	-0.05	0.25
Frame	Visual aids	0.03	-0.01	0.08

Explained versus unexplained heterogeneity. Table 5 presents the meta-regression results. Despite the factors discussed above, it is important to note that only 21.5% of the heterogeneity in likelihood sensitivity is explained. This indicates that a significant portion of the variation in estimates remains unexplained, highlighting the complexity and diversity of the factors influencing γ. As with other parameters, the elicitation method emerges as the most significant predictor of variation across studies in our dataset, accounting for 35.8% of the explained variation.

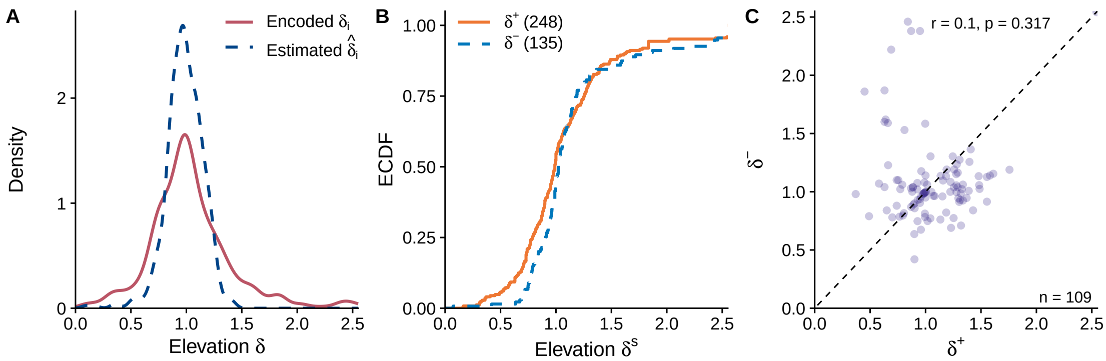

FIGURE 8: (A) Density plot comparing encoded δ_i values to their corresponding estimates $\widehat{\delta}_i$. (B) Empirical CDFs of the elevation parameter, separated by gain and loss domains. (C) Scatter plot of parameter estimates for studies reporting both gain and loss values. *Notes*: The x-axis is truncated for improved visualization. However, density estimations include all observations, including those beyond the displayed range.

5.3 Elevation δ

For the elevation parameter δ, we exclude 383 observations associated with the two one-parameter PWF specifications, Prelec I and TK. The Prelec I specification fixes δ at 1, while the TK function does not include this parameter. Additionally, due to the limited number of estimates for the Power and Gul functions and their similarity, we combine them for further heterogeneity analysis.

Descriptive statistics. We begin by examining the overall distribution of the elevation parameter. For gains, this parameter reflects optimism when $\delta > 1$ and pessimism when $\delta < 1$ (with the opposite interpretation for losses). As shown in Panel A of Figure 8 (solid red curve), the mean and median values are 1.14 and 1.00, respectively, with the IQR of $[0.83, 1.23]$. The distribution is skewed to the right due to a small number of exceptionally high estimates (26 values exceed 2, with the highest reaching 5.9).

Parameter distributions. Panel A compares the distribution of the raw data parameters to that of the estimated latent parameters. Due to shrinkage effects, the densities of the estimated values, $\widehat{\delta}_i$, exhibit a noticeable shift, with higher density concentrated around 1. The analysis shows that the mean of δ is 0.98, with a 95% CrI of $[0.95, 1.02]$, suggesting only a limited degree of elevation, if any.

Gain-loss comparison. Figure 8 compares encoded δ estimates across outcome domains. As shown in Panel B, the empirical CDFs for gains and losses are nearly

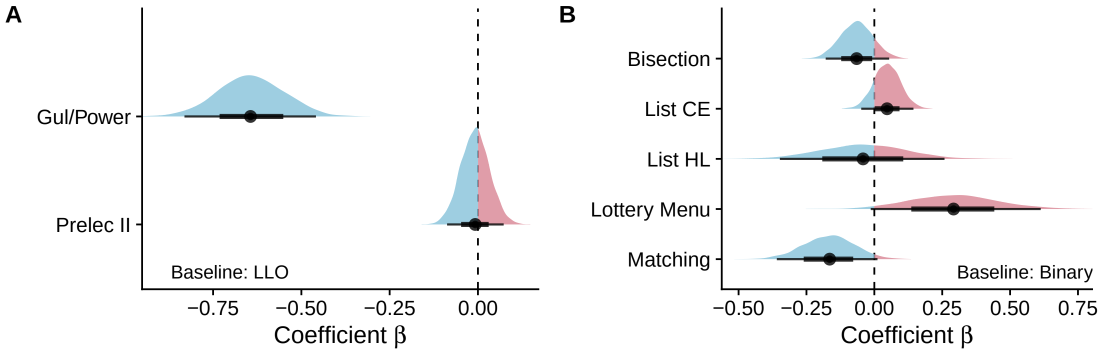

FIGURE 9: Posterior distributions of meta-regression coefficients. (A) Different PWF forms against LLO. (B) Various elicitation methods against Binary Choices. *Notes*: The posterior distributions of the Bayesian random-effects meta-regression coefficient(s) β, along with the posterior medians (represented by a black dot), 66% credible intervals (indicated by thick solid lines), and 95% credible intervals (shown as thin solid lines), are displayed.

indistinguishable. This is reflected in their respective central values: for δ^+, the mean and median are 1.13 and 0.99; for δ^-, they are 1.21 and 1.02. Our meta-regression analysis (Table 6) likewise reveals no evidence of a sign effect for δ, with the coefficient being both economically and statistically insignificant. We also examine studies that report elevation parameter estimates for both gain and loss domains. The pattern in Panel C reinforces our earlier findings: the mean of δ^+ is 1.09, while the mean of δ^- is 1.13 ($p = 0.34$), indicating no significant difference. Additionally, the correlation between gain and loss parameters is low and statistically insignificant ($r = 0.097$, $p = 0.317$).

Functional forms and procedure invariance. Panel A of Figure 9 shows that parameter estimates from the two most widely used two-parameter PWFs are statistically indistinguishable. In contrast, estimates based on the power or Gul weighting functions suggest stronger probability distortion (i.e., more pronounced depression). As shown in Panel B of Figure 9, our meta-regression results indicate that elevation parameters are unaffected by the elicitation method. Table 6 presents the full meta-regression results, including all covariates.

Explained versus unexplained heterogeneity. The proportion of explained heterogeneity for elevation is 20.9%, notably lower than that for utility curvature (approximately 50%), but comparable to that for likelihood sensitivity (21.8%). This finding highlights the complexity of the factors shaping individual probability perception, suggesting that a broad range of influences, beyond the study

TABLE 6: Meta-regression analysis of elevation.

Category	Variable	Median	2.5%	97.5%
Sign	Loss	0.02	-0.05	0.09
Utility	Expo	0.06	-0.04	0.17
PWF	LLO	Baseline		
	Gul/Power	-0.64	-0.83	-0.46
	Prelec I			
	Prelec II	-0.01	-0.09	0.07
	TK			
Elicitation	Binary Choice	Baseline		
	Bisection	-0.06	-0.18	0.05
	List CE	0.05	-0.04	0.15
	List HL	-0.05	-0.35	0.26
	Matching	-0.16	-0.34	0.03
	Lottery Menu	0.28	-0.04	0.62
Estimate	Aggregate	Baseline		
	Ind. Mean	0.04	-0.05	0.13
	Ind. Median	-0.01	-0.11	0.08

Category	Variable	Median	2.5%	97.5%
Subject	Univ. Pop.	Baseline		
	General	0.18	0.01	0.34
	Other	0.10	-0.03	0.23
Reward	Real Money	Baseline		
	Hypo. Money	-0.07	-0.16	0.03
	Other	0.15	-0.02	0.32
Data	Lab	Baseline		
	Class	0.06	-0.15	0.28
	Field	-0.20	-0.39	-0.01
	Online	-0.19	-0.41	0.04
Continent	Europe	Baseline		
	Africa	0.32	0.13	0.51
	Asia	0.05	-0.06	0.16
	C/S-America	0.08	-0.09	0.25
	North America	0.05	-0.05	0.16
	Oceania	0.03	-0.24	0.31
Frame	Visual Aids	0.06	-0.04	0.15

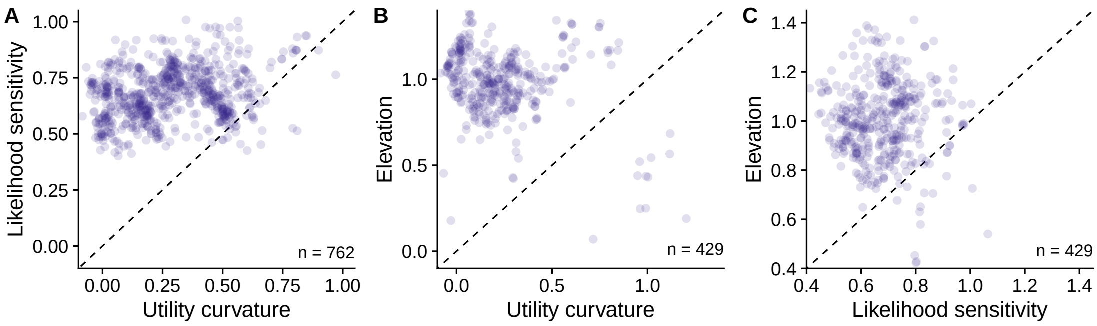

FIGURE 10: Scatter plots of the parameter estimates. Spearman's correlation coefficient: (A) $r = 0.216$, $p < 0.001$, (B) $r = -0.140$, $p = 0.006$, (C) $r = 0.066$, $p = 0.181$. *Notes*: The x- and y-axes are truncated for improved visualization.

characteristics considered here, may drive variation in both likelihood sensitivity and elevation.

