



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Nonlinear decision weights or moment-based preferences? A model competition involving described and experienced skewness

Leonidas Spiliopoulos*, Ralph Hertwig

Max Planck Institute for Human Development, Center for Adaptive Rationality, Lentzeallee 94, 14195 Berlin, Germany



ARTICLE INFO

Keywords:

Risk
 Uncertainty
 Moment-based preferences
 Prospect theory
 Bayesian hierarchical mixture (latent class) modeling
 Decisions from description and experience

ABSTRACT

The predictive power of cumulative prospect theory and expected utility theory is typically compared using decisions from description, where lotteries' outcome values and probabilities are explicitly stated. In decisions from experience, individuals sample (in the sampling paradigm without cost) from the return distributions to learn outcome values and their relative frequencies; here cumulative prospect theory and expected utility theory require the calculation of probabilities from experience. Individuals, however, may be more attuned to the experienced moments of outcome distributions, rather than the probabilities. We therefore test the mean–variance–skewness model, and retrieve the proportion of expected utility theory (over income), cumulative prospect theory, and mean–variance–skewness populations using a latent-class hierarchical Bayesian model across six large datasets. For simple lotteries (with 1–2 outcomes), we find a mixture of cumulative prospect theory and mean–variance–skewness populations in decisions from both description and experience. For more complex lotteries (with 2–3 outcomes), all participants are classified as cumulative prospect theory types in decisions from description, but as mean–variance–skewness types in decisions from experience. This suggests that in decisions from experience with more complex return distributions, preferences for skewness are more predictive than nonlinear probability weighting.

1. Introduction

Many return distributions in the wild are heavily skewed (or asymmetric) due to an ecological negative correlation between probabilities and the magnitude of payoffs (see Pleskac & Hertwig, 2014). This negative relationship is often found in both market and nonmarket environments. Examples include lottery returns and gambling, health outcomes and disease, and asset returns. For this reason, decision makers are well advised to not only attend to the expected value of an option, but also to the skewness in its returns. Skewness is a measure of the asymmetry of a distribution with respect to its mean. A negatively skewed distribution will typically exhibit an extended or longer tail on the left side of the distribution than on the right side, and vice versa for a positively skewed distribution (see Fig. 1). Skewed distributions therefore exhibit different probabilities and magnitudes of extreme events on either side of the distribution (low vs. high outcomes) due to asymmetry, whereas the normal distribution is perfectly symmetric and therefore has zero skew.

Currently, two quite different theoretical frameworks describe how the effects of distribution skewness shape choices. The first framework is that of moment-based preference models, specifically the

mean–variance model (Markowitz, 1952; Tobin, 1958), which can be extended to explicitly model skewness preferences. The second framework is cumulative prospect theory (CPT, Tversky & Kahneman, 1992). Although it was not originally devised with skewness specifically in mind, it captures its effects indirectly, primarily through nonlinear decision weights. Cumulative prospect theory's original application was to decision making under risk (e.g., choice between monetary gambles with known and stated probabilities) but it has recently also been applied to asset pricing. In contrast, moment-based preferences have mostly been applied to asset pricing in financial markets, with very little application in individual decision making under risk and uncertainty. These distinct home turfs may explain why these two frameworks have coexisted without having been rigorously pitted against each other.

Our first major goal is to address this state of affairs by running a competitive model comparison. While there are a multitude of models of decision making under risk and uncertainty, our primary goal is to compare how well individuals' responses to skewness are captured by moment-based preferences and the nonlinear decision weights in cumulative prospect theory. Furthermore, we are interested in how the parameterizations of these models may depend on the decision-making

* Corresponding author.

E-mail addresses: spiliopoulos@mpib-berlin.mpg.de (L. Spiliopoulos), hertwig@mpib-berlin.mpg.de (R. Hertwig).

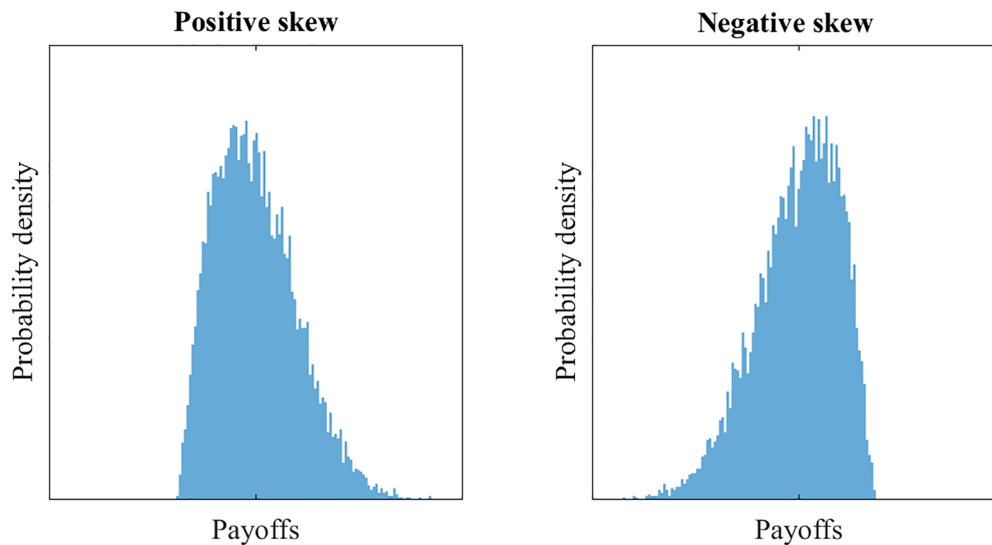


Fig. 1. Examples of positively and negatively skewed distributions.

context—specifically, whether the options’ outcomes and their probabilities (relative frequencies) are described or experienced. For this reason, we will restrict our focus to models that can be applied to decisions from description as well as decisions from experience. This means that we will not include accounts that have proven successful in modeling decisions from experience such as associative learning models (Weber, Shafir, & Blais, 2004), similarity-based contingent averaging rules (Plonsky, Teodorescu, & Erev, 2015), and the instance-based learning model (Gonzalez & Dutt, 2011; Gonzalez, Lerch, & Lebiere, 2003). However, note that while the latter was successful in the Technion Prediction Tournament (Erev et al., 2010), Glöckner, Hilbig, Henninger, and Fiedler (2016) find that the instance-based learning model performs significantly worse than cumulative prospect theory in four other datasets. Also, in Appendix E we show that our models’ predictions are on par with those of the top performing model in the experience–sampling paradigm of the Erev et al. (2010) tournament. As a benchmark we will also include expected utility theory over income (EUT) in the model competition. We employ data from six different existing (and with one exception, published) experimental studies and take advantage of a Bayesian hierarchical latent-class model that includes cumulative prospect theory, a mean–variance–skewness model (MVS), and expected utility theory over income. These studies examine decisions from both description and experience, a distinction to which we return shortly. By bringing together six datasets from different studies rather than running a new study we benefit from a large number of datapoints that would be nearly impossible to collect in a single new choice experiment in the laboratory. Based on this rich dataset, we can infer participant heterogeneity, use cross-validation techniques to address overfitting and examine the out-of-sample predictive ability of the models. Furthermore, the data cover a large range of the potential properties of risky options, that is, combinations of outcomes and probabilities, which improves model generalizability and the robustness of our conclusions. However, because each individual study examined different decision models and used different statistical inference techniques, it is nearly impossible to compare their results and draw meaningful conclusions. Our analysis, a meta-analysis of sorts, unifies the statistical inference technique and the decision models, thereby rendering comparisons feasible and meaningful. Such an approach does not eliminate the methodological differences across studies. Where relevant, we will highlight them and discuss whether they may impact the findings.

Our second major goal is to capture the extent of heterogeneity in individuals’ behavior using a latent-class hierarchical Bayesian model.

Participant heterogeneity is found in a diverse range of tasks studied in psychology and economics—for example, in static individual decision models (e.g., Abdellaoui, 2000), models of different levels of sophistication in strategic behavior (e.g., Costa-Gomes, Crawford, & Broseta, 2001), and learning models (e.g., Spiliopoulos, 2012). Estimating models on the basis of pooled data when individuals prove heterogeneous is highly problematic in terms of parameter recovery and model selection (Cohen, Sanborn, & Shiffrin, 2008; Estes & Maddox, 2005).

1.1. The description–experience gap in risky choice

For many decades, the standard tool of economists and psychologists for studying individual decision making under risk has been choice between lotteries. These lotteries are event-contingent outcomes (e.g., win \$100 million with certainty; win \$500 million with 10%, \$100 million with 89%, and nothing otherwise; Allais, 1953). In these lotteries, the individual is either provided with the outcomes’ objective probabilities from the outset (a condition referred to as *risk*), or has to assign subjective likelihoods to events (a condition referred to as *ambiguity* in the economic literature; e.g., win \$100 if home team wins, otherwise nothing).

Numerous empirical findings suggest a tendency for people to overweight small probabilities and underweight moderate to high probabilities under risk in response to described probabilities and outcomes (e.g., Tversky & Kahneman, 1992); these are called decisions from description. In the last decade, however, researchers have begun to extensively and systematically compare decisions from description with choices in which properties of the payoff distributions are experienced rather than described. In decisions from experience, possible outcomes and their relative frequencies must be learned through sampling—that is, through repeated draws with replacement from payoff distributions unknown to the decision maker (Barron & Erev, 2003; Erev & Roth, 2014; Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009). The decision maker thus makes choices on the basis of statistical rather than a priori probabilities (in the terminology of Knight, 1921). Statistical probabilities cannot be deduced or determined exactly, but must be assessed in an empirical manner. Research on decisions from experience has employed a simple experimental tool, a “computerized money machine.” Respondents see two urns or buttons on a computer screen, each representing a payoff distribution that is initially completely unknown. Three variations of this tool have commonly been employed (see Hertwig & Erev, 2009). In the *sampling paradigm*, people first sample as many outcomes as they wish

before deciding on a distribution from which to make a single draw. In the *full-feedback paradigm*, each draw contributes to people's earnings and they receive draw-by-draw feedback on the obtained and the foregone payoffs (i.e., payoff they would have received had they selected the other option). Finally, the *partial-feedback paradigm* differs from the full-feedback paradigm only in that people merely learn about the obtained payoffs. In the partial-feedback paradigm, unlike the other two paradigms, respondents face an exploitation–exploration trade-off. They must strike a balance between two goals associated with every choice: to obtain a desired outcome (exploitation) or to gather new information about other, perhaps better, actions (exploration). In our analyses, we are concerned with datasets that were collected using the sampling paradigm.

The starting point of the decisions from experience research was the question of the extent to which choices and value maximization systematically change when people rely on direct experience of the structure of the probabilistic world rather than simply being told about it. In the latter case, people receive symbolic or graphic descriptions of the options' possible outcomes and probabilities, and no learning is required. Numerous studies have systematically pitted decisions based on experience against decisions based on descriptions, mostly using monetary lotteries. Taken together, these investigations have revealed a systematic and robust difference between the two kinds of decisions, namely a description–experience gap in the choice proportions. The overall pattern obtained across the three experimental paradigms can be summarized as follows. In decisions from experience, people behave as if rare events have less impact than they deserve according to their *objective probabilities*, whereas in decisions from description, people behave as if rare events have more impact than they deserve (Hertwig & Erev, 2009). To what extent this portrayal of the shape of the probability weighting function and its reversal (inferred from the choices observed in decisions from description and experience) is appropriate will be discussed later.