5.4 Parameter Correlations

Figure 10 displays the correlations among the three PT parameters. Utility curvature (or outcome insensitivity) is positively correlated with likelihood sensitivity ($r = 0.22$, $p < 0.01$), but negatively correlated with elevation ($r = -0.14$, $p < 0.01$). No statistically significant correlation is found between the two weighting function parameters at conventional significance levels ($p = 0.18$).

Correlations between PT parameters have received relatively little attention in the literature. One potential source of such correlations is measurement and econometric noise. Specifically, when measurement error is present, parameters can be difficult to disentangle (Zeisberger, Vrecko and Langer, 2012); for instance, elevation and utility curvature may reflect overlapping motivational constructs, and utility curvature may be interdependent with likelihood sensitivity. Correlations may also arise from uncontrolled between-study heterogeneity or from substantive, structural relationships rooted in the underlying psychological or behavioral processes that drive the observed parameter values (Vieider, 2024*b*).

5.5 Testing for Publication Bias

Meta-analysis offers a powerful means to quantitatively synthesize findings from the literature. However, its inferences are only as reliable as the data fed into the model. A key challenge arises when certain types of results are more likely to be reported by authors, or published by editors, than others.[15] For example, the early focus on inverse-S shaped probability weighting may have discouraged the reporting or publication of findings suggesting S-shaped functions. Whether this occurred in the PT literature remains unclear; indeed, at some point, results deviating from standard findings may have become more publishable than yet another replication of common patterns. Moreover, many PT estimates appear in papers with different primary objectives, and the presence of multiple parameters makes it unclear which ones—if any—might be subject to publication bias (to wit, convex utility for gains and concave utility for losses are both fairly common in the literature). These considerations make it all the more essential to test for potential publication bias in the estimates included in our dataset.

Figure 11 presents funnel plots for the three PT parameters, a standard visual tool for assessing potential publication bias (Borenstein et al., 2009; Stanley and Doucouliagos, 2012). The plots include all complete observations with both parameter estimates and associated standard errors. In the absence of publication bias, the data points should be symmetrically distributed around the "true" effect size, indicated by the vertical solid line. Less precise studies (i.e., those with larger

[15] This phenomenon is often referred to as "publication bias" or the "file-drawer problem." In the context of PT parameters, two main sources of such bias are possible. First, journals may favor parameter estimates that align with canonical values, such as the utility curvature of 0.12 reported in Tversky and Kahneman (1992), and exhibit skepticism toward deviations from this benchmark. For bias to occur, the favored estimate must systematically diverge from the "true" value that would emerge from a broader population of studies (Borenstein et al., 2009). A second form of bias involves editorial preference for statistically significant results, often defined by a p-value below 0.05, which signals rejection of a null hypothesis (Andrews and Kasy, 2019; Brodeur, Cook and Heyes, 2020; Chopra et al., 2024).

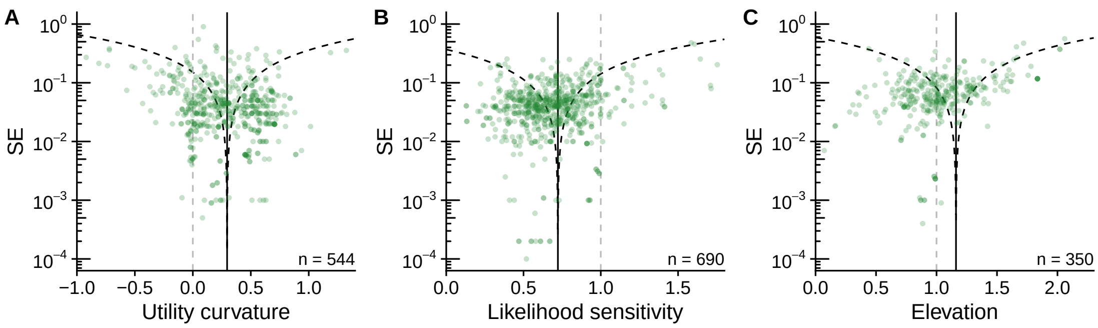

FIGURE 11: Relationship between reported parameter estimates and associated standard errors. *Notes*: Only estimates accompanied by standard errors are included. The solid vertical line represents the aggregate mean θ, while the dashed gray line indicates the neutrality benchmark ($x = 0$ or $x = 1$). The two dashed curves mark the boundaries for statistically significant deviations from the mean parameter value. The x-axis is truncated and the y-axis is displayed on a log scale for improved visualization.

standard errors) are expected to scatter more widely due to sampling variability but should still do so symmetrically around the true value. Asymmetry among these less precise estimates—appearing higher on the y-axis—is commonly interpreted as evidence of publication bias. In such cases, the average estimate may no longer reflect the true underlying effect size.

Egger's test reveals no significant funnel plot asymmetry for utility curvature ($p = 0.38$).[16] For likelihood sensitivity, the test similarly fails to detect significant asymmetry ($p = 0.16$). However, when retaining the top 1% of outliers, the test indicates a significant bias in favor of larger γ estimates. Such estimates are relatively common in our dataset and tend to be associated with larger standard errors, possibly due to specific measurement methods. To investigate this, we re-ran Egger's test while controlling for study-level characteristics. Although the magnitude and statistical significance of the asymmetry are reduced, the effect remains significant at conventional levels. Finally, we observe strong asymmetry in Panel C for the elevation parameter δ ($p < 0.01$). While selection based on optimism or pessimism is arguably less likely, the observed asymmetry may stem from the truncated distribution of estimates, i.e., the failure of the normality assumption implicit in both the funnel plot and Egger's test.

[16]Egger's test is a parametric method that formally assesses funnel plot asymmetry (Egger et al., 1997). It involves a weighted regression of effect size estimates on their precision (typically the inverse of the standard error or its logarithm). To enhance robustness, we trimmed the top 1% of outliers before conducting the analysis.

6 Discussion

Prospect theory has emerged as a remarkably successful framework for understanding behavior under risk, as evidenced by the 166 empirical papers included in this meta-analysis. Our study offers a rigorous quantitative synthesis of its core model parameters. On average, utility curvature estimates reveal diminishing sensitivity to increases in wealth (gains) and, to a somewhat lesser extent, to decreases in wealth (losses). The elevation parameter of the probability weighting function centers around neutral values, suggesting a general absence of optimism or pessimism. Importantly, the meta-analysis provides clear evidence of likelihood insensitivity—the tendency for relative risk aversion to increase systematically with the probability of winning or losing. These central tendencies strongly support the stylized behavioral patterns that originally motivated the development of prospect theory.

At the same time, our findings raise several challenges to prospect theory. Chief among them is the influence of the measurement or elicitation method, which emerged as the most significant predictor of variation in parameter estimates. Specifically, we observed notable differences in utility curvature and sensitivity parameters across methods, such as choice lists versus binary choices, bisection, and direct matching, as well as between choice lists varying outcomes versus probabilities. These inconsistencies violate the principle of procedure invariance, which prospect theory implicitly assumes—namely, that preference functionals should yield stable responses regardless of how choices are presented. However, these findings should be interpreted with caution: measurement methods may correlate with unobserved study characteristics that we did not code or control for, precluding causal inference. A more appropriate interpretation is that these results call for rigorously controlled experiments to identify underlying causal mechanisms.

Another mystery from our meta-analysis is the substantial unexplained heterogeneity in parameter estimates. Significant variability remains even after accounting for a wide range of study characteristics, including outcome domain, functional forms, measurement method, study population, and incentive structures. This suggests that PT parameter estimates may be sensitive to subtle experimental details. From the point of view of a preference-based model such as prospect theory, it would seem desirable to specifically investigate what might be driving such differences across studies. Elements such as the use of visual aids to represent probabilities, the size of the urn used to convey risk, or even the numerical scale of outcomes may help account for the observed variability.

Since prospect theory does not explicitly incorporate the effects of such contextual factors, a promising direction for future research lies in models that endogenize PT-like preference parameters. A growing body of work adopts this approach. Rather than seeing parameters governing choice processes as "preferences" (or at least as exogenous parameters) like prospect theory, these models typically depict observed choice regularities as an outgrowth of cognitive frictions affecting the decision process, and at least in some cases, optimal ways of dealing with such frictions. Several studies have attempted to explain decreasing sensitivity towards changes in wealth (e.g., Robson, 2001; Netzer, 2009; Khaw, Li and Woodford, 2021), probability weighting (e.g., Zhang and Maloney, 2012; Steiner and Stewart, 2016; Enke and Graeber, 2023; Herold and Netzer, 2023; Frydman and Jin, 2023; Oprea and Vieider, 2024), or both (e.g., Khaw, Li and Woodford, 2023; Vieider, 2024*b*). Some of these model furthermore make predictions that are specific to the choice context: while Vieider (2024*b*) presents a model of probability distortions in binary choice, the models proposed by Khaw, Li and Woodford (2023) and Bouchouicha et al. (2024) are specific to valuations or choice lists. We hope that the continued development and empirical testing of such models will help illuminate the more puzzling patterns in our findings, particularly the persistent heterogeneity across parameter estimates.

References

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

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review*, 109(8): 2766–2794.

Bernoulli, Daniel. 1738/1954. "Exposition of a New Theory on the Measurement." *Econometrica*, 22(1): 23–36.