What causes this description–experience gap in choices? Several contributing factors have been proposed (see Hertwig & Erev, 2009, for a review). Based on people's limited search in the sampling paradigm, reliance on small samples has been proposed as a key factor in attenuating the impact of rare events. In fact, in a meta-analysis of 55 data sets of the sampling paradigm (Wulff, Mergenthaler-Cansec, & Hertwig, 2018) and involving 40,246 trials (tasks \times participants), the median sample size was 14 for all choices involving a risky and a safe option (15,054 trials) and 22 for all choices involving two risky options (30,185 trials). In about one third of the trials (36%), at least one of the outcomes was not experienced. The small samples that people, on average, rely on exact a sampling error and this sampling error systematically changes the experienced options (relative to the objective payoff distributions).

Several approaches have been taken to examine whether the gap observed in the sampling paradigm can indeed be reduced to individuals' reliance on small samples and, by extension, sampling error. If sampling error were the sole culprit, then reducing the error by extending individuals' search effort should attenuate and eventually eliminate the gap. Numerous methods have been employed to achieve this goal, including higher monetary incentives (Hau, Pleskac, Kiefer, & Hertwig, 2008), fixed large samples (e.g., Camilleri & Newell, 2011; Hau et al., 2008), and sampling without replacement (e.g., Ungemach, Chater, & Stewart, 2009) to yoking of description to experience (e.g., Hau, Pleskac, & Hertwig, 2010; Rakow & Demes, 2008). Taken together, these methods, on average, reduced the magnitude of the gap but did not eliminate it (see Hertwig & Erev, 2009; Wulff et al., 2018). This suggests that sampling error (as a result of small samples) is a key, but not the sole, contributor to the description–experience gap. In fact, in their meta-analysis, Wulff et al. (2018) found that across trials in which individuals happen to experience outcomes as frequently as the objective (described) probabilities would suggest, a systematic description–experience gap persisted. Its magnitude was comparable in

size to the gap when including all trials (even those in which the experienced relative frequencies diverged from the described probabilities). However, the size of the gap is mediated by characteristics of the lotteries, such as whether two risky lotteries are compared to each other or one risky lottery is compared to a certain outcome (e.g., Glöckner et al., 2016; Wulff et al., 2018). The gap is typically substantially larger in the latter compared to the former. Measured in terms of proportion of risky choices, the gap is robust across a wide range of situations, although other specifications of the gap based on the shape of the probability weighting function are less so (see Regenwetter & Robinson, 2017; Wulff et al., 2018).

The description–experience gap in choice behavior has been observed both for experimental designs in which search was cost-free (the sampling paradigm) and costly (the partial-feedback paradigm; see, e.g., Hertwig & Erev, 2009). Here we focus on the sampling paradigm but want to emphasize that the term “decisions from experience” generally refers to both cost-free and costly accumulation of experience of outcome and probability information. Our key concern is to investigate the extent to which people respond differently to variance and skewness when they are described versus when they need to be experienced through sampling. Our examination is related to the work by Weber et al. (2004), who observed that people respond differently to risk (outcome variance) depending on whether they encounter it through description or experience. Specifically, they found that the risk preferences of both animals (i.e., foraging birds and insects) and humans are better predicted by the coefficient of variation (CoV)¹ than by the variability of an option's possible outcomes. Furthermore, they found that humans' risk preferences become strongly proportional to the CoV when they learn about choice options like other animals, namely by experiential sampling over time rather than description. While these results are intriguing, we do not examine CoV as it suffers from an important restriction: It can only be generally applied to positive outcomes (gains) and not to mixed lotteries.²

1.2. The mean–variance framework and skewness

The mean–variance framework originally proposed by Markowitz (1952) and Tobin (1958) is the cornerstone not only of the capital asset pricing model (Sharpe, 1964), but also of the finance literature. This framework is compatible with a risk-averse expected utility maximizer as long as one of two conditions hold: returns are normally distributed (and thus are completely specified by their first and second moments) or utility is a quadratic function of wealth. However, asset returns are significantly non-normal, exhibiting both nonzero skewness and excess kurtosis (e.g., Albuquerque, 2012). In response, arguments in favor of mean–variance–skewness (MVS) models for asset pricing have been put forth (e.g., Kraus & Litzenberger, 1976).³ Owing to the rise in popularity of cumulative prospect theory in economics, other studies in finance have broken ranks with the historical prevalence of moment-based preference modeling and have turned to cumulative prospect

¹ CoV is a measure of the relative variability of risky choice options that is calculated by dividing the standard deviation of outcomes by their expected value; it is often multiplied by 100 to express it as a percentage. It is a measure of relative risk—risk per unit of expected return—rather than a measure of absolute risk such as the unstandardized variance of the outcomes.

² The problem arises with the denominator of CoV, which is the expected value. Allowing lotteries with negative outcomes will in some cases lead to expected values close to zero—as the expected value approaches zero, the CoV approaches infinity regardless of the standard deviation.

³ See also Arditti (1967), Harvey and Siddique (2000), Harvey, Liechty, Liechty, and Müller (2010), and Rubinstein (1973) for further arguments. Note that despite earlier claims based on the derivation of moment preferences from a Taylor expansion of a utility function (e.g., Arditti, 1967), Brockett and Kahane (1992) prove that expected utility and moment-based preferences are generally incompatible for arbitrary utility functions.

Table 1
Experimental datasets and their characteristics.

Dataset source [Abbrev.]	Participants× Tasks	Outcomes per option	Domain (gain,loss,mixed)	Payment (#)	Details
Erev et al. (2010) [<i>D_{EEER}</i>]	40 × 60	(1, 2)	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	1	One option was a sure outcome
Hey and Orme (1994) [<i>D_{HO}</i>]	80 × 200	(≤ 3, ≤ 3)	(1, 0, 0)	2	Pie-chart presentation
Harrison and Rutström (2009) [<i>D_{HR}</i>]	158 × 60	(≤ 3, ≤ 3)	(0.4, 0.23, 0.37)	3	Pie-chart presentation, participants assigned only to one domain
Erev et al. (2010) [<i>E_{EEER}</i>]	80 × 30	(1, 2)	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	1	One option was a sure outcome
Frey et al. (2015, Exp. 2) [<i>E_{HH}</i>]	80 × 30	(3, 3)	$(\frac{2}{3}, \frac{1}{3}, 0)$	All	Significant variation in lotteries' expected values
Hau and Hertwig [<i>E_{FMH}</i>]	70 × 72	(3, 3)	$(\frac{1}{2}, \frac{1}{2}, 0)$	All	Significant variation in lotteries' expected values and variance

theory instead, (e.g., Barberis & Huang, 2008).

1.3. Cumulative prospect theory, expected utility theory, and skewness

In contrast to asset pricing in markets, another strand of research has examined the effects of skewed distributions in individual decision making (that is, with exogenously provided probabilities and outcomes). Expected utility theory struggles to explain two concurrent behaviors that are often observed in the context of real-world choices between skewed options. The canonical examples are gambling (purchasing lottery tickets) and purchasing insurance—the former is an example of positive skewness-seeking, the latter of negative skewness-aversion. Expected utility theory can only capture both of these behaviors by assuming that the utility function is concave below the current wealth point but convex above the current wealth.

Although not explicitly investigating skewness, empirical studies fitting cumulative prospect theory to choices between simple lotteries—often lotteries with a maximum of two outcomes per option—indirectly capture the effects of skewed distributions as the majority of lotteries by design exhibit nonzero skewness. Cumulative prospect theory can indirectly capture skewness seeking/aversion through the interaction of the curvature of the probability weighting and value functions. Indeed, this is one of its primary advantages over expected utility theory in terms of descriptive power. Although cumulative prospect theory is compatible with both negative and positive skewness-seeking, in practice the range of empirically derived parameter estimates for decisions from description are compatible with positive skewness-seeking; we return to this in more detail in the General Discussion.

Positive skewness-seeking has been found to be the modal behavior, although to varying degrees in decision from description experiments and gambling in the field (see Appendix C for an in-depth discussion and references). Furthermore, Nursimulu and Bossaerts (2014), Symmonds, Bossaerts, and Dolan (2010), Symmonds, Wright, Bach, and Dolan (2011), and Wu, Bossaerts, and Knutson (2011) are the only studies that estimated moment-based preference models including skewness from empirical data; they also searched for neural correlates of such preferences. These studies find support for moment-based preferences both in term of observed choices and neural data. However, due to the constraints imposed by the fMRI scanning methodology they had a small number of participants and/or tasks.

To summarize, there is abundant evidence both from the field and from experiments that the skewness of return distributions matters to choice. However, relatively little attention has been paid to explicitly modeling choice in response to skewness and, in particular, to the direct comparison of nonlinear decision weights and explicit skewness preferences as competing explanations. Furthermore, existing studies have not rigorously compared competing models using large datasets. This is our objective. We next describe the experiments and the data we use for our model competition. We then present the three decision models that

we competitively test: cumulative prospect theory, mean–variance–skewness, and expected utility theory, followed by the specification of the Bayesian hierarchical latent-class model. After turning to the results we conclude with a discussion of the implications of our findings.

2. The experimental datasets

We use a diverse set of five datasets from published studies and one unpublished set of experimental data. The latter dataset was collected by Hau and Hertwig and has been made available as supplemental material; methods and instructions are provided in Appendix F. Of these six datasets, three are decisions from description and the other three are decisions from experience using the sampling paradigm. Datasets were chosen primarily using the criterion that they have a large number of observations relative to other published studies, both in terms of participants and decisions per individual (see Table A1 in Wulff et al., 2018), and ideally included risky options with two or even three outcomes. The six datasets have a total of 37,720 observations. Table 1 presents the sources of the datasets (the published studies and the abbreviations we have assigned to each dataset for ease of reference) and summarizes important characteristics of the six datasets, including the number of participants and properties of the lotteries (see Appendix A for more details).

All datasets presented participants with two options; however, for *D_{EEER}* and *E_{EEER}* one of the options was a sure outcome, and the other option had two possible outcomes. By contrast, *E_{HH}* and *E_{FMH}* had two options with three outcomes each, while *D_{HO}* and *D_{HR}* included options with up to three outcomes each.⁴ Differences between decision from description experiments also exist in terms of how information is presented—*D_{EEER}* presented probabilities numerically, whereas *D_{HO}* and *D_{HR}* used a pie-chart presentation. Also, a random lottery incentive mechanism was used for four of the six experiments, but for the remaining two experiments all choices were incentivized.⁵

⁴ Lopes and Oden (1999) collected data on decisions from description using lotteries with five outcomes. While we argue that such multiple-outcome lotteries are important and understudied, we could not find another study using four–five outcomes in decisions from experience, which would render a comparison of the two feasible. Since three outcomes per lottery was the highest number found in both the decision from description and decision from experience literature, we set this as our limit. However, we will be addressing this in future research with experiments using more than three outcomes in both decision from description and decision from experience environments.

⁵ In the decision from experience literature, it is common to pay out all of a participant's choices rather than a single choice or subset of choices. This is to provide incentives throughout for participants to continue sampling sufficiently from the distributions, even though the act of sampling does not directly contribute to earnings. This raises the possibility of wealth effects (see Appendix F.2 for a further discussion and an analysis concluding that wealth effects did not play a role in the *E_{HH}* dataset).

Table 2
The relationship between choices and lotteries’ mean, variance and skew.