Bleichrodt, Han, and Jose Luis Pinto. 2000. "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis." *Management Science*, 46(11): 1485–1496.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis.* John Wiley & Sons.

Bouchouicha, Ranoua, and Ferdinand M. Vieider. 2017. "Accommodating Stake Effects Under Prospect Theory." *Journal of Risk and Uncertainty*, 55(1): 1–28.

Bouchouicha, Ranoua, Ryan Oprea, Ferdinand M. Vieider, and Jilong Wu. 2024. “Is Prospect Theory Really a Theory of Choice?” Working paper.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. “Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics.” *American Economic Review*, 110(11): 3634–3660.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer. 2024. “Meta-analysis of Empirical Estimates of Loss-Aversion.” *Journal of Economic Literature*, 62(2): 485–516.

Camerer, Colin F., and Teck-Hua Ho. 1994. “Violations of the Betweenness Axiom and Nonlinearity in Probability.” *Journal of Risk and Uncertainty*, 8(2): 167–196.

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.

Cheung, Stephen L., and Lachlan Johnstone. 2017. “True Overconfidence, Revealed Through Actions: An Experiment.” IZA Discussion Paper No. 10545.

Chopra, Felix, Ingar Haaland, Christopher Roth, and Andreas Stegmann. 2024. “The Null Result Penalty.” *Economic Journal*, 134(657): 193–219.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. “Bias in Meta-Analysis Detected by a Simple, Graphical Test.” *BMJ*, 315(7109): 629–634.

Enke, Benjamin, and Thomas Graeber. 2023. “Cognitive Uncertainty.” *Quarterly Journal of Economics*, 138(4): 2021–2067.

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

Frydman, Cary, and Lawrence J. Jin. 2023. “On the Source and Instability of Probability Weighting.” SSRN. `http://dx.doi.org/10.2139/ssrn.4493313`.

Gabry, Jonah, Rok Češnovar, Andrew Johnson, and Steve Bronder. 2023. “`cmdstanr`: R Interface to CmdStan.” `https://mc-stan.org/cmdstanr`.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis.* Vol. 3, Boca Raton, FL:CRC Press.

Goldstein, William M., and Hillel J. Einhorn. 1987. “Expression Theory and the Preference Reversal Phenomena." *Psychological Review*, 94(2): 236–254.

Gonzalez, Richard, and George Wu. 1999. "On the Shape of the Probability Weighting Function." *Cognitive Psychology*, 38(1): 129–166.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica*, 59(3): 667–686.

Herold, Florian, and Nick Netzer. 2023. "Second-Best Probability Weighting." *Games and Economic Behavior*, 138: 112–125.

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): 263–292.

Karmarkar, Uday S. 1978. "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model." *Organizational Behavior and Human Performance*, 21(1): 61–72.

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." NBER Working Paper No. 30417.

Köbberling, Veronika, and Peter P. Wakker. 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119–131.

L'Haridon, Olivier, and Ferdinand M. Vieider. 2019. "All over the Map: A Worldwide Comparison of Risk Preferences." *Quantitative Economics*, 10(1): 185–215.

Netzer, Nick. 2009. "Evolution of Time Preferences and Attitudes Toward Risk." *American Economic Review*, 99(3): 937–955.

Oprea, Ryan. 2024. "Decisions Under Risk Are Decisions Under Complexity." *American Economic Review*, 112(4): 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.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica*, 66(3): 497–527.

Preston, Malcolm G., and Philip Baratta. 1948. "An Experimental Study of the Auction-Value of an Uncertain Outcome." *American Journal of Psychology*, 61(2): 183–193.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rieger, Marc Oliver, Mei Wang, and Thorsten Hens. 2017. "Estimating Cumulative Prospect Theory Parameters from an International Survey." *Theory and Decision*, 82(4): 567–596.

Robson, Arthur J. 2001. "Why Would Nature Give Individuals Utility Functions?" *Journal of Political Economy*, 109(4): 900–914.

Schoemaker, Paul J. H. 1990. "Are Risk-Attitudes Related across Domains and Response Modes?" *Management Science*, 36(12): 1451–1463.

Stanley, Tom D., and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business.* London:Routledge.

Steiner, Jakub, and Colin Stewart. 2016. "Perceiving Prospects Properly." *American Economic Review*, 106(7): 1601–31.

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*a*. "Bayesian Estimation of Decision Models." https://fvieider.quarto.pub/bstats.

Vieider, Ferdinand M. 2024*b*. "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:Edward Elgar Publishing.

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

Wulff, Dirk U., Max Mergenthaler-Canseco, and Ralph Hertwig. 2018. "A Meta-Analytic Review of Two Modes of Learning and the Description-Experience Gap." *Psychological Bulletin*, 144(2): 140–176.

Zeisberger, Stefan, Dennis Vrecko, and Thomas Langer. 2012. "Measuring the Time Stability of Prospect Theory Preferences." *Theory and Decision*, 72(3): 359–386.

Zhang, Hang, and Laurence T. Maloney. 2012. "Ubiquitous Log Odds: A Common Representation of Probability and Frequency Distortion in Perception, Action, and Cognition." *Frontiers in Neuroscience*, 6: 1.

Online Appendix

Meta-Analysis of Risk-Taking Propensities

Taisuke Imai   Salvatore Nunnari   Jilong Wu   Ferdinand M. Vieider

Contents

A Data **1**

A.1 Paper Search and Screening 1

A.2 Approximation and Imputation of Missing Standard Errors . . 2

A.3 Coded Variables 3

A.4 Collected Studies for Meta-Analysis 5

A.5 Journals 19

A.6 Global Map 21

B Standard Error Imputation **22**

B.1 Balance Check 24

B.2 Robustness Check 25

A Data

A.1 Paper Search and Screening

We conducted a search for pertinent literature within the Web of Science, a scientific citation indexing database. After several rounds of trial and error to refine our approach, we settled on the following combination of query terms for our search.

```
(    "prospect theory"
 OR  "probability weighting" OR "probability distortion"
 OR  ("risk preference" AND ("risk attitudes" OR "ambiguity attitudes")) )
                               AND
           (estimat* OR measur* OR experiment* OR survey)
```

FIGURE A.1: Keywords used in the search.

The initial search, conducted in the spring of 2023, yielded 2,034 papers. In the first phase of paper identification, we reviewed the titles and abstracts, eliminating 1,453 papers that were evidently not relevant to our study. Subsequently, we thoroughly examined the remaining papers, applied our inclusion criteria focusing on content, and proceeded to code the information. Additionally, we utilized IDEAS/RePEc and Google Scholar to locate unpublished working papers.

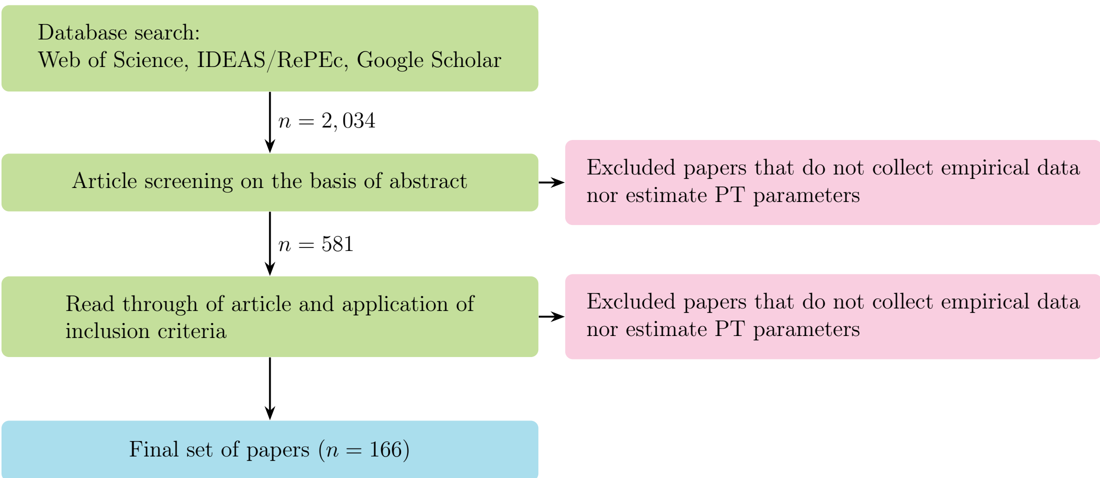

FIGURE A.2: Paper search and data construction.

A.2 Approximation and Imputation of Missing Standard Errors

There are different standard error information sources related to the parameter estimates. We explain below how we calculate SEs using different sources.

- 387 estimates are reported with exact SEs values in the paper.
- 170 estimates are accompanied by standard deviations. We derived $se = SD/\sqrt{n}$.
- 32 SEs were derived from 95% confidence intervals $[lb, ub]$ as $se = (ub - lb)/3.92$.
- 12 SEs are calculated from the effect size (ES) and t-value: $se = |ES|/t$.
- Interquartile ranges are reported for 52 estimates, for which we approximated SEs using $se \approx IQR/(1.35 \times \sqrt{n})$.
- 95% credible intervals or 95% highest density intervals are reported for 31 estimates. We treated these intervals as confidence intervals and followed the above-mentioned approach, $se \approx (ub - lb)/3.92$.
- 12 observations have provided p-value, from which we calculated SEs according to $se \approx |ES|/\Phi^{-1}(1 - p)$, where Φ^{-1} is the quantile function of the standard normal distribution.
- For 3 estimates that provide the maximal and minimal values, $se \approx (Max - Min)/(4 \times \sqrt{n})$.

Note that these approximation formulas are deemed valid when the parameters exhibit a normal distribution within the population. We acknowledge this is a strong assumption for our dataset. Despite this assumption, opting for an "approximated" standard error was considered preferable to discarding the observation altogether or resorting to alternative, potentially stronger assumptions to retain the observation.

A.3 Coded Variables

Table A.1: List of coded variables.

Variable	Description
Article meta data	
title	Title of the paper
author_lastnames	Last names of the authors
author_firstnames	First names of the authors
published	1 = published paper
journal	Journal
year	Year of publication
num_subject	Number of subjects
num_choice	Number of choices each subject made
num_list	Number of choices list each subject made
Estimates	
res_est_u1	Reported utility function curvature
res_est_u2	Reported extra utility function curvature
res_est_pwf_alpha	Reported PWF parameter γ
res_est_pwf_beta	Reported PWF parameter δ
res_est_la	Reported loss aversion coefficient
res_err_u1	SE of utility function curvature
res_err_u2	SE of extra utility function curvature
res_err_pwf_alpha	SE of PWF parameter γ
res_err_pwf_beta	SE of PWF parameter δ
res_err_la	SE of loss aversion coefficient
Model features	
u_form	Utility function specification adopted
u_common_gain_loss	1 = common u is assumed for gains and losses
pwf_form	PWF specification adopted
pwf_common_gain_loss	1 = common PWF is assumed for gains and losses
pwf_num_parameters	Number of parameters of the PWF
Type of data	
exp_lab	1 = the data is collected in lab
exp_field	1 = the data is collected in field
exp_class	1 = the data is collected in classroom

Continued on next page.

Variable	Description
`exp_online`	1 = the data is collected online or via survey
Type of elicitation	
`choice_bisection`	1 = the bisection setup is used
`choice_binary`	1 = the binary choice setup is used
`choice_list`	1 = the choice list setup is used
`choice_matching`	1 = the direct matching is used
`choice_menu`	1 = the lottery menu is used
Level of measurement	
`est_aggregate`	1 = aggregate level estimate
`est_aggregate_median_data`	1 = aggregate level ("median subject") estimate
`est_individual`	1 = individual level estimate
`est_mixture`	1 = mixture model estimation
Subject pool	
`subject_uni`	1 = university subjects are recruited
`subject_general`	1 = general subjects are recruited
`subject_other`	1 = special subjects are recruited
`subject_other_type`	specify the population when `subject_other` = 1
Reward type	
`reward_money`	1 = real monetary reward
`reward_hypothetical`	1 = hypothetical monetary choices
`reward_other`	1 = non monetary reward (other)
Location of the experiment/survey	
`location_country`	Country location of the experiment
`location_continent`	Continent location of the experiment
Estimation Strategy	
`est_strategy`	Description of estimation strategy
`est_loss`	1 = parameter estimate for loss domain
Estimation Strategy	
`interface_visual_aids`	1 = lotteries represented with visual aids
`interface_iconic_express`	1 = lotteries represented with iconic visual aids

A.4 Collected Studies for Meta-Analysis

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

Abdellaoui, Mohammed, Ahmed Driouchi, and Olivier L'Haridon. 2011. "Risk Aversion Elicitation: Reconciling Tractability and Bias Minimization." *Theory and Decision*, 71(1): 63–80.

Abdellaoui, Mohammed, and Emmanuel Kemel. 2014. "Eliciting Prospect Theory When Consequences Are Measured in Time Units: "Time Is Not Money"." *Management Science*, 60(7): 1844–1859.

Abdellaoui, Mohammed, Aurélien Baillon, Lætitia Placido, and Peter P. Wakker. 2011. "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation." *American Economic Review*, 101(2): 695–723.

Abdellaoui, Mohammed, Emmanuel Kemel, Amma Panin, and Ferdinand M. Vieider. 2019. "Measuring Time and Risk Preferences in an Integrated Framework." *Games and Economic Behavior*, 115: 459–469.

Abdellaoui, Mohammed, Enrico Diecidue, and Ayse Öncüler. 2011. "Risk Preferences at Different Time Periods: An Experimental Investigation." *Management Science*, 57(5): 975–987.

Abdellaoui, Mohammed, Enrico Diecidue, Emmanuel Kemel, and Ayse Öncüler. 2022. "Temporal Risk: Utility vs. Probability Weighting." *Management Science*, 68(7): 5162–5186.

Abdellaoui, Mohammed, Olivier L'Haridon, and Corina Paraschiv. 2011. "Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications?" *Management Science*, 57(10): 1879–1895.

Abdellaoui, Mohammed, Olivier L'Haridon, and Corina Paraschiv. 2013. "Individual vs. Couple Behavior: An Experimental Investigation of Risk Preferences." *Theory and Decision*, 75(2): 175–191.

Ackert, Lucy F., Richard Deaves, Jennifer Miele, and Quang Nguyen. 2020. "Are Time Preference and Risk Preference Associated with Cognitive Intelligence and Emotional Intelligence?" *Journal of Behavioral Finance*, 21(2): 136–156.

Alam, Jessica, Konstantinos Georgalos, and Harrison Rolls. 2022. "Risk Preferences, Gender Effects and Bayesian Econometrics." *Journal of Economic Behavior & Organization*, 202: 168–183.

Andersen, Steffen, Amalia Di Girolamo, Glenn W. Harrison, and

Morten I. Lau. 2014. “Risk and Time Preferences of Entrepreneurs: Evidence from a Danish Field Experiment.” *Theory and Decision*, 77(3): 341–357.

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(2): 537–555.

Andreoni, James, and Charles Sprenger. 2011. “Uncertainty Equivalents: Testing the Limits of the Independence Axiom.” NBER Working Paper No. 17342.

Åstebro, Thomas, José Mata, and Luís Santos-Pinto. 2015. “Skewness Seeking: Risk Loving, Optimism or Overweighting of Small Probabilities?” *Theory and Decision*, 78(2): 189–208.

Attema, Arthur E., Han Bleichrodt, and Olivier L'Haridon. 2018. “Ambiguity Preferences for Health.” *Health Economics*, 27(11): 1699–1716.

Attema, Arthur E., Lisheng He, Alasdair J. C. Cook, and Victor J. Del Rio Vilas. 2019. “Unbiased Assessment of Disease Surveillance Utilities: A Prospect Theory Application.” *PLOS Neglected Tropical Diseases*, 13(5): e0007364.

Aydogan, Ilke. 2021. “Prior Beliefs and Ambiguity Attitudes in Decision from Experience.” *Management Science*, 67(11): 6934–6945.

Aydogan, Ilke, and Yu Gao. 2020. “Experience and Rationality Under Risk: Re-Examining the Impact of Sampling Experience.” *Experimental Economics*, 23(4): 1100–1128.

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(1): 93–112.

Baillon, Aurélien, Laure Cabantous, and Peter P. Wakker. 2012. “Aggregating Imprecise or Conflicting Beliefs: An Experimental Investigation Using Modern Ambiguity Theories.” *Journal of Risk and Uncertainty*, 44(2): 115–147.

Baláž, Vladimír, Viera Bacova, Eva Drobna, Katarina Dudekova, and Kamil Adamik. 2013. “Testing Prospect Theory Parameters.” *Ekonomický časopis*, 61(7): 655–671.

Barreda-Tarrazona, Iván, Ainhoa Jaramillo-Gutiérrez, Daniel Navarro-Martínez, and Gerardo Sabater-Grande. 2011. “Risk Attitude Elicitation Using a Multi-Lottery Choice Task: Real vs. Hypothetical Incentives.” *Spanish Journal of Finance and Accounting*, 40(152): 613–628.

Bauermeister, Golo Friedrich, Daniel Hermann, and Oliver Musshoff. 2018. "Consistency of Determined Risk Attitudes and Probability Weightings Across Different Elicitation Methods." *Theory and Decision*, 84(4): 627–644.

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(4): 1069–1099.

Bernedo Del Carpio, María, 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.

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica*, 88(4): 1363–1409.

Birchby, Jeff, Gary Gigliotti, and Barry Sopher. 2011. "Consistency and Aggregation in Individual Choice Under Uncertainty." Working paper.

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(2): 161–194.

Bleichrodt, Han, and Jose Luis Pinto. 2000. "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis." *Management Science*, 46(11): 1485–1496.

Bleichrodt, Han, Olivier L'Haridon, and David Van Ass. 2018. "The Risk Attitudes of Professional Athletes: Optimism and Success Are Related." *Decision*, 5(2): 95–118.

Bocquého, Geraldine, Julien Jacob, and Marielle Brunette. 2023. "Prospect Theory in Multiple Price List Experiments: Further Insights on Behaviour in the Loss Domain." *Theory and Decision*, 94(4): 593–636.

Booij, Adam S., Bernard M. S. Van Praag, and Gijs Van De Kuilen. 2010. "A Parametric Analysis of Prospect Theory's Functionals for the General Population." *Theory and Decision*, 68(1-2): 115–148.

Bouchouicha, Ranoua, and Ferdinand M. Vieider. 2017. "Accommodating Stake Effects Under Prospect Theory." *Journal of Risk and Uncertainty*, 55(1): 1–28.

Bougherara, Douadia, Lana Friesen, and Céline Nauges. 2021. "Risk Taking with Left- and Right-Skewed Lotteries." *Journal of Risk and Uncertainty*, 62(1): 89–112.