	D_{EER}	D_{HO}	D_{HR}	E_{EER}	E_{HH}	E_{FMH}
δEV	1.0539	0.5075	0.8776	1.2031	0.5227	0.3617
δVar	-0.0224	-0.0207	-0.0192	-0.0254	-0.0020	-0.0016
$\delta Skew$	0.0060	0.0002	0.0009	0.0016	0.0013	-0.0006
DIC	2852.126	18997.43	12196.79	2453.596	2307.134	5493.562
log(ML)	-1459.934	-9539.872	-6134.682	-1260.226	-1189.823	-2786.153

2.1. Preliminary analysis

To ascertain whether the skewness of lotteries (and their mean and variance) impact choices, we employ a Bayesian logistic regression. The dependent variable is a binary indicator of lottery choice and the independent variables are the difference in the expected value of the two lotteries δEV , the difference in the variance δVar , and the difference in the skew $\delta Skew$. Positive coefficients for these three variables imply a preference for higher expected values, higher variance and more positive (or less negative) skewness. The results are summarized in Table 2 and reported in full in Appendix B (Table B.7). The 95% highest posterior density interval⁶ (HPDI) of the parameters excluded zero in all cases except for δVar in E_{HH} ; therefore we can speak of significant effects. As expected, participants exhibited a preference for higher expected values and an aversion to variance in all datasets. Importantly, participants exhibited a preference for positive-skewness in all studies except E_{FMH} . Overall these findings are in accord with existing research, and highlight the importance of lottery skewness on choices. Consequently, we argue that this justifies further investigation regarding the mechanisms through which skewness may operate: either directly through preferences for statistical moments or indirectly through non-linear decision weights.

3. The decision models

We refer to the models used by participants as “decision models” and reserve the term “model” for the statistical model (which will include multiple decision models). All decision models employ the same stochastic choice rule (logit, or softmax) with a best-response parameter τ . The higher τ is, the more likely a person is to best respond, that is, choose the option with the highest value; at the other extreme, if $\tau = 0$ then choice is random. The probability of choosing an arbitrary lottery L_1 from a set of K lotteries for a decision rule that values lottery k as $V(L_k)$ is given in Eq. (1). We consider other stochastic parameterizations to be promising (e.g., Blavatsky, 2014c; Wilcox & Wilcox, 2011), but it is impractical for us to simultaneously competitively test these. Consequently, we chose the logit stochastic choice rule as it is still widely used. Throughout the rest of the article, the reader should bear in mind that for datasets of decision from description, all models use the described probabilities, but for datasets of decision from experience the experienced relative frequencies are used instead. For example, suppose the participant observed an outcome of 100 ten times while sampling, an outcome of 50 five times, and an outcome of 10 another five times. The experienced probabilities for the outcomes 100, 50, and 10 would be 0.5, 0.25, and 0.25 respectively.

$$p(L_1) = \frac{e^{\tau V(L_1)}}{\sum_k e^{\tau V(L_k)}} \tag{1}$$

⁶ The 95% HPDI is the smallest possible interval that contains 95% of a distribution’s density—this can be literally interpreted as the probability that the parameter lies within this interval.

3.1. Cumulative Prospect Theory (CPT)

We parameterized the probability-weighting function—denoted by $w(p)$ —using the linear in log-odds functional form, first used by Goldstein and Einhorn (1987). The γ parameter influences the curvature of the probability-weighting function: $\gamma < 1$ implies an inverse S-shaped curvature, whereas $\gamma > 1$ implies an S-shaped curvature. If $\gamma \neq 1$ then the probability-weighting function concurrently exhibits overweighting and underweighting of probabilities for different ranges of p . The δ parameter affects the elevation of the probability-weighting function. If $\gamma = 1$ then $\delta > 1$ implies overweighting of all probabilities (often attributed to optimism) and $\delta < 1$ implies underweighting (pessimism). If $\gamma = \delta = 1$ then the probability-weighting function is a linear function identical to the described or experienced probabilities. Recall that in decisions from experience, the experienced relative frequencies are used in the probability-weighting function. Note that since D_{HO} involves lotteries with gains only, cumulative prospect theory is equivalent to rank-dependent expected utility for this dataset.

The value function is the standard power utility function (CRRA) with risk parameter α except for the addition of loss aversion captured by λ . The latter is not included in datasets without mixed-outcome lotteries (D_{HO} , E_{HH} , E_{FMH}). Assuming that the outcomes x_i with probabilities p_i of an option are ordered by value such that $x_1 \leq \dots \leq x_k \leq 0 \leq x_{k+1} \leq \dots \leq x_n$, then cumulative prospect theory is defined as follows in Eq. (2). We have refrained from using different parameters in the gain and loss domains due to the problems with parameter interaction and recoverability in cumulative prospect theory models (Broomell & Bhatia, 2014); see Appendix D.3 for our reasoning and supporting empirical evidence.

$$V_{cpt}(L) = \sum_{i=1}^k \omega_i^- v(x_i) + \sum_{j=k+1}^n \omega_j^+ v(x_j) \tag{2}$$

$$v(x) = \begin{cases} x_i^\alpha & x_i \geq 0 \\ -\lambda(-x_i)^\alpha & x_i < 0 \end{cases}$$

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma} \quad \gamma, \delta > 0$$

$$\begin{aligned} \omega_i^- &= w(p_1) \\ \omega_n^+ &= w(p_n) \\ \omega_i^- &= w(p_1 + \dots + p_i) - w(p_1 + \dots + p_{i-1}) \quad 1 < i \leq k \\ \omega_j^+ &= w(p_j + \dots + p_n) - w(p_{j+1} + \dots + p_n) \quad k < j < n \end{aligned}$$

3.2. Expected Utility Theory (EUT)

The utility of an outcome x_i is parameterized using the CRRA utility function. This is the most commonly used value function for cumulative prospect theory, and therefore we chose the same function for expected utility theory for the sake of comparability and to put both models on a similar footing. A value of $\rho = 1$ implies risk neutrality, $\rho > 1$ implies risk seeking, and $0 < \rho < 1$ implies risk aversion. As is common in the literature, our specification is one of expected utility over income rather than wealth: There is no asset integration and behavior is reference-point dependent. Cox and Sadiraj (2006) have argued in favor of this specification on the grounds that it is immune to the Rabin (2000)

calibration critique. Furthermore, empirical evidence finds that participants are better explained by choice models over lottery outcomes rather than final outcomes over wealth (e.g., Harrison, List, & Towe, 2007; Heinemann, 2008). We note that expected utility theory over income is a special case of cumulative prospect theory with the following parameter restrictions: $\gamma = \delta = \lambda = 1$.

$$EU(L) = \sum_i p_i u(x_i) \tag{3}$$

$$u(x) = \begin{cases} x_i^\rho & x_i \geq 0 \\ -(-x_i)^\rho & x_i < 0 \end{cases}$$

3.3. The Mean–Variance–Skewness Model (MVS)

The mean–variance–skewness model shown in Eq. (4) directly models preferences for the first three central (unstandardized) moments of the distribution of lottery outcomes L ; the parameters m_1 , m_2 , m_3 correspond to the mean, variance, and skewness respectively.⁷ The inclusion of all three moments does not preclude the inference of simpler decision models. For example, if β^s is close to zero, then this implies that participants are using a mean–variance model. Similarly, if both β^s and β^v are close to zero, then this implies an expected value calculation only, ignoring both the variance and skewness of lotteries.

$$V_{mvs}(L) = m_1 + \beta^v m_2 + \beta^s m_3 = \bar{L} + \beta^v Var(L) + \beta^s Skew(L) \tag{4}$$

$$m_1 = \sum_i (p_i x_i) = E(L)$$

$$m_2 = \sum_i (p_i (x_i - \bar{x})^2) = E(L - m_1)^2$$

$$m_3 = \sum_i (p_i (x_i - \bar{x})^3) = E(L - m_1)^3$$

4. The statistical model

We employ a latent-class hierarchical Bayesian approach for statistical inference related to the decision models (see Appendix D.1 for a discussion of the advantages of this approach). This allows for two kinds of heterogeneity, both of which have been found to be important in behavioral decision research. The first is *parameter heterogeneity*, which allows individuals to have different parameter values for the same decision model. For example, two people who both can be captured by cumulative prospect theory may have a different loss-aversion parameter λ . The second is *model heterogeneity*, which allows individuals to use different decision models. For example, one person may best be captured by a cumulative prospect theory decision model whereas another by the mean–variance–skewness model. Model heterogeneity is captured by using a latent-class or mixture model approach, that is, by inferring the proportion of individuals belonging to each latent-class or decision model. Ours is the first paper to use both parameter and model heterogeneity simultaneously for decisions from description and decisions from experience in a Bayesian framework—see Appendix D.2 for a comparison to other published models incorporating some form of heterogeneity. Our hierarchical model permits for a maximum of three latent classes from the set of decision models EUT, CPT, and MVS, including all possible subsets, {CPT,MVS,EUT}, {CPT,MVS}, {CPT,EUT}, {MVS,EUT}, {CPT}, {MVS}, {EUT}. This allows us to determine both the number of latent-classes in

⁷ We selected this specification from a range of alternatives based on its parsimony and prior evidence. For example, the third-order moment could be further broken down into its upper and lower moments—that is, one could estimate a different parameter for positive skewness and negative skewness. However, Ebert (2015) and Symmonds et al. (2011) did not find evidence for differences in the strength of preference for negative and positive skewness.

the participant pool and the identity of the decision models in the classes.

4.1. Method

We employ the following hierarchical model. The prior distribution for the probability of using one of C decision models $\pi = (\pi_1, \dots, \pi_C)$ is distributed as a uniform, symmetric Dirichlet function (Eq. (5)): Each combination of probabilities in the $(C - 1)$ -simplex is equally likely.⁸ A categorical distribution is then used to assign an individual to one of C classes or decision models indexed by the variable c_i (Eq. (6)). The probability of choosing lottery L_1 by individual i in task j depends on the value of the specific decision model V_{c_i} used by an individual and is denoted by $p_{i,j}$ (Eq. (7)). The binary variable $a_{i,j}$ denotes which of the two lotteries was actually chosen and follows a Bernoulli distribution with probability $p_{i,j}$ (Eq. (8)).

$$\pi \sim \text{Dirichlet}(\mathbf{1}_C) \tag{5}$$

$$c_i \sim \text{Categorical}(\pi) \tag{6}$$

$$p_{i,j} = \frac{e^{\tau_j V_{c_i}(L_1)}}{\sum_k e^{\tau_j V_{c_i}(L_k)}} \tag{7}$$

$$a_{i,j} \sim \text{Bernoulli}(p_{i,j}) \tag{8}$$

We specified weakly informative prior distributions for the parameters of each decision model—that is, we did not incorporate the prior information about the most likely parameter values available from previous studies, but we did inform our priors about the expected range of parameter values and scaled these appropriately. The prior distributions are presented in detail in Appendix D.4.

Each model—one for each possible combination of the three latent classes—was computed using the MCMC algorithm implemented in JAGS for five different chains. After discarding a burn-in of 50,000 samples per chain, we collected 10,000 samples per chain to approximate the posterior distributions of the parameters of interest. We chose not to thin the samples based on new evidence (Link & Eaton, 2012), contrary to existing beliefs about the value of thinning. Convergence of the chains was checked visually and using the \hat{R} -statistic (Gelman & Rubin, 1992).

A k -fold ($k = 3$) cross-validation criterion (based on the out-of-sample average likelihood, \overline{lcV}) was employed for model selection due to issues with other common techniques such as the deviance information criterion, which is problematic for selection between mixture models (Celeux, Forbes, Robert, & Titterton, 2003). Cross-validation automatically accounts for the possibility of overfitting the models to the training data by directly examining the predictive performance of the model on unseen data. Typically, models are ranked according to \overline{lcV} and the best performer is declared a winner; however, this neglects two important issues. The first is that \overline{lcV} is a random variable, dependent on the partitioning of the dataset; models with close performance may therefore not be statistically significantly different. The second is that even if two models are statistically significantly different, this does not mean that they are economically significantly different—that is, that we gain a significant increase (in terms of magnitude) in predictive power. We define two models as economically significantly different if the difference in their \overline{lcV} is at least 0.01: For 100 choices one of the models should on average predict at least one additional choice correctly. We considered both types of significance and ultimately chose the most parsimonious model from the set of models not economically significantly different from the top-performing model. Further details on

⁸ Our choice of the Dirichlet distribution with parameters $(\mathbf{1}_C)$ assigns a greater prior likelihood to the existence of heterogeneity. This is consistent with prior evidence from studies estimating the proportion of decision model use (see Appendix D.2).