Bougherara, Douadia, Xavier Gassmann, Laurent Piet, and Arnaud Reynaud. 2017. "Structural Estimation of Farmers' Risk and Ambiguity Preferences: A Field Experiment." *European Review of Agricultural Economics*, 44(5): 782–808.

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(2): 79–100.

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

Bruhin, Adrian, Luís Santos-Pinto, and David Staubli. 2018. "How Do Beliefs About Skill Affect Risky Decisions?" *Journal of Economic Behavior & Organization*, 150: 350–371.

Bruner, David M. 2009. "Changing the Probability Versus Changing the Reward." *Experimental Economics*, 12: 367–385.

Cabantous, Laure, and Cedric Ditzler. 2004. "Is Probability Weighting Sensitive to the Quality of Probabilistic Information? An Experimental Investigation on the Impact of Vagueness and Conflict on Probabilistic Weighting Functions." Working paper.

Campos-Vazquez, Raymundo M., and Emilio Cuilty. 2014. "The Role of Emotions on Risk Aversion: A Prospect Theory Experiment." *Journal of Behavioral and Experimental Economics*, 50: 1–9.

Capra, C. Monica, Bing Jiang, Jan B. Engelmann, and Gregory S. Berns. 2013. "Can Personality Type Explain Heterogeneity in Probability Distortions?" *Journal of Neuroscience, Psychology, and Economics*, 6(3): 151–166.

Charupat, Narat, Richard Deaves, Travis Derouin, Marcelo Klotzle, and Peter Miu. 2013. "Emotional Balance and Probability Weighting." *Theory and Decision*, 75(1): 17–41.

Cheung, Stephen L. 2020. "Eliciting Utility Curvature in Time Preference." *Experimental Economics*, 23(2): 493–525.

Cheung, Stephen L., and Lachlan Johnstone. 2017. "True Overconfidence, Revealed Through Actions: An Experiment." IZA Discussion Paper No. 10545.

Choi, Syngjoo, Jeongbin Kim, Eungik Lee, and Jungmin Lee. 2022. "Probability Weighting and Cognitive Ability." *Management Science*, 68(7): 5201–5215.

Clarke, Daniel, and Gautam Kalani. 2011. "Microinsurance Decisions: Evidence from Ethiopia." Working paper.

Conte, Anna, John D. Hey, and Peter G. Moffatt. 2011. "Mixture Models of Choice Under Risk." *Journal of Econometrics*, 162(1): 79–88.

Conte, Anna, M. Vittoria Levati, and Chiara Nardi. 2018. "Risk Preferences and the Role of Emotions." *Economica*, 85(338): 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(2): 193–210.

Cubitt, Robin, Orestis Kopsacheilis, and Chris Starmer. 2022. "An Inquiry into the Nature and Causes of the Description–Experience Gap." *Journal of Risk and Uncertainty*, 65(2): 105–137.

de Brauw, Alan, and Patrick Eozenou. 2014. "Measuring Risk Attitudes Among Mozambican Farmers." *Journal of Development Economics*, 111: 61–74.

Donkers, Bas, Bertrand Melenberg, and Arthur Van Soest. 2001. "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *The Journal of Risk and Uncertainty*, 22(2): 165–195.

Drichoutis, Andreas C., and Jayson L. Lusk. 2014. "Judging Statistical Models of Individual Decision Making Under Risk Using in- and Out-of-Sample Criteria." *PLOS ONE*, 9(7): e102269.

Engelmann, Jan B., Sara Moore, C. Monica Capra, and Gregory S. Berns. 2012. "Differential Neurobiological Effects of Expert Advice on Risky Choice in Adolescents and Adults." *Social Cognitive and Affective Neuroscience*, 7(5): 557–567.

Enke, Benjamin, and Cassidy Shubatt. 2023. "Quantifying Lottery Choice Complexity." NBER Working Paper No. 31677.

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(3): 169–203.

Erev, Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan M. Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West, and Christian Lebiere. 2010. "A Choice Prediction Competition: Choices from Experience and from Description." *Journal of Behavioral Decision Making*, 23(1): 15–47.

Erner, Carsten, Alexander Klos, and Thomas Langer. 2013. "Can Prospect Theory Be Used to Predict an Investor's Willingness to Pay?" *Journal of Bank-*

ing & Finance, 37(6): 1960–1973.

Etchart-Vincent, Nathalie. 2004. "Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses." *Journal of Risk and Uncertainty*, 28(3): 217–235.

Etchart-Vincent, Nathalie. 2009. "Probability Weighting and the 'Level' and 'Spacing' of Outcomes: An Experimental Study over Losses." *Journal of Risk and Uncertainty*, 39(1): 45–63.

Fan, Yuyu, David V. Budescu, and Enrico Diecidue. 2019. "Decisions with Compound Lotteries." *Decision*, 6(2): 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(2): 147–180.

Fehr-Duda, Helga, Manuele De Gennaro, and Renate Schubert. 2006. "Gender, Financial Risk, and Probability Weights." *Theory and Decision*, 60(2-3): 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 & Organization*, 78(1-2): 14–24.

Freudenreich, Hanna, and Oliver Musshoff. 2022. "Experience of Losses and Aversion to Uncertainty – Experimental Evidence from Farmers in Mexico." *Ecological Economics*, 195: 107379.

Galarza, Francisco B. 2009. "Choices Under Risk in Rural Peru." Working paper.

Gao, Xiaoxue Sherry, Glenn W. Harrison, and Rusty Tchernis. 2023. "Behavioral Welfare Economics and Risk Preferences: A Bayesian Approach." *Experimental Economics*, 26(2): 273–303.

Gawryluk, Katarzyna, and Michal Krawczyk. 2019. "Additional Deliberation Reduces Pessimism: Evidence from the Double-Response Method." *Journal of the Economic Science Association*, 5(1): 51–64.

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(4): 1533–1562.

Glöckner, Andreas, and Thorsten Pachur. 2012. "Cognitive Models of Risky Choice: Parameter Stability and Predictive Accuracy of Prospect Theory." *Cognition*, 123(1): 21–32.

Glöckner, Andreas, Baiba Renerte, and Ulrich Schmidt. 2020. "Violations of Coalescing in Parametric Utility Measurement." *Theory and Decision*, 89(4): 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(4): 486–508.

González-Jiménez, Víctor. 2022. "Incentive Contracts When Agents Distort Probabilities." Working paper.

Gonzalez, Richard, and George Wu. 1999. "On the Shape of the Probability Weighting Function." *Cognitive Psychology*, 38(1): 129–166.

Gonzalez, Richard, and George Wu. 2022. "Composition Rules in Original and Cumulative Prospect Theory." *Theory and Decision*, 92(3-4): 647–675.

Goyal, Shruti, and Krishna P. Miyapuram. 2019. "Feedback Influences Discriminability and Attractiveness Components of Probability Weighting in Descriptive Choice Under Risk." *Frontiers in Psychology*, 10: 962.

Güth, Werner, Matteo Ploner, and Ivan Soraperra. 2016. "Buying and Selling Risk: An Experiment Investigating Evaluation Asymmetries." *Review of Behavioral Economics*, 3(3-4): 389–419.

Haffke, Peter, and Ronald Hübner. 2014. "Effects of Different Feedback Types on Information Integration in Repeated Monetary Gambles." *Frontiers in Psychology*, 5: 1597.

Hajimoladarvish, Narges. 2017. "Very Low Probabilities in the Loss Domain." *Geneva Risk and Insurance Review*, 42(1): 41–58.

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(4): 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(1): 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(2): 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(3): 552–568.

Hey, John D., Andrea Morone, and Ulrich Schmidt. 2009. "Noise and Bias in Eliciting Preferences." *Journal of Risk and Uncertainty*, 39(3): 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(7): 2231–2237.

Kairies-Schwarz, Nadja, Johanna Kokot, Markus Vomhof, and Jens Weßling. 2017. "Health Insurance Choice and Risk Preferences Under Cumulative Prospect Theory – An Experiment." *Journal of Economic Behavior & Organization*, 137: 374–397.

Kean-Siang, Ch'ng, Suresh Narayanan, and Tan Scaik Lin. 2021. "Gender Differences in Weighing Probability and Payoffs in Risky Prospects: Experimental Evidence from Malaysia." *New Zealand Economic Papers*, 55(3): 263–276.

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: 101258.

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.

Kemel, Emmanuel, and Corina Paraschiv. 2018. "Deciding About Human Lives: An Experimental Measure of Risk Attitudes Under Prospect Theory." *Social Choice and Welfare*, 51(1): 163–192.

Kemel, Emmanuel, and Corina Paraschiv. 2023. "Risking the Future? Measuring Risk Attitudes Towards Delayed Consequences." *Journal of Economic Behavior & Organization*, 208: 325–344.

Kemel, Emmanuel, Antoine Nebout, and Bruno Ventelou. 2020. "To Test or Not to Test? Risk Attitudes and Prescribing by French GPs." Working paper.

Königsheim, Christian, Moritz Lukas, and Markus Nöth. 2019. "Salience Theory: Calibration and Heterogeneity in Probability Distortion." *Journal of Economic Behavior & Organization*, 157: 477–495.

Kontek, Krzysztof, and Michael H. Birnbaum. 2019. "The Impact of Middle Outcomes on Lottery Valuations." *Journal of Behavioral and Experimental*

Economics, 78: 30–44.

Kopsacheilis, Orestis. 2018. "The Role of Information Search and Its Influence on Risk Preferences." *Theory and Decision*, 84(3): 311–339.

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(6): 1487–1505.

Lampe, Immanuel, and Matthias Weber. 2021. "Intertemporal Prospect Theory." SSRN. http://dx.doi.org/10.2139/ssrn.3849330.

Laury, Susan K., Melayne Morgan McInnes, and J. Todd Swarthout. 2012. "Avoiding the Curves: Direct Elicitation of Time Preferences." *Journal of Risk and Uncertainty*, 44(3): 181–217.

Lejarraga, Tomás, Thorsten Pachur, Renato Frey, and Ralph Hertwig. 2016. "Decisions from Experience: From Monetary to Medical Gambles." *Journal of Behavioral Decision Making*, 29(1): 67–77.

Liebenehm, Sabine, and Hermann Waibel. 2014. "Simultaneous Estimation of Risk and Time Preferences Among Small-Scale Cattle Farmers in West Africa." *American Journal of Agricultural Economics*, 96(5): 1420–1438.

Ligneul, Romain, Guillaume Sescousse, Guillaume Barbalat, Philippe Domenech, and Jean-Claude Dreher. 2013. "Shifted Risk Preferences in Pathological Gambling." *Psychological Medicine*, 43(5): 1059–1068.

Lipman, Stefan A., Werner B. F. Brouwer, and Arthur E. Attema. 2019. "QALYs Without Bias? Nonparametric Correction of Time Trade-Off and Standard Gamble Weights Based on Prospect Theory." *Health Economics*, 28(7): 843–854.

Liu, Elaine M., and Ji Kun Huang. 2013. "Risk Preferences and Pesticide Use by Cotton Farmers in China." *Journal of Development Economics*, 103(1): 202–215.

Lobel, Robert Eugene, Marcelo Cabus Klotzle, Paulo Vitor Jordao Da Gama Silva, and Antonio Carlos Figueiredo Pinto. 2017. "Prospect Theory: A Parametric Analysis of Functional Forms in Brazil." *Revista de Administração de Empresas*, 57(5): 495–509.

l'Haridon, Olivier, and Ferdinand M. Vieider. 2019. "All over the Map: A Worldwide Comparison of Risk Preferences." *Quantitative Economics*, 10(1): 185–215.

Maafi, Hela. 2011. "Preference Reversals Under Ambiguity." *Management Sci-*

ence, 57(11): 2054–2066.

Millroth, Philip, Håkan Nilsson, and Peter Juslin. 2018. "Examining the Integrity of Evaluations of Risky Prospects Using a Single-Stimuli Design." *Decision*, 5(4): 362–377.

Murad, Zahra, Martin Sefton, and Chris Starmer. 2016. "How Do Risk Attitudes Affect Measured Confidence?" *Journal of Risk and Uncertainty*, 52(1): 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(1): 308–326.

Newell, Anthony. 2020. "Is Your Heart Weighing Down Your Prospects? Interoception, Risk Literacy and Prospect Theory." Working paper.

Nguyen, Quang, Marie Claire Villeval, and Hui Xu. 2016. "Trust Under the Prospect Theory and Quasi-Hyperbolic Preferences: A Field Experiment in Vietnam." *Economic Development and Cultural Change*, 64(3): 545–572.

Nilsson, Håkan, Jörg Rieskamp, and Eric-Jan Wagenmakers. 2011. "Hierarchical Bayesian Parameter Estimation for Cumulative Prospect Theory." *Journal of Mathematical Psychology*, 55(1): 84–93.

Ojala, Karita E., Lieneke K. Janssen, Mahur M. Hashemi, Monique H. M. Timmer, Dirk E. M. Geurts, Niels P. Ter Huurne, Roshan Cools, and Guillaume Sescousse. 2018. "Dopaminergic Drug Effects on Probability Weighting During Risky Decision Making." *eNeuro*, 5(2).

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*, 2: e475.

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(2): 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(5): 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(1): 16115.

Patalano, Andrea L., Alexandra Zax, Katherine Williams, Liana Math-

ias, Sara Cordes, and Hilary Barth. 2020. “Intuitive Symbolic Magnitude Judgments and Decision Making Under Risk in Adults.” *Cognitive Psychology*, 118: 101273.

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(6): 1820–1829.

Qiu, Jianying, and Eva-Maria Steiger. 2011. “Understanding the Two Components of Risk Attitudes: An Experimental Analysis.” *Management Science*, 57(1): 193–199.

Razo, Michael, and Song Gao. 2013. “A Rank-Dependent Expected Utility Model for Strategic Route Choice with Stated Preference Data.” *Transportation Research Part C: Emerging Technologies*, 27: 117–130.

Reynaud, Arnaud, and Cécile Aubert. 2020. “Does Flood Experience Modify Risk Preferences? Evidence from an Artefactual Field Experiment in Vietnam.” *Geneva Risk and Insurance Review*, 45(1): 36–74.

Riddel, Mary, and Sonja Kolstoe. 2013. “Heterogeneity in Life-Duration Preferences: Are Risky Recreationists Really More Risk Loving?” *Journal of Risk and Uncertainty*, 46(2): 191–213.

Rieger, Marc Oliver, Mei Wang, and Thorsten Hens. 2017. “Estimating Cumulative Prospect Theory Parameters from an International Survey.” *Theory and Decision*, 82(4): 567–596.

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(8): 1241–1255.

Robinson, Peter John, and W. J. Wouter Botzen. 2020. “Flood Insurance Demand and Probability Weighting: The Influences of Regret, Worry, Locus of Control and the Threshold of Concern Heuristic.” *Water Resources and Economics*, 30: 100144.

Sagemüller, Frederik, and Oliver Mußhoff. 2020. “Effects of Household Shocks on Risk Preferences and Loss Aversion: Evidence from Upland Smallholders of South East Asia.” *Journal of Development Studies*, 56(11): 2061–2078.

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(2): 391–407.

Schley, Dan, Alina Ferecatu, Hang-Yee Chan, and Manissa Gunadi. 2023. "How Categorization Shapes the Probability Weighting Function." SSRN. http://dx.doi.org/10.2139/ssrn.3959751.

Scholten, Marc, and Daniel Read. 2014. "Prospect Theory and the "Forgotten" Fourfold Pattern of Risk Preferences." *Journal of Risk and Uncertainty*, 48(1): 67–83.

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: 981.

Schwanen, Tim, and Dick Ettema. 2009. "Coping with Unreliable Transportation When Collecting Children: Examining Parents' Behavior with Cumulative Prospect Theory." *Transportation Research Part A: Policy and Practice*, 43(5): 511–525.

Somasundaram, Jeeva, and Vincent Eli. 2022. "Risk and Time Preferences Interaction: An Experimental Measurement." *Journal of Risk and Uncertainty*, 65(2): 215–238.

Spiliopoulos, Leonidas, and Ralph Hertwig. 2019. "Nonlinear Decision Weights or Moment-Based Preferences? A Model Competition Involving Described and Experienced Skewness." *Cognition*, 183: 99–123.

Spitmaan, Mehran, Emily Chu, and Alireza Soltani. 2019. "Salience-Driven Value Construction for Adaptive Choice Under Risk." *Journal of Neuroscience*, 39(26): 5195–5209.

Stolk-Vos, Aline C., Arthur E. Attema, Michele Manzulli, and Joris J. van de Klundert. 2022. "Do Patients and Other Stakeholders Value Health Service Quality Equally? A Prospect Theory Based Choice Experiment in Cataract Care." *Social Science & Medicine*, 294: 114730.

Stott, Henry P. 2006. "Cumulative Prospect Theory's Functional Menagerie." *Journal of Risk and Uncertainty*, 32(2): 101–130.

Sun, Qingzhou, Evan Polman, and Huanren Zhang. 2021. "On Prospect Theory, Making Choices for Others, and the Affective Psychology of Risk." *Journal of Experimental Social Psychology*, 96: 104177.

Sun, Qingzhou, Huanren Zhang, Liyang Sai, and Fengpei Hu. 2018. "Self-Distancing Reduces Probability-Weighting Biases." *Frontiers in Psychology*, 9: 611.

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(4): 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(4): e0122475.

Suzuki, Ayako, and Koichi Kume. 2008. "Aging, Probability Weighting, and Reference Point Adoption: An Experimental Study." ISER Discussion Paper No. 720.

Takahashi, Hidehiko, Hiroshi Matsui, Colin Camerer, Harumasa Takano, Fumitoshi Kodaka, Takashi Ideno, Shigetaka Okubo, Kazuhisa Takemura, Ryosuke Arakawa, Yoko Eguchi, et al. 2010. "Dopamine D1 Receptors and Nonlinear Probability Weighting in Risky Choice." *Journal of Neuroscience*, 30(49): 16567–16572.

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: 778.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review*, 100(1): 557–571.

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: 101477.

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(3): 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(2): 879–894.

Vieider, Ferdinand M., Peter Martinsson, Pham Khanh Nam, and Nghi Truong. 2019. "Risk Preferences and Development Revisited." *Theory and Decision*, 86(1): 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(2): 209–218.

Vollmer, Elisabeth, Daniel Hermann, and Oliver Musshoff. 2019. "The Disposition Effect in Farmers' Selling Behavior: An Experimental Investigation." *Agricultural Economics*, 50(2): 177–189.

Wang, Mei, and Paul S. Fischbeck. 2004. "Incorporating Framing into Prospect Theory Modeling: A Mixture-Model Approach." *Journal of Risk and Uncertainty*, 29(2): 181–197.