Table 3
Prevalence of CPT, MVS, and EUT latent classes in six data sets (D = Description, E = Experience).

		D_{EER}	D_{HO}	D_{HR}	E_{EER}	E_{HH}	E_{FMH}
π [95% HPDI]	CPT	0.55 [0.50,0.60]	1	1	0.76 [0.70,0.82]	0	0
	MVS	0.45 [0.40,0.50]	0	0	0.24 [0.19,0.31]	1	1
	EUT	0	0	0	0	0	0
\bar{l}_{cv}		0.675	0.712	0.641	0.759	0.728	0.682

the model selection procedure can be found in [Appendix D.5](#).

5. Results

5.1. Model heterogeneity

[Table 3](#) presents the best model according to our selection criteria for each dataset, including the proportion of participants belonging to each latent class (π) and the average cross-validation likelihood (\bar{l}_{cv}). The complete set of results for every possible combination of decision models can be found in [Appendix D.6](#). We found the following major results.

Result 1. *Expected utility theory performs poorly and is never included in the best model for any of the datasets; there is no population of EUT individuals independent of the CPT and MVS populations.* Neither the statistical nor economic significance criterion ever chose a model including an expected utility theory population—see [Table D.8](#) in [Appendix D.6](#). Furthermore, the difference between the \bar{l}_{cv} of the best-performing econometric model and a single-class EUT model is substantially different for all six datasets: 0.114, 0.018, 0.025, 0.046, 0.032 and 0.021 for D_{EER} , D_{HO} , D_{HR} , E_{EER} , E_{HH} , and E_{FMH} respectively.

Strictly speaking, our result regarding expected utility theory concerns the existence of an EUT population that is independent of a CPT population. Cumulative prospect theory nests expected utility theory as a special case if $\gamma = \delta = \lambda = 1$, therefore a CPT population with significant density around these parameter values will include individuals whose behavior would approximate expected utility theory. Similarly, if $\beta^v = \beta^s = 0$ then mean–variance–skewness nests expected utility theory under the assumption of risk neutrality. The existing decision from description literature that permits for model heterogeneity only (and not parameter heterogeneity) typically finds roughly 20% of individuals as belonging to expected utility theory ([Bruhin, Fehr-Duda, & Epper, 2010](#); [Conte, Hey, & Moffatt, 2011](#)). Our results are generally in accord with this, depending on how strictly one defines approximate EUT behavior within a CPT or MVS population. We will return to this in more detail when examining parameter heterogeneity. However, we can already conclude that expected utility theory is too rigid to accommodate the parameter heterogeneity exhibited across participants. Consequently, any EUT types may be subsumed under the CPT or MVS populations.

Result 2. *There is significant model heterogeneity for two datasets, D_{EER} and E_{EER} . The 2-class CPT–MVS model is the best performer for both of these datasets.* Participants in D_{EER} and E_{EER} are classified as using one of two decision models, cumulative prospect theory or mean–variance–skewness. In the D_{EER} dataset, the probability of participants belonging to these classes are 0.55 and 0.45 respectively. In the E_{EER} dataset, the probabilities of participants belonging to cumulative prospect theory and mean–variance–skewness are 0.76 and 0.24 respectively. The necessity of allowing for both classes is evident not only from a comparison of \bar{l}_{cv} ([Table D.8](#)) but also from the fact that the 95% HPDIs of the class probabilities do not include (and

are very far from) a value of zero. Participants were classified with high posterior probabilities to each class, indicating that the two latent classes are necessary and that the statistical technique efficiently recovered these classes (see [Appendix D.7](#) for details).

Result 3. *No significant model heterogeneity was found in D_{HO} , D_{HR} , E_{HH} , or E_{FMH} . The single-class CPT model is the best model in the two description datasets, whereas the single-class MVS model is the best model in the two experience datasets.* In contrast to [Result 2](#), we do not find significant model heterogeneity in the D_{HO} , D_{HR} , E_{HH} , and E_{FMH} datasets (where the two lotteries have 2–3 outcomes each; see [Tables 1 and 3](#)). Consequently, parameter heterogeneity—discussed in [Section 5.2](#)—was adequate in capturing participants’ diverse behavior without the need for model heterogeneity. We are unaware of any other study that permitted for model heterogeneity in decisions from experience to which we could compare this result.

5.1.1. Summary and discussion

A significant proportion of participants were classified as using the MVS model in four out of six datasets—one decision from description dataset and all three decision from experience datasets. Furthermore, in the two decision from experience datasets with options comprised of two or three outcomes each (without a lottery consisting of a sure outcome), every single participant was classified as MVS. We consider this to be evidence of the existence of MVS types of participants, particularly in decisions from experience. By extension, we conclude that nonlinear probability weighting is inadequate for capturing the behavior of participants in decisions from experience, at least in multi-outcome lotteries (in the interior of the Marschak-Machina triangle), and that explicit preferences for skewness play an important role here. On the other hand, in four out of six datasets, a significant proportion of participants were classified as CPT—one decision from experience dataset and all three decision from description datasets. Our analysis suggests that, particularly for multi-outcome lotteries, the scope of the cumulative prospect theory model may be limited primarily to decisions from description, whereas the scope of the mean–variance–skewness model is limited to decisions from experience. Note that if we had not included the MVS decision model in the analysis, we would have concluded that the single-class CPT model outperformed EUT (and the 2-class CPT and EUT models) in predicting behavior in all four of these datasets. Consequently, our hypothesis, and the confirmation thereof, that MVS types exist is an important contribution to the existing literature. The following section will highlight another important result regarding cumulative prospect theory—in three out of the four datasets where a CPT population was found, in only one case did participants consistently exhibit the oft-advocated inverse S-shaped probability-weighting function.

As we have noted, the studies did not all implement their experiments identically. It is therefore possible that the above conclusions are confounded with some of the differences across studies. The first main difference concerns the properties of the lotteries in the various datasets. In the E_{HH} and E_{FMH} datasets, where we found a single MVS population, the first and third moments of the lotteries (calculated from the experienced relative frequencies and outcomes) are positively related, whereas in the other four datasets they are negatively related (see [Table A.6](#)). This difference in the lotteries’ properties, however, cannot fully explain the differences in the statistical models because we still find a significant MVS population in the D_{EER} and E_{EER} datasets, whose lotteries exhibit a negative mean correlation between the first and third moments. The second main difference concerns the incentive scheme. Again, although the decision from experience experiments in E_{HH} and E_{FMH} paid out all the decisions, we still found significant MVS populations in the D_{EER} and E_{EER} datasets, which only paid out a single decision. Also, recall that no wealth effects were found in the E_{HH} dataset—that is, participants treated each outcome separately. Consequently, these two variables cannot fully explain the differences

in the scope of the CPT and MVS models. We attribute the differences primarily to the complexity of the lotteries and the nature of learning, whether through description or experience.

5.2. Parameter heterogeneity (within-decision model)

In this section, we examine whether significant heterogeneity exists in individuals' parameters within a decision model. We restrict our analysis to the decision models that are present in the best model for each dataset. Hence, we present the posterior distributions of the parameters of cumulative prospect theory in datasets D_{EER} , D_{HO} , D_{HR} , and E_{EER} and the parameters of mean–variance–skewness in the D_{EER} , E_{EER} , E_{HH} , and E_{FMH} datasets. Parameters were found to be well identified in the models (see Appendix D.8 for the evidence).

5.2.1. Cumulative prospect theory

We present the posterior distributions for the CPT model in the four datasets where a nonzero proportion of participants were classified as CPT. The results are contrasted to those of existing decision from description studies⁹ that estimated the same probability-weighting function empirically using the mean or median values from non-pooled—that is, individual or hierarchical—estimation. Table 4 contains the descriptive statistics of the posterior parameter distributions, including the HPDI and the probability that the posterior distribution function of any parameter θ (conditional on the data D) is greater than 1, $p_{>1}^{\theta} = \int_1^{\infty} f(\theta|D)d\theta$. Recall, that these are the group-level distributions of each parameter and therefore $p_{>1}$ is equal to the proportion of participants whose individual parameter values are greater than 1. Fig. 2 is a contour plot of the joint distribution of the cumulative prospect theory parameters γ and δ . This figure will be useful in our discussion of whether the CPT population includes a significant proportion of EUT types. The latter implies that the joint distributions in the first column of Fig. 2 should exhibit significant density at—and be centered on—the point where the two lines intersect ($\gamma = \delta = 1$). Alternatively, this implies that the inferred probability-weighting functions in the second column of Fig. 2 should not systematically deviate from the 45° line that represents the linear weighting inherent in expected utility theory. Posterior marginal distributions of all cumulative prospect theory parameters are available in Fig. D.5 in Appendix D.9.

Result 4. Cumulative prospect theory parameters are not stable across datasets. The standard inverse-S shaped probability-weighting function is observed only in one decision from description dataset (D_{EER}). In another decision from description dataset (D_{HO}) the function is symmetrically distributed around the linear probability-weighting function and in two other datasets (D_{HR} , E_{EER}) the observed probability-weighting function is clearly inconsistent with an inverse-S shaped probability-weighting function. There are two hallmark characteristics of the standard probability-weighting function in the existing literature: the inverse-S shape ($\gamma < 1$) and a crossover point (where the probability-weighting function crosses the 45° line) significantly below 0.5 ($\delta < 1$). These two properties are consistent with the lower-left quadrants in Fig. 2. The inverse-S shape has been consistently defended using both parametric estimates (Camerer & Ho, 1994; Tversky & Kahneman, 1992) and nonparametric characterizations (Gonzalez & Wu, 1999; Wu & Gonzalez, 1996). The median estimate of γ in prior studies using the same function ranged from 0.38 to 0.96. However, our results raise doubts about the stability of this posited function in all but one of our datasets. The probability of participants exhibiting both these characteristics ($\gamma, \delta < 1$) in the three decision from description

⁹See Abdellaoui (2000), Abdellaoui, L'Haridon, and Paraschiv (2011), Abdellaoui, Vossman, and Weber (2005), Bruhin et al. (2010), Erner, Klos, and Langer (2013), Glöckner and Pachur (2012), Gonzalez and Wu (1999), Scheibehenne and Pachur (2014), Stott (2006), and Tversky and Fox (1995).

Table 4
Posterior statistics for cumulative prospect theory parameters.

		Mean	SD	Median	95% HPDI		$p_{>1}$
D_{EER}	γ	0.35	0.43	0.25	0.01	0.91	0.04
	δ	0.80	0.32	0.76	0.25	1.40	0.18
	α	0.87	0.14	0.86	0.60	1.15	0.13
	λ	1.13	0.59	1.03	0.29	2.12	0.53
	τ	1.11	0.58	1.01	0.26	2.12	
D_{HO}	γ	0.94	0.16	0.94	0.63	1.27	0.31
	δ	1.06	0.44	1.00	0.27	1.89	0.50
	α	0.42	0.19	0.37	0.14	0.82	0.02
	λ	–	–	–	–	–	–
	τ	13.54	10.95	10.70	0.60	34.72	
D_{HR}	γ	1.10	0.39	1.03	0.46	1.87	0.54
	δ	1.15	0.93	0.90	0.10	2.85	0.44
	α	0.68	0.35	0.61	0.14	1.37	0.17
	λ	1.02	0.99	0.74	0.04	2.75	0.35
	τ	5.17	5.83	3.34	0.10	15.56	
E_{EER}	γ	1.19	0.49	1.10	0.40	2.14	0.60
	δ	1.57	0.83	1.43	0.52	2.88	0.84
	α	1.26	0.63	1.13	0.31	2.50	0.60
	λ	1.89	2.53	1.11	0.02	6.07	0.54
	τ	3.76	4.07	2.62	0.10	10.60	

Note. The loss aversion parameter λ cannot be identified in D_{HO} .

datasets D_{EER} , D_{HO} , and D_{HR} is 0.79, 0.34, and 0.25, respectively. Note that in the sole decision from experience dataset where a CPT population was found, the relevant probability is also extremely low, 0.05. Consequently, only the D_{EER} dataset supports the prior findings of an inverse-S-shaped function and a crossover point below 0.5. Since this is the only dataset that involved choices between one risky (two-outcome) lottery and a sure outcome, our finding suggests that the inverse-S shape is primarily driven by the certainty effect (see Glöckner et al., 2016, for a discussion of the role that the kind of choice plays in decisions from experience). Further work should be directed towards determining whether the scope of cumulative prospect theory, and specifically an inverse-S shaped function, is limited to choice between simple lotteries with only one or two outcomes (i.e., on the boundary of the Marschak-Machina triangle; see also Wulff et al., 2018). Wilcox (2015) similarly found nonstandard (optimistic) probability-weighting functions in decision from description tasks and attributes this to context-dependent weighting based on the salience of events.

Examining the curvature of the probability-weighting function separately from the elevation, the mean/median values of γ for D_{EER} participants are 0.35/0.25 and $p_{>1}^{\gamma} = 0.96$, indicating that almost all participants had an inverse-S shaped function in this dataset. However, the evidence from the other two decision from description studies (D_{HO} , D_{HR}) stands in stark contrast. The mean/median values of γ for D_{HO} , D_{HR} are 0.94/0.94 and 1.1/1.03, respectively—these values are effectively identical to a linear function. The probability of participants exhibiting an inverse-S shaped probability-weighting function are 0.31 and 0.54 for D_{HO} and D_{HR} , respectively. Similarly, the mean/median values of γ for E_{EER} participants are 1.19/1.10 and $p_{>1}^{\gamma} = 0.60$, indicating that the majority of participants had a weakly S-shaped probability-weighting function in the sole decision from experience dataset where a CPT population was found (see Camilleri & Newell (2011) for a similar finding). Note, that since experienced relative frequencies were used for the decision from experience experiments, this finding cannot be explained by sampling error.

The elevation of the probability-weighting function captured by δ exhibited a similar pattern. Let the superscripts $+$ and $-$ refer to separate parameter estimates for the gain and loss domains, respectively. Median estimates of δ^+ in nine out of 10 decision from description studies were all lower than 1, ranging from 0.63 to 0.98, with one exception (1.40)

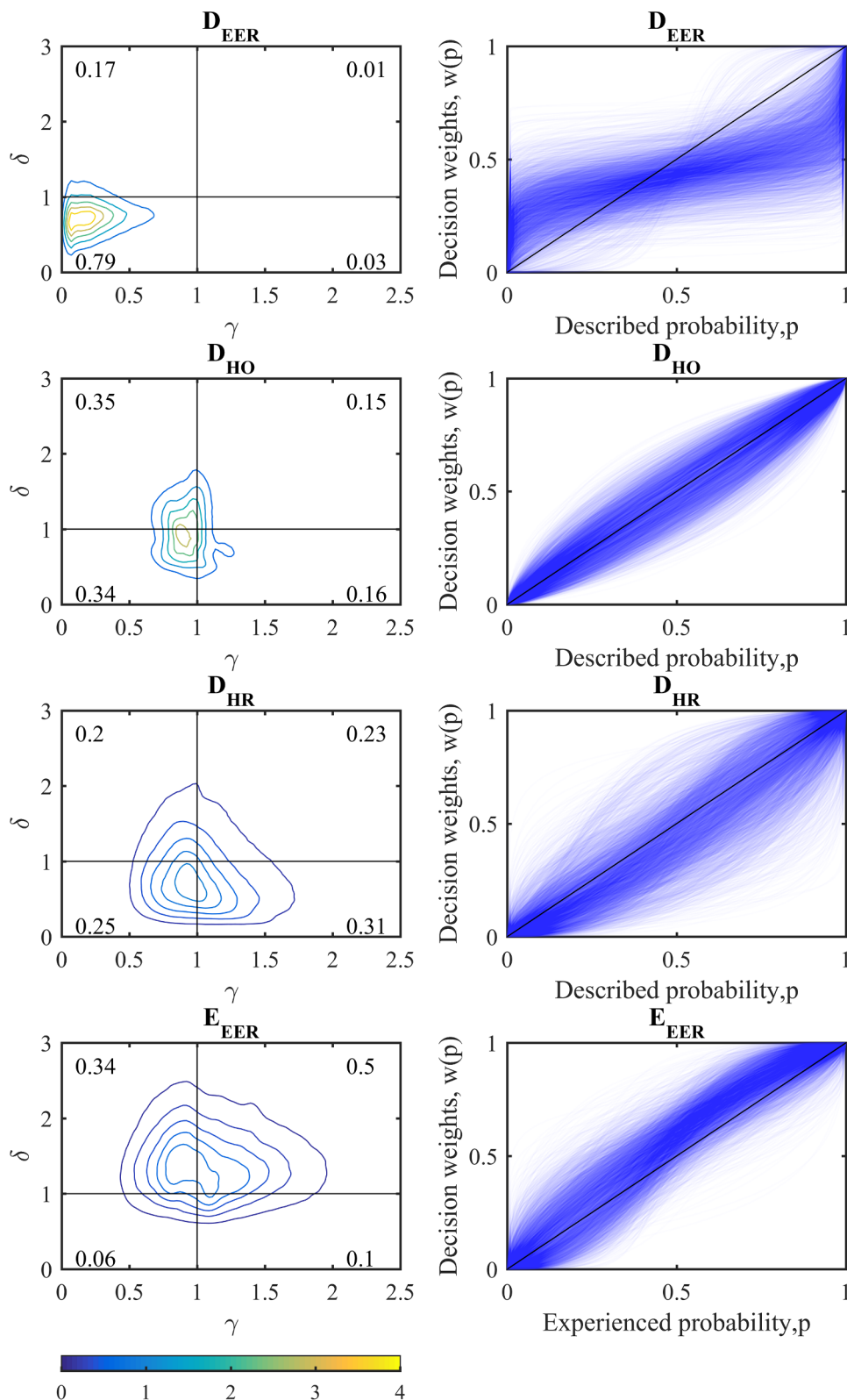


Fig. 2. Joint posterior distributions for cumulative prospect theory parameters. Note. The curvature and elevation parameters of the probability-weighting function are denoted as γ and δ , respectively. Experienced probabilities are the experienced sample-based relative frequencies. Numbers in the graphs represent the proportion of the population in each quadrant.

found in Stott (2006). For δ^- values range from 0.78 to 1.87 with four out of seven studies finding $\delta^- > 1$. The mean value in D_{EER} is 0.8, largely consistent with prior findings such as Nilsson, Rieskamp, and Wagenmakers (2011), which found a median value of 0.89 in decision

from description tasks using a one-parameter probability-weighting function. However, in D_{HO} and D_{HR} the posterior means/medians of δ are 1.06/1.0 and 1.15/0.9, respectively—these values are significantly closer to those implied by a linear function. Finally, the mean/median

values in E_{EER} are 1.57/1.43, pointing to an important, qualitatively different result.

The mean parameter value $\alpha = 0.87$ indicated mild value-function concavity in D_{EER} and 87% of participants exhibited concave value functions ($\alpha < 1$). Median estimates from eight out of nine studies using the same probability-weighting function found α ranging from 0.19 to 0.94, with the exception of Erner et al. (2013), where α was 1.15. In E_{EER} , the value function was predominantly mildly convex for 60% of the population, with a mean value of $\alpha = 1.26$. According to the findings from the description–experience gap, participants are more likely to chose the risky prospect in decisions from experience than in decisions from description. Our finding that the value function is, on average, convex in E_{EER} suggests that one mechanism for the increase in risk-averse behavior in decisions from experience can be attributed to the differential curvature of the value function, beyond any other implications for risk behavior arising from probability weighting and loss aversion. In conjunction with the (on average) S-shaped probability-weighting function that we identified above for this dataset, this leads to the reverse reflection effect, which is consistent with findings for decisions from experience.

Loss aversion ($\lambda > 1$) was found for 53% of participants in D_{EER} , while the mean/median values of λ were 1.13/1.03 indicating loss neutrality on average in the population. Similarly, the mean/median values were 1.02/0.74 in D_{HR} —only 35% of the population exhibited loss aversion. In E_{EER} , the posterior mean of $\lambda = 1.89$ but the median value was close to loss-neutrality, 1.11. Existing studies with the same probability-weighting function found values ranging from 1.05 to 2.51. Using a one-parameter probability-weighting function, Nilsson et al. (2011) recovered a median loss-aversion value of 1.02 in a hierarchical Bayesian model. Also, similarly to the D_{EER} results, only 54% of the population were loss-averse ($\lambda > 1$). Yechiam and Hochman (2013) reviewed the evidence regarding the existence of loss aversion and concluded that it is mixed (and task-dependent), and suggested that losses have other attentional, rather than weighting, effects on behavior (see also Lejarraga, Schulte-Mecklenbeck, Pachur, & Hertwig, submitted for publication).

Finally, we examine whether the CPT populations might include approximate expected utility theory types. The location and density of the joint distributions of γ and δ in Fig. 2 imply that EUT types were virtually nonexistent in D_{EER} . In the remaining datasets, the proportion of approximate EUT types would be nonzero (the highest in D_{HO}) but small as the density in the vicinity of the origin is very low. Our analysis here is an upper bound on the prevalence of EUT types as we have not simultaneously accounted for loss aversion. The exact proportion of approximate EUT types is relatively subjective, depending on how strictly one defines the boundaries of the approximation (in terms of intervals on γ , δ , and λ). However, the evidence suggests that this would be relatively low. Furthermore, as noted earlier, despite the existence of a small proportion of approximate EUT types, the EUT decision model was not able to adequately capture parameter heterogeneity across participants.

5.2.2. Mean–Variance–Skewness

The posterior statistics of the mean–variance–skewness parameters are presented in Table 5, along with the 95% HPDI and the probability that any parameter is greater than zero, $p_{>0}$. The posterior joint distributions of the β^v and β^s parameters are presented as contour plots in Fig. 3; also displayed is the proportion of participants belonging to each quadrant. The posterior marginal distributions of the MVS parameters can be found in Fig. D.6 in Appendix D.9.

Existing empirical evidence points towards variance-aversion and positive skewness-seeking as the modal preferences (see Appendix C), but considerable heterogeneity is evident. Our results are consistent with these findings. In D_{EER} , E_{EER} , and E_{HH} , the highest proportion of participants belonged to the variance-aversion/positive skewness-seeking quadrant. An exception occurs for F_{FMH} , where participants

Table 5
Posterior statistics for mean–variance–skewness parameters. ($\times 10^{-2}$)

		Mean	SD	Median	95% HPDI	$p_{>0}$	
D_{EER}	β^v	−2.99	4.46	−3.01	−11.98	5.50	0.24
	β^s	0.37	0.36	0.39	−0.44	1.10	0.88
	τ	196.64	89.51	182.54	57.77	360.86	
E_{EER}	β^v	−8.84	9.95	−10.10	−26.83	13.60	0.17
	β^s	0.88	0.79	0.91	−0.73	2.44	0.90
	τ	480.18	2513.44	223.31	0.03	1396.11	
E_{FMH}	β^v	−0.31	2.23	−0.38	−4.86	4.03	0.43
	β^s	−0.10	0.26	−0.13	−0.61	0.44	0.31
	τ	53.39	33.01	46.82	6.14	114.73	
E_{HH}	β^v	−0.11	5.75	−0.60	−10.79	12.59	0.45
	β^s	0.30	0.46	0.28	−0.56	1.20	0.73
	τ	89.17	58.64	75.47	14.90	197.22	

were most likely to belong to the variance-aversion/negative skewness-seeking quadrant. We find that across all datasets the majority of participants are variance averse (see Table 5). However, for E_{FMH} and E_{HH} the proportion of variance-averse participants was higher and close to equiprobable: 0.43 and 0.45 respectively. Furthermore, the vast majority of participants had a preference for positive skewness in the E_{EER} (0.88), D_{EER} (0.90), and E_{HH} (0.73) datasets, but not in E_{FMH} (0.31).

With respect to the possibility that the MVS population includes some approximately risk-neutral EUT types we note the following. In D_{EER} and E_{EER} , the joint posterior distribution of β^v and β^s is positioned very far from the origin ($\beta^v = \beta^s = 0$) and with very low density around this point. In E_{FMH} , the joint distribution is located closer to the origin than in the previous datasets, but the distribution is quite highly peaked. Therefore, the density drops off quite quickly to an intermediate value in the vicinity of the origin—we conclude that a small proportion of the population would be approximated reasonably well by risk-neutral expected utility theory. In E_{HH} , the distribution is located quite close to the origin, but the distribution is quite flat and spread out. Consequently, the density around the origin is not particularly high, despite it being close to the peak of the distribution. We conclude again that a small proportion of the population in E_{HH} can be interpreted as approximating a risk-neutral EUT type. We highlight that despite these results the expected utility theory decision model cannot accommodate the heterogeneity found in participants’ preferences for moments—this is why we do not find a distinct EUT population in these datasets. Finally, we note in both E_{FMH} and E_{HH} , all the lotteries were paid out to participants. This may explain why the MVS joint posterior distributions are positioned relatively closer to the origin than in E_{EER} and D_{EER} —that is, average behavior (ignoring the significant heterogeneity we found) is closer to a risk-neutral EUT type, or equivalently an expected value maximizer. This warrants further investigation in future work.

6. General discussion

Our model comparison produced the following results based on tests of out-of-sample predictive power. First, regardless of the data set and type of decisions (from description or from experience), we did not find a population of individuals behaving according to expected utility theory over income. Second, we found a mixture of CPT and MVS participants in two out of six datasets (one dataset with decisions from description and one with decisions from experience, each featuring a sure outcome and a lottery with two outcomes). Third, in the two remaining decisions from experience datasets, every single participant was classified as MVS. Fourth, in the two remaining datasets of decisions from description every single participant was classified as CPT.

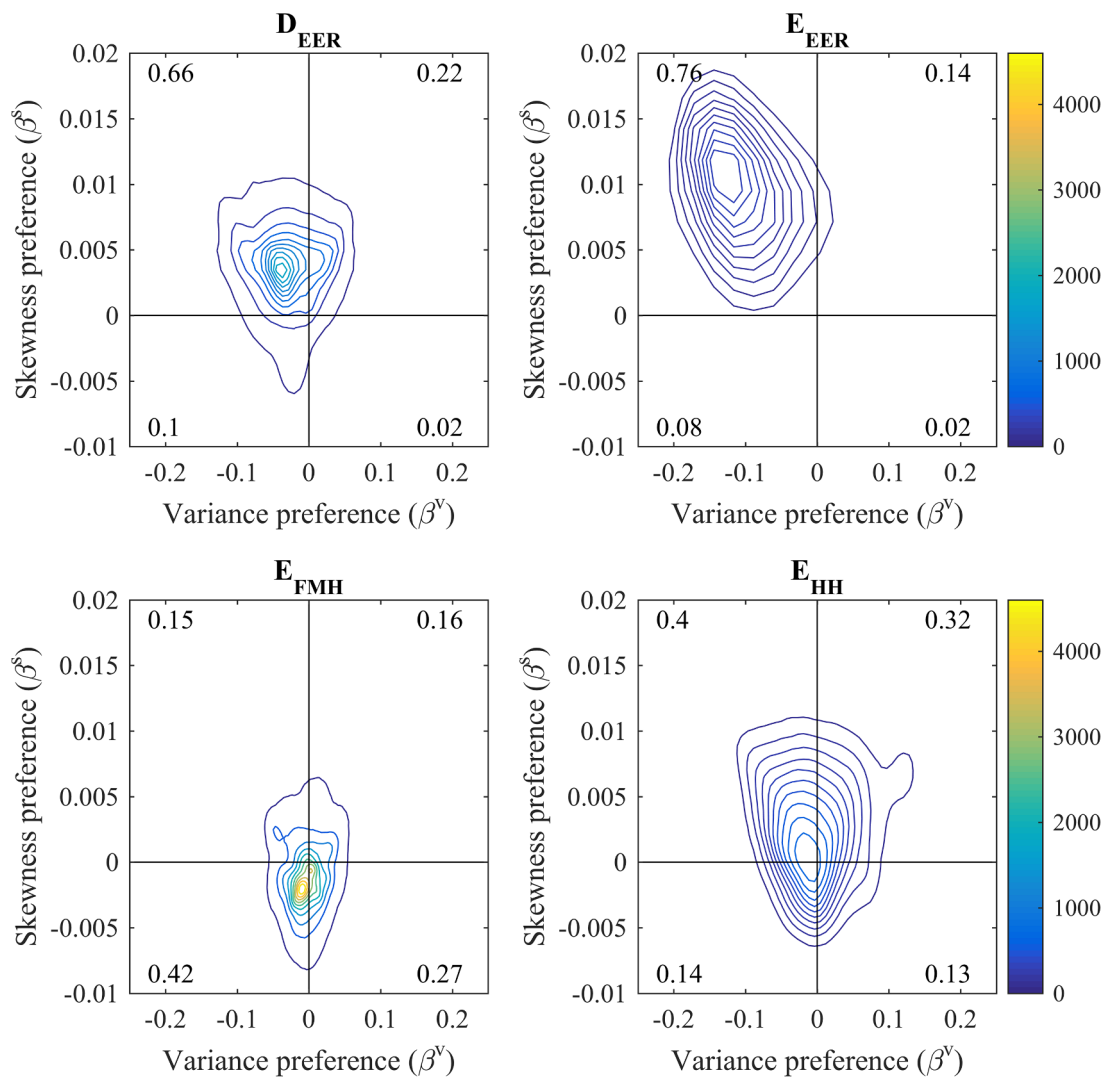


Fig. 3. Joint posterior distributions for mean–variance–skewness parameters. Note. Numbers in the graphs represent the proportion of the population in each quadrant.

Importantly, these last two results involved datasets with complex lotteries of two or three outcomes per option and no sure outcome. For such complex lotteries our findings suggest that cumulative prospect theory is better than mean–variance–skewness at predicting behavior in decisions from description, but that mean–variance–skewness is better at decisions from experience. We now discuss conceptual differences between cumulative prospect theory and mean–variance–skewness and identifiable deviations in their predictions.

6.1. Nonlinear decision weights and moment-based preferences

The intuition behind why cumulative prospect theory is more successful than expected utility theory in modeling the purchasing of insurance and gambling is the following: Relative to a linear probability-weighting function, the typical (for decisions from description) inverse-S shaped probability-weighting function overweights extreme outcomes with low probabilities and underweights extreme outcomes with high probabilities. This is consistent with positive skewness seeking, such as gambling, and negative skewness aversion, such as insurance. At the same time, the standard cumulative prospect theory value function exhibits concavity over gains and convexity over losses, thereby implicitly reducing the importance of extreme outcomes. Cumulative prospect theory’s ability to model gambling and insurance behavior therefore hinges on the relative magnitudes of the curvature of the

weighting and value functions; also, when dealing with mixed lotteries the asymmetric effect of loss aversion must be taken into account. This implies that whether cumulative prospect theory predicts positive or negative skewness seeking is conditional on the specific functional forms and parameter values used and the characteristics of the decision tasks—that is, probabilities and values of outcomes. Consequently, a single individual may exhibit both positive and negative skewness-seeking behavior. This stands in stark contrast to the stable preferences in the mean–variance–skewness model, where given parameter values β^v and β^s , a participant is always consistent in their preferences for a distribution’s moments regardless of the underlying lottery probabilities and outcome values.

The possible interactions between all the cumulative prospect theory parameters makes estimating their combined influence on skewness seeking/aversion complicated. However, Neilson and Stowe (2002) concluded that if the probability of an extreme outcome is less than 0.24, then Tversky and Kahneman (1992)’s parameter estimates can accommodate both gambling and insurance—this is the task characteristic dependence to which we alluded earlier. Therefore, decision makers facing skewed distributions with probabilities of extreme outcomes greater than 0.24 may exhibit the opposite skewness preferences than for qualitatively similarly skewed distributions that include less likely outcomes. Furthermore, they conclude that the standard functional forms for the probability-weighting and value functions are

unable to simultaneously predict both gambling (on low probability high outcomes) and Allais paradox behavior, and can lead to large risk premiums. Law and Peel (2009) also contended that cumulative prospect theory is compatible both with positive skewness-seeking and aversion conditional on the probabilities of a fixed stake lottery. Deck and Schlesinger (2010) found that using Tversky and Kahneman (1992)'s estimates in their tasks led to a preference for positive skewness. Similarly, Ågren (2006) found that for a normal inverse Gaussian distribution of returns, a CPT investor calibrated at Tversky and Kahneman (1992)'s estimates exhibits a preference for positive skewness; however, imposing a linear probability weighting leads to a preference for negative skewness. It is the documented ability of cumulative prospect theory to mimic skewness preference that leads us to argue that cumulative prospect theory should be further competitively tested against moment-based preference models.

The standard cumulative prospect theory and mean–variance–skewness models also differ in terms of their reference points—these are explicit in the former but implicit in the latter. In most empirical studies, the reference point in cumulative prospect theory—defining what constitutes a gain or a loss—is set at zero; however, endogenous reference points are an important extension (e.g., Koszegi & Rabin, 2006). Mean–variance–skewness models have an implicit reference point that determines the value around which second- and higher-order moments are calculated. This is typically set at the mean value of a distribution, hence the use of central moments of the distribution—we have also followed this convention. Consequently, in mean–variance–skewness models skewness preferences are defined relative to the mean of the distribution. However, the influence of the loss aversion parameter in cumulative prospect theory on skewness seeking/aversion is relative to an outcome of zero. Therefore, predictions of the two models can diverge conditionally on whether lotteries include mixed outcomes, or are in the domain of losses or gains only.

6.2. Conceptual issues: model parsimony, interpretability, and computational requirements

Beyond these empirical results, let us also highlight some conceptual arguments supporting the mean–variance–skewness model over cumulative prospect theory and expected utility theory. Due to the lack of an EUT population, we take cumulative prospect theory to be the main competitor to mean–variance–skewness and focus on differences between the latter two, turning to three conceptual arguments: model parsimony, interpretability, and computational requirements.

The various implementations of the cumulative prospect theory model, dependent on the parameterizations of the probability-weighting and value functions, require more free parameters than the mean–variance–skewness model. The latter has two parameters capturing the preferences for variance and skewness, compared to four parameters for a standard implementation of cumulative prospect theory. The minimum number of parameters for cumulative prospect theory would be three, if a one-parameter probability-weighting function were used. However, most modelers advise against this because such a function cannot separately capture elevation and curvature, and thus runs the risk of serious misspecification (see Chapter 7.2 in (Wakker, 2010) for a critical discussion of probability-weighting function specifications). More often than not even more parameters are estimated as modelers include separate parameters for negative and positive values (e.g., separate utility and probability-weighting function parameters for losses and gains), increasing the total number of parameters to seven.

The parameters of the mean–variance–skewness model are also more easily interpretable and do not interact to the degree that cumulative prospect theory parameters do. For example, the degree of risk aversion in cumulative prospect theory arises from the interaction of three effects: concavity/convexity of the value function, the shape of the probability weighting function, and the degree of loss aversion.