Ward, Patrick S., and Vartika Singh. 2015. "Using Field Experiments to Elicit Risk and Ambiguity Preferences: Behavioural Factors and the Adoption of New Agricultural Technologies in Rural India." *Journal of Development Studies*, 51(6): 707–724.

Wölbert, Eva, and Arno Riedl. 2013. "Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity." CESifo Working Paper Series No. 4339.

Wu, George, and Alex B. Markle. 2008. "An Empirical Test of Gain-Loss Separability in Prospect Theory." *Management Science*, 54(7): 1322–1335.

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

Wu, Tienhua. 2022. "Perception Bias Effects on Healthcare Management in COVID-19 Pandemic: An Application of Cumulative Prospect Theory." *Healthcare*, 10(2): 226.

Young, Diana L., Adam S. Goodie, and Daniel B. Hall. 2011. "Modeling the Impact of Control on the Attractiveness of Risk in a Prospect Theory Framework." *Journal of Behavioral Decision Making*, 24(1): 47–70.

Zeisberger, Stefan, Dennis Vrecko, and Thomas Langer. 2012. "Measuring the Time Stability of Prospect Theory Preferences." *Theory and Decision*, 72(3): 359–386.

Zhou, Wenting, and John Hey. 2018. "Context Matters." *Experimental Economics*, 21(4): 723–756.

A.5 Journals

Table A.2 provides a summary of the journals in which the papers included in our dataset were published. Journal categories are based on the classification provided by The Master Journal List (`https://mjl.clarivate.com/home`).

TABLE A.2: List of journals.

	Journal	Category
1	Agricultural Economics	Economics
2	American Economic Review	Economics
3	American Journal of Agricultural Economics	Economics
4	Attention, Perception & Psychophysics	Psychology
5	BioPsychoSocial Medicine	Psychology, Multidisciplinary
6	Cognition	Psychology, Experimental
7	Cognitive Psychology	Psychology
8	Decision	Management
9	Decision Support System	Operations Research & Management Science
10	Ecological Economics	Economics
11	Econometrica	Economics
12	Economic Development and Cultural Change	Economics
13	Economic Inquiry	Economics
14	Economica	Economics
15	Ekonomický časopis	Economics
16	eNeuro	Neurosciences
17	Environmental and Resource Economics	Economics
18	European Review of Agricultural Economics	Economics
19	Experimental Economics	Economics
20	Frontiers In Psychology	Psychology, Multidisciplinary
21	Games and Economic Behavior	Economics
22	Geneva Risk and Insurance Review	Economics
23	Health Economics	Economics
24	Healthcare	Health Policy & Services
25	International Economic Review	Economics
26	Journal of Banking & Finance	Economics
27	Journal of Behavioral and Experimental Economics	Economics
28	Journal of Behavioral Decision Making	Psychology, Applied
29	Journal of Behavioral Finance	Economics
30	Journal of Development Economics	Economics
31	Journal of Development Studies	Economics
32	Journal of Econometrics	Economics
33	Journal of Economic Behavior & Organization	Economics
34	Journal of Economic Theory	Economics
35	Journal of Experimental Psychology: General	Psychology, Experimental
36	Journal of Experimental Psychology: Learning	Psychology
37	Journal of Experimental Social Psychology	Psychology, Social
38	Journal of Mathematical Psychology	Psychology, Mathematical
39	Journal of Money and Economy	N.A.
40	Journal of Neuroscience	Neurosciences
41	Journal of Risk and Uncertainty	Economics

Continued on next page.

	Journal	Category
42	Journal of the Economic Science Association	Economics
43	Judgment and Decision Making	Psychology, Multidisciplinary
44	Journal of Behavioral Decision Making	Psychology, Applied
45	Management Science	Management
46	Medical Decision Making	Health Policy & Services
47	New Zealand Economic Papers	Economics
48	Operations Research	Management
49	Organizational Behavior and Human Decision Processes	Management
50	PeerJ	Multidisciplinary Sciences
51	PLOS Neglected Tropical Diseases	Parasitology
52	PLOS ONE	Multidisciplinary Sciences
53	Psychological Medicine	Psychology
54	Psychological Science	Psychology, Multidisciplinary
55	Psychonomic Bulletin & Review	Psychology, Experimental
56	Quantitative Economics	Economics
57	Review of Behavioral Economics	Economics
58	Review of Economics and Statistics	Economics
59	Revista de Administração de Empresas	Management
60	Scientific Reports	Multidisciplinary Sciences
61	Social Choice and Welfare	Economics
62	Social Cognitive and Affective Neuroscience	Neurosciences
63	Social Science & Medicine	Social Sciences, Biomedical
64	Southern Economic Journal	Economics
65	Spanish Journal of Finance and Accounting	Business, Finance
66	Theory and Decision	Economics
67	Transportation Research Part A	Economics
68	Transportation Research Part B	Operations Research & Management Science
69	Transportation Research Part C	Transportation Science & Technology
70	Water Resource and Economics	Economics

A.6 Global Map

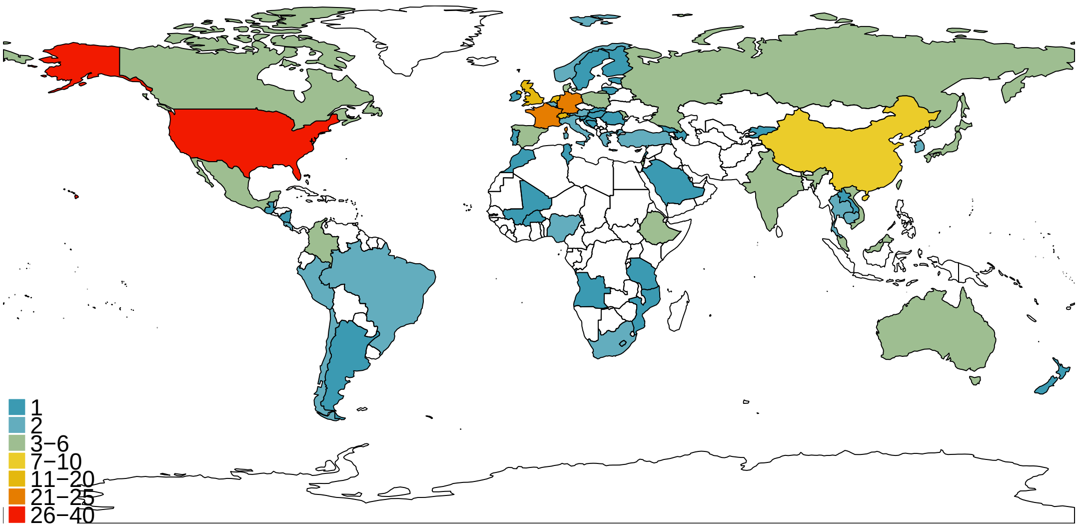

FIGURE A.3: Study location. *Notes*: This map was created using R (https://www.r-project.org/) on a base world map obtained from Natural Earth (https://www.naturalearthdata.com/).

B Standard Error Imputation

In our dataset, 113 out of 812 estimates lack associated standard errors. To address those missing SEs, the fundamental approach involved estimating the parameters characterizing their distribution in the data, represented as $\log(\boldsymbol{se}_i) \sim \mathcal{N}(\boldsymbol{Z}_i\boldsymbol{\xi}, \boldsymbol{\Omega})$. Here, $\log(\boldsymbol{se}_i)$ be a vector of element-wise natural logarithms of the standard errors, which serve to enforce non-negativity in the definition of standard errors, $\boldsymbol{Z}_i$ is a $1 \times M$ vector of characteristics of study i predictive of its standard errors, $\boldsymbol{\xi}$ is a $M \times 3$ matrix of coefficients, and $\boldsymbol{\Omega}$ is a covariance matrix with variances on its main diagonal, and covariances in its off-diagonal cells. In terms of predictors in $\boldsymbol{Z}$, we include characteristics—the square root of the number of subjects, experiment location, parameter sign, continent, utility function forms, and probability weighting function forms—in addition to a column of 1s to capture the intercept. In total, these characteristics prove to accommodate standard error variations well: 87.9%, 66.5%, and 98.5% of total variations explained for the three parameters, respectively.

Moreover, the vector $\log(\boldsymbol{se}_i)$ has a distinctive structure: it includes the logarithm of observed standard errors for studies where these are available, and the logarithm of *predicted* standard errors for cases where the standard errors are missing and must be inferred. The model thus fulfills two roles: (1) it estimates the regression coefficients $\boldsymbol{\xi}$ from the observed study characteristics in $\boldsymbol{Z}$; and (2) it uses these characteristics, along with the estimated coefficients $\boldsymbol{\xi}$, to *predict* (or impute) standard errors where they are not observed. The correlation structure encoded in $\boldsymbol{\Omega}$ ensures that the dependencies among errors are properly accounted for during the imputation process. As Figure B.1 shows, the mutual correlation of these three PT parameter estimates' standard error exist, which support the validity of our approach.