Assuming distinct parameters for gains and losses further increases the complexity. The multitude of ways of capturing various behavioral regularities also leads to difficulty in recovering parameter values as the model may not be well-identified (i.e., a flat minimum may exist where a large range of parameter values fit equally well).

It is important to emphasize that we are not arguing that cumulative prospect theory and mean–variance–skewness are models that literally capture the underlying decision processes. However, the relatively good performance of mean–variance–skewness suggests that people are sensitive to properties of the payoff distribution such as variance and skewness, which are not explicitly represented within cumulative prospect theory. Our viewpoint is more in line with the notion of decision makers as “intuitive statisticians” (Peterson & Beach, 1967), who have been reported to be surprisingly accurate when estimating statistical quantities such as proportions, means, variances, and correlations—both of samples and of populations. Moreover, preliminary neuroeconomic evidence of the independent neural encoding of both variance and skewness supports this view (Symmonds et al., 2011, 2013). Although the exact computations required to calculate higher order statistical moments can be demanding, approximate estimates of these moments (perhaps based on heuristic measures of the volatility and asymmetry of a distribution) would considerably reduce the computational burden, particularly for multi-outcome or continuous-outcome lotteries. The mean–variance–skewness model has an advantage over cumulative prospect theory if one considers the information compression achieved for multi-outcome distributions by approximately characterizing them using only their first three moments, instead of encoding all outcomes and associated probabilities, and then integrating them. Similar arguments have been put forth by d’Acromont and Bossaerts (2008) in a comparison of expected utility theory and the mean–variance preference model. Further work should be directed at empirically deriving the precise prescriptive measures of a distribution’s moments that decision makers use, and the processes required to compute them. Candidates for such simpler measures of skewness include Pearson’s mode skewness, defined as $(\text{mean} - \text{mode})/SD$, median skewness defined as $3(\text{mean} - \text{median})/SD$, or other measures based on the asymmetry of lower and upper quantiles with respect to the median.

6.3. Future directions

The mean–variance–skewness model we used—and especially the way it was updated—in decisions from experience was highly parsimonious. There are reasonable extensions or modifications to this basic model that may further boost its predictive power and should be investigated in future work. For example, in calculating the moments of the payoff distributions in decisions from experience, we assumed that all samples carry the same weight in forming the ‘experienced probabilities’. Alternatively, one can also assume that more recent observations are more heavily weighted, for both the mean–variance–skewness and cumulative prospect theory models in decisions from experience. In general, refining the model of how experienced relative frequencies are related to sampling may be warranted, as Barron and Yechiam (2009) found that participants’ reported probability assessments of a rare event’s ($p = 0.15$) likelihood from experience were on average higher than the experienced frequencies. Furthermore, the assumption that central moments—calculated relative to the mean—are the relevant measure can be relaxed. In fact, the point around which moments are calculated can be derived directly from the data. Another modification could involve standardized moments in place of the raw moments that we (and the existing literature) predominantly use. This would serve to align the mean–variance–skewness model with evidence of neuronal adaptation to the range of inputs (Wark, Lundstrom, & Fairhall, 2007). Raw statistical moments, especially of higher-order, are sensitive to outliers and estimates of these moments may deviate significantly from the population for small samples in decisions from experience. While this may be undesirable

from a normative perspective, the empirical question of whether decision makers exhibit sensitivity to outliers and small samples deserves further attention. If they do not, other measures of variability and skewness may be more appropriate. For example, Weber et al. (2004) found that comparing two measures of risk, coefficient of variation and variance, the former—which is normalized by the magnitude of the mean return—is more predictive for decisions from experience. Furthermore, as a first investigation we have modeled choices in decisions from experience based on the experienced relative frequencies, but have not modeled the sampling process itself. Reformulating mean–variance–skewness preferences as a process model including the sampling decisions would permit predictions using the objective probabilities only. Finally, it would be worthwhile to explore specifications that do not violate first-order stochastic dominance, such as the theoretical model in Blavatsky (2014b). It should be noted, however, that from a normative perspective the principle of stochastic dominance is less important in decisions from experience. This is because experienced relative frequencies will generally not be equivalent to the objective probabilities of outcomes, and therefore stochastic dominance in terms of the latter will often be violated even if individuals obey dominance for experienced relative frequencies.

In our view, the results of our model competition and the conceptual arguments constitute a serious case for the further examination of moment-based preferences in decisions under risk and uncertainty, particularly when decision makers learn probabilities and outcomes from experience rather than description. Due to our focus on comparisons between decisions from description and decisions from experience our analysis was restricted to models that can be applied to both decision contexts—further work should be directed toward more intensive model comparisons between moment-based preferences and other proposed models of decisions from experience. Furthermore, we advocate the use of a broader range of decision tasks extending beyond the typical decisions from description with lotteries of one or two possible outcomes. We found the typical inverse-S shaped probability weighting function only for such simple lotteries in decisions from description. The varying degrees of complexity of the lotteries may also explain why there are conflicting results of probability weighting even in decisions from description. van de Kuilen and Wakker (2011) summarized the state of affairs thusly: “Although we believe that inverse-S is the prevailing phenomenon [in studies with stated probabilities], it is certainly not universal” (p. 594).

We have already moved toward examining decisions with more complex lotteries of up to three outcomes each. The inclusion of tasks of

greater diversity should be further extended, particularly in the direction of continuous payoff distributions (or discrete distributions with many possible outcomes). In this case, learning the probabilities of occurrence of a multitude of outcomes will be cognitively taxing. It may turn out that summarizing such distributions by their statistical moments can be an effective cognitive representation and choice solution. Such extensions are relevant to most real-world decision making, such as asset allocation decisions, where decision makers typically learn online from experience and face a particularly complex, noisy environment with a large outcome space.

Finally, further work should be directed toward the more accurate specification of the scope of different decision models. For instance, for what types of decision problems does cumulative prospect theory capture behavior better than mean–variance–skewness (or other models) and vice versa—and why? Our findings suggest that moment-based preferences are particularly relevant for decision from experiences. By contrast, in decisions from description we classified participants primarily as CPT, indicating that this model plays an important role under this representation format. While we have unified the statistical modeling across six experimental datasets to make the results comparable, experimental implementation details still differ across these datasets. Ultimately, a more systematic exploration is required to clearly delineate the explanatory scope of cumulative prospect theory and mean–variance–skewness, and further determine the stability of model parameters for choices between lotteries with different characteristics. The practical difficulties associated with collecting new experimental data of the order of magnitude that we have used in this large-scale analysis greatly impedes such an effort. Collective effort may be necessary to overcome this problem.

Acknowledgments

The authors would like to thank (in alphabetic order) Ido Erev, Eyal Ert, Renato Frey, Glenn Harrison, Robin Hau, John Hey, Michael Lee, Rui Mata, Chris Orme, Alvin Roth, and Elisabet Rutström for providing the raw data from their experiments; Mohammed Abdellaoui, Pavlo Blavatsky, Adrian Bruhin, Andreas Ortmann, and Peter Wakker for valuable comments; Deborah Ain for editing the manuscript; and participants at the 2015 SPUDM conference and the WZB workshop on Interdisciplinary Perspectives on Decision Making.

Funding: This work was supported by the Alexander von Humboldt Foundation (Leonidas Spiliopoulos, Research Fellowship for Experienced Researchers).

Appendix A. Characteristics of the datasets

Fig. A.4 displays the distribution of the first three central moments (m_1 =mean, m_2 =variance, m_3 =skew) arising from all the lotteries (for each option) in every experimental dataset. For D_{EER} and E_{EER} only the distribution for the risky option is shown, as the safe option occurs with $p = 1$, therefore the second- and third-order moments are zero. Note, the moments for DfD datasets are calculated using the *described* probabilities, whereas for DfE datasets the *experienced* probabilities are used. From Fig. A.4, it is clear that the central moments are all of the same order of magnitude across the different experiments, rendering comparisons meaningful. The only exception is E_{EER} where the variance and skewness of the experienced outcomes are concentrated more closely near or around zero. This occurred because participants did not sample much, leading to the underrepresentation of low probability events. Since the low probability events are associated with extreme outcomes by construct, low sampling leads to an underestimation of the variance and skewness of the underlying lottery. The datasets with three outcomes for each of the two options were better at producing more variability in the experienced volatility and skewness. Ultimately, we consider datasets with two-three outcomes for each option (without a sure outcome) as significantly more informative; however, we have included the D_{EER} and E_{EER} datasets as many CPT studies have used options and outcomes similar to these. Furthermore, identifiability of the MVS model requires that the first three central moments are not strongly correlated—Table A.6 presents the correlation between moments for each option and experimental dataset. The first two moments are either uncorrelated or weakly correlated in the lotteries.¹⁰ The correlation of the second and third moments are, at worst, moderately correlated, and include both positive and negative correlations ranging from -0.56 in D_{EER} to 0.54 for the first option in E_{FMH} . Covering a wide range of possible task parameterizations/characteristics was the main driver behind our decision to use many datasets, as it imparts robustness to our results and greater generalizability.

¹⁰ The reader is also reminded that Bayesian inference can handle correlated regressors elegantly, as this will be incorporated into the joint posterior distribution of the parameters, that is, parameters will show a strong negative correlation.

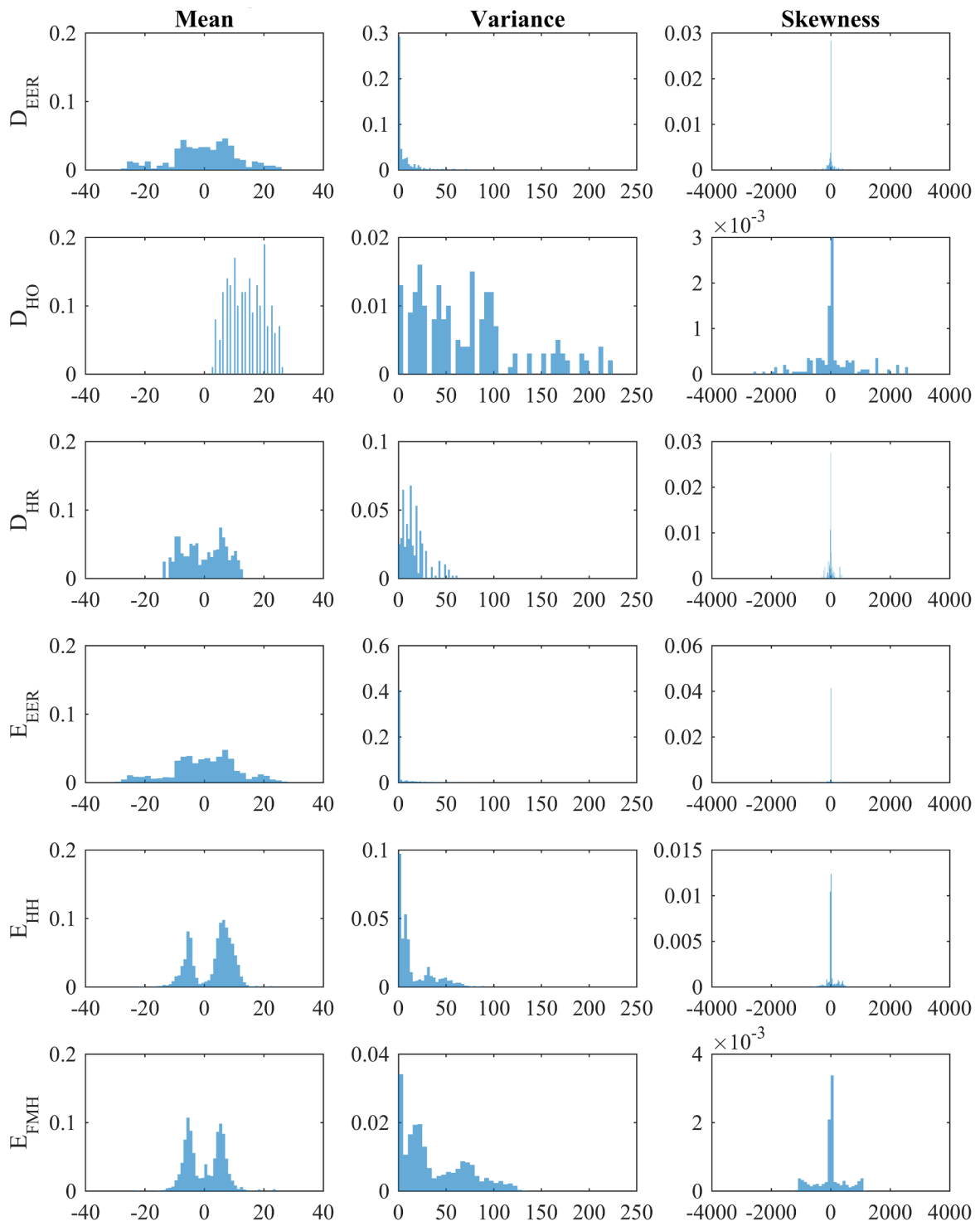


Fig. A.4. The mean, variance and skewness of lotteries in each dataset. The statistical moments are calculating using *experienced* probabilities in DFE datasets.

Table A.6
Correlation between the moments of each dataset's options

D_{EER}				D_{HO}				D_{HR}			
	m_1	m_2	m_3		m_1	m_2	m_3		m_1	m_2	m_3
m_1	1.00	-0.02	-0.40	m_1	1.00	0.03	-0.55	m_1	1.00	0.02	-0.22
m_2	-0.02	1.00	0.21	m_2	0.03	1.00	0.11	m_2	0.02	1.00	0.18
m_3	-0.40	0.21	1.00	m_3	-0.55	0.11	1.00	m_3	-0.22	0.18	1.00
E_{EER}				E_{HH}				E_{FMH}			
	m_1	m_2	m_3		m_1	m_2	m_3		m_1	m_2	m_3
m_1	1.00	-0.01	-0.22	m_1	1.00	-0.04	0.32	m_1	1.00	-0.04	0.52
m_2	-0.01	1.00	0.22	m_2	-0.04	1.00	0.25	m_2	-0.04	1.00	-0.03
m_3	-0.22	0.22	1.00	m_3	0.32	0.25	1.00	m_3	0.52	-0.03	1.00

Appendix B. A Bayesian logistic regression of choices on lotteries' statistical moments

Table B.7 presents the posterior parameter statistics of a Bayesian logistic regression of choice on the difference in the statistical moments of the lotteries. All prior distributions were set to be uninformative, $N \sim (0, 1 \times 10^6)$.

Table B.7
The relationship between choices and differences in lotteries' first three statistical moments

Dataset	Variables	Mean	Std. Dev.	MCSE	Median	95% HPDI
D_{EER}	δEV	1.0539	0.0793	0.0033	1.0522	0.9073 1.2060
	δVar	-0.0224	0.0036	0.0001	-0.0223	-0.0296 -0.0158
	$\delta Skew$	0.0060	0.0004	0.0000	0.0060	0.0052 0.0068
D_{HO}	δEV	0.5075	0.0181	0.0007	0.5080	0.4696 0.5403
	δVar	-0.0207	0.0005	0.0000	-0.0206	-0.0216 -0.0196
	$\delta Skew$	0.0002	0.0000	0.0000	0.0002	0.0001 0.0002
D_{HR}	δEV	0.8776	0.0447	0.0014	0.8777	0.7901 0.9640
	δVar	-0.0192	0.0024	0.0001	-0.0192	-0.0240 -0.0149
	$\delta Skew$	0.0009	0.0002	0.0000	0.0008	0.0005 0.0012
E_{EER}	δEV	1.2031	0.0622	0.0017	1.2027	1.0837 1.3237
	δVar	-0.0254	0.0047	0.0002	-0.0254	-0.0342 -0.0164
	$\delta Skew$	0.0016	0.0006	0.0000	0.0016	0.0005 0.0029
E_{HH}	δEV	0.5227	0.0234	0.0007	0.5212	0.4766 0.5673
	δVar	-0.0020	0.0021	0.0001	-0.0020	-0.0059 0.0023
	$\delta Skew$	0.0013	0.0002	0.0000	0.0013	0.0008 0.0018
E_{FMH}	δEV	0.3617	0.0127	0.0006	0.3623	0.3343 0.3841
	δVar	-0.0016	0.0006	0.0000	-0.0016	-0.0028 -0.0005
	$\delta Skew$	-0.0006	0.0000	0.0000	-0.0006	-0.0007 -0.0005

MCSE = Monte Carlo standard errors.

Appendix C. Skewness preferences: empirical evidence and theoretical models

We summarize below the empirical findings from the literature on preferences skewness. Symmonds et al. (2010) estimate the MVS model for sixteen participants and found that all were variance-averse and positive skewness-seeking. Similarly, Symmonds et al. (2011) estimated the model for 23 participants concluding that 70% of participants were variance-averse whereas 35% were positive skewness-seeking. Burke and Tobler (2011) and Wu et al. (2011) connect skewness with neural activity and affective responses. Burke and Tobler (2011) find based on choice data (without model fitting) that participants on average exhibited variance-aversion and positive skewness-seeking behavior. Nursimulu and Bossaerts (2014) presented participants with the opportunity to pay a price to participate in various gambles under time pressure (1s, 3s and 5s to decide whether to participate or not). They find that under time pressure, participants were variance-averse in the 1s treatment but variance-neutral in 3s and 5s treatments. Participants were positive skewness-averse in the 1s treatment and their aversion was greater in the 3s and 5s treatments. Also, Strait and Hayden (2013) find a preference for positive skewness in rhesus monkeys.

Other articles have examined skewness preferences without explicitly estimating an MVS model. Deck and Schlesinger (2010) found weak evidence of positive skewness seeking. Ebert (2015) concluded that 64% of participants' choices were positive skewness seeking compared to 77% in

Ebert and Wiesen (2011). Åstebro, Mata, and Santos-Pinto (2015) also found significant evidence of positive skewness-seeking in the laboratory, and explored whether this arises from the curvature of the utility function or of the probability weighting function. They concluded that optimism and likelihood insensitivity are the main drivers of positive skewness preference. Unlike our analysis they did not explore the possibility of moment-based preferences. In a meta-analysis of studies, Weber et al. (2004) found that positive skewness was a significant predictor of human (and animal) risk-taking behavior, and in Shafir, Bechar, and Weber (2003) map out implications for plant pollination. Grossman and Eckel (2015) devised an innovative elicitation mechanism for skewness preferences in the laboratory and found that participants were predominantly positive-skewness seeking. Furthermore, participants willingly chose to take on more risk as the outcome distribution became more positively skewed *ceteris paribus*. Coricelli (2016) related preferences for positive-skewness with aspiration levels, and disentangled the two experimentally. They concluded that aspiration levels are often used as a simple non-compensatory heuristic for choice, and that in the cases where they cannot be used, a preference for positive-skewness has strong predictive power.

The existence of positive skewness-seeking preferences is also supported in studies that remain model-agnostic regarding the specific decision model but classify behavior via revealed preferences over lottery pairs. The concept of prudence, although model independent, is related to skewness preferences in the context of our MVS model (Eeckhoudt & Schlesinger, 2006). In a utility framework, prudence is characterized by a positive third derivative of the utility function, that is, with a preference for positive skewness. Deck and Schlesinger (2010) found that 61% of participants' responses were consistent with prudence overall; however, this increased to 70% for higher stakes. Ebert and Wiesen (2011) reported that 65% of participants' choices were prudent, whereas for Ebert and Wiesen (2014) prudent choices were even more prevalent (88%). Maier and Rieger (2011) found that in the domain of gains (losses) that 60% (55%) of choices were prudent. Tarazona-Gomez (2005) found that 63% of choices were prudent. Åstebro et al. (2015) also found a preference for positive skew on average as participants were willing to expose themselves to more risk in return for greater positive skewness.

On the theoretical front, Blavatsky (2014b) provided an axiomatization of an extension to EUT that relates disappointment and elation—arising from the difference between *ex post* realizations and *ex ante* expectations of the lottery—to measures of deviation and skewness of lottery outcomes. Interestingly, this model does not violate first-order stochastic dominance for a wide range of parameterizations, unlike mean-variance preferences—see Borch (1969) for an early discussion and also Cox (2008) and Johnstone and Lindley (2013). Blavatsky (2014a) proposed a new 2-parameter probability weighting function for CPT, relating its curvature and elevation to preferences for the second- and third-order *L*-moments of lotteries.

Appendix D. A detailed presentation of the Bayesian hierarchical model

D.1. Advantages of hierarchical models

There are numerous advantages to using a hierarchical model compared to a set of individual models. Hierarchical models can efficiently capture the effects of groups or clusters that exhibit correlation within them, thereby violating an assumption of independence. For example, multiple observations of a single participant may be correlated if the same decision model (and parameterization) is used across observations, that is, there exists within-participant correlation. At a higher level, participants that belong to a specific group (in this case decision model or latent class) may also exhibit between-participant (but within-group) correlation in their parameterizations of the decision model. Consequently, parameter values for one individual may be informative for other individuals in the same group. A hierarchical approach involves specifying a distribution for each parameter from which individuals' parameters are drawn from, that is, it specifies the group/population level of the distribution. Therefore, the introduction of an appropriate hierarchical structure constrains individual parameters in a beneficial manner, leading to a simpler model. This occurs because individuals' parameters are constrained by the group distribution of said parameter; extreme individual values are pulled in closer to areas of higher density. This stands in contrast to individual modeling where extreme individual parameter values may occur even for large datasets (Wilcox, 2007). Ultimately, hierarchical modeling can be viewed as a compromise between pooled and individual modeling, with the trade off between the two being mediated by the data and the relevance of the hierarchical structure assumed by the modeler.

D.2. Existing models of heterogeneous decision makers

In recent work, Bruhin et al. (2010) and Harrison and Rutström (2009) allowed only for model heterogeneity using mixtures of Prospect Theory and EUT, and CPT and EUT classes respectively, with point estimates of parameters within each decision model. Nilsson et al. (2011) allowed only for parameter heterogeneity within CPT. Stahl (2014) re-examined the D_{HO} and D_{HR} datasets but permitted only for parameter heterogeneity in the Anticipated Utility Model of Quiggin (1982). He concluded that once heterogeneity is accounted for, EUT did a good job at predicting behavior. Parameter heterogeneity was modeled in von Gaudecker, van Soest, and Wengström (2011) using a random coefficients specification, however they estimated only a single Krep-Porteus model, that is, they did not include model heterogeneity. Scheibehenne and Pachur (2014) allowed for parameter heterogeneity using a Bayesian hierarchical approach but they implemented CPT and the transfer-of-attention-exchange models separately, that is, they did not allow for a mixture of these two models. The closest paper to ours is Conte et al. (2011), who allowed for two classes of decision models EUT and RDEU, whilst also allowing for parameter heterogeneity for some, but not all parameters. However, there are still many points of departure. Our analysis followed a Bayesian inference approach (with parameter heterogeneity for *all* parameters) instead of a frequentist (maximum likelihood estimation) approach, included the MVS model, and analyzed both DfD and DfE in six datasets compared to one DfD dataset.

D.3. Choice of CPT model parameters

We present the empirical findings regarding the use of different parameters for gains and losses. Nilsson et al. (2011) conducted a parameter-recovery study and found that the estimation of separate risk-parameters α for gains and losses leads to problems