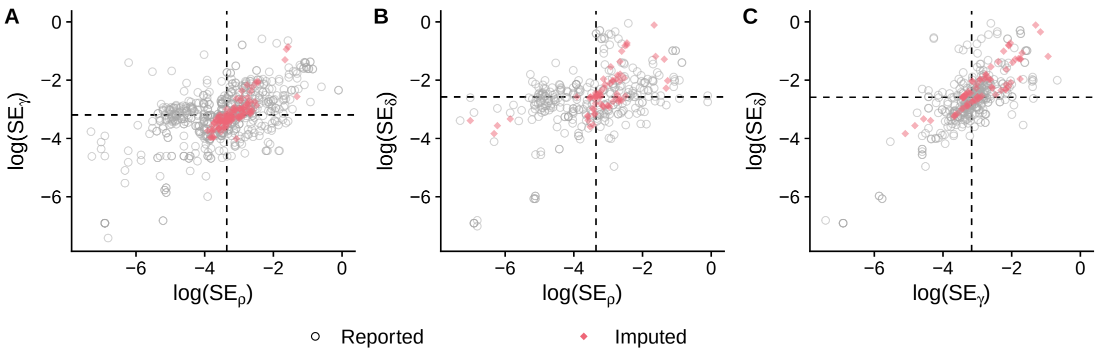

FIGURE B.1: Scatter plot of reported PT parameter and log(SE). *Notes*: Gray open circles represent reported standard errors (SEs), while red diamonds indicate imputed SEs. Dashed lines denote the medians of the corresponding variables.

B.1 Balance Check

A key question is whether the studies that report estimates' standard errors differ from those that do not report. Figure B.2 demonstrates the scatter plot of PT parameter estimates and their associated standard errors. Gray open circles represent complete observations, while red diamonds mark incomplete observations with imputed standard errors. With eyeballing, we can see that those incomplete observations are located evenly along the whole range of parameter values. Regarding the central value, the difference between complete and incomplete observations is generally mild: 0.26 vs. 0.31 for ρ, 0.72 vs. 0.73 for γ, and 1.16 vs. 1.05 for δ.[17]

Further, according to Wilcox test results, we see a significant difference in reported estimates of utility curvature ($p = 0.01$) and elevation ($p = 0.06$), while an insignificant one in those of likelihood sensitivity ($p = 0.34$). However, it is notable that this could be affected by the compound effect of heterogeneity of study characteristics as we documented in the main text. To partially address this, we choose to examine the difference by looking at the subsample that a power utility function is assumed. Now, the difference in ρ becomes insignificant ($p = 0.33$), and the difference in δ is smaller, though still significant (1.08 vs. 1.07; $p < 0.01$). This remaining significance can be caused by other characteristics other than utility function forms. To eliminate this concern, in the next subsection, we provide the meta-analysis results, in which we only include those complete observations. As we can see, the results are essentially unchanged.

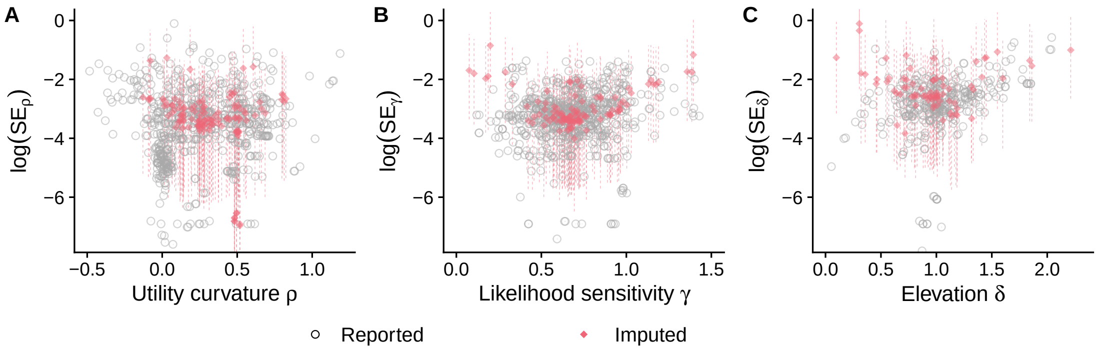

FIGURE B.2: Scatter plot of reported PT parameter and log(SE). *Notes*: Gray open circles represent complete observations, while red diamonds mark observations with imputed standard errors. Dashed red lines represent the percentile range from 2.5% to 97.5% across the 4,000 posterior draws for SE_i^{imputed}.

[17]In t-test, we only find a significant difference for likelihood sensitivity γ.

B.2 Robustness Check

To ensure that our results are not biased by our standard error imputation practice, this subsection reports the results of the meta-analysis, which only includes estimates that report associated standard errors. Table B.1 reports the weighted average of parameter estimates across gains and losses, while Tables B.2, B.3, and B.4 reports corresponding meta-regression results.

Table B.1: Meta-analysis of complete PT estimates.

	ρ		γ		δ	
	Mean	95% CrI	Mean	95% CrI	Mean	95% CrI
Gains	0.33	[0.30, 0.36]	0.67	[0.65, 0.70]	0.96	[0.91, 1.01]
Losses	0.27	[0.23, 0.32]	0.69	[0.65, 0.73]	0.93	[0.85, 1.01]

Table B.2: Robustness of meta-regression analysis of utility curvature (cf. Table 4).

Category	Variable	Median	2.5%	97.5%
Sign	Loss	-0.06	-0.09	-0.03
Utility	Expo	-0.24	-0.29	-0.19
PWF	LLO	Baseline		
	Gul/Power	0.60	0.46	0.74
	Prelec I	0.13	0.06	0.19
	Prelec II	0.10	0.04	0.16
	TK	0.00	-0.05	0.05
Elicitation	Binary choice	Baseline		
	Bisection	-0.06	-0.12	0.00
	List CE	-0.21	-0.26	-0.16
	List HL	0.00	-0.09	0.09
	Matching	-0.02	-0.09	0.05
	Lottery Menu	0.36	0.26	0.47
Estimate	Aggregate	Baseline		
	Ind. mean	0.02	-0.03	0.06
	Ind. median	-0.02	-0.07	0.04

Category	Variable	Median	2.5%	97.5%
Subject	Univ. pop.	Baseline		
	General	0.09	0.01	0.18
	Other	0.01	-0.06	0.09
Reward	Real money	Baseline		
	Hypo. money	-0.08	-0.13	-0.04
	Other	-0.11	-0.17	-0.04
Data	Lab	Baseline		
	Class	0.04	-0.05	0.13
	Field	0.36	0.28	0.45
	Online	-0.11	-0.2	-0.03
Continent	Europe	Baseline		
	Africa	0.18	0.10	0.25
	Asia	0.03	-0.02	0.09
	C/S-America	0.12	0.06	0.19
	North America	0.11	0.06	0.16
	Oceania	0.16	0.03	0.29
Frame	Visual aids	0.08	0.04	0.13

Table B.3: Robustness of meta-regression analysis of likelihood sensitivity (cf. Table 5).

Category	Variable	Median	2.5%	97.5%
Sign	Loss	0.04	-0.00	0.07
Utility	Expo	0.10	0.04	0.16
PWF	LLO		Baseline	
	Gul/Power			
	Prelec I	-0.03	-0.10	0.04
	Prelec II	0.06	-0.00	0.11
	TK	0.04	-0.02	0.10
Elicitation	Binary choice		Baseline	
	Bisection	-0.13	-0.20	-0.06
	List CE	-0.24	-0.30	-0.19
	List HL	-0.04	-0.14	0.06
	Matching	-0.06	-0.13	0.02
	Lottery Menu	-0.10	-0.22	0.03
Estimate	Aggregate		Baseline	
	Ind. mean	0.02	-0.03	0.06
	Ind. median	-0.02	-0.08	0.04

Category	Variable	Median	2.5%	97.5%
Subject	Univ. pop.		Baseline	
	General	-0.17	-0.26	-0.07
	Other	-0.15	-0.24	-0.07
Reward	Real money		Baseline	
	Hypo. money	-0.08	-0.14	-0.03
	Other	-0.19	-0.27	-0.11
Data	Lab		Baseline	
	Class	0.24	0.14	0.35
	Field	-0.07	-0.16	0.01
	Online	0.13	0.04	0.23
Continent	Europe		Baseline	
	Africa	-0.08	-0.18	0.03
	Asia	-0.03	-0.08	0.03
	C/S-America	0.01	-0.07	0.09
	North America	0.12	0.06	0.18
	Oceania	0.13	-0.03	0.27
Frame	Visual aids	-0.00	-0.05	0.05

Table B.4: Robustness of meta-regression analysis of elevation (cf. Table 6).

Category	Variable	Median	2.5%	97.5%
Sign	Loss	-0.03	-0.1	0.05
Utility	Expo	0.10	-0.02	0.21
PWF	LLO		Baseline	
	Gul/Power	-0.65	-0.85	-0.46
	Prelec I			
	Prelec II	-0.08	-0.17	0.01
	TK			
Elicitation	Binary choice		Baseline	
	Bisection	-0.06	-0.20	0.07
	List CE	0.03	-0.09	0.15
	List HL	-0.06	-0.36	0.25
	Matching	-0.17	-0.38	0.03
	Lottery Menu	0.46	0.02	0.88
Estimate	Aggregate		Baseline	
	Ind. mean	0.02	-0.08	0.13
	Ind. median	-0.08	-0.19	0.04

Category	Variable	Median	2.5%	97.5%
Subject	Univ. pop.		Baseline	
	General	0.18	0.00	0.37
	Other	0.23	0.08	0.37
Reward	Real money		Baseline	
	Hypo. money	-0.08	-0.18	0.03
	Other	0.21	0.01	0.40
Data	Lab		Baseline	
	Class	-0.20	-0.41	0.01
	Field	0.23	-0.01	0.47
	Online	-0.27	-0.49	-0.03
Continent	Europe		Baseline	
	Africa	0.29	0.09	0.49
	Asia	0.06	-0.06	0.19
	C/S-America	0.09	-0.08	0.28
	North America	0.15	0.03	0.27
	Oceania	-0.01	-0.33	0.30
Frame	Visual aids	0.11	0.01	0.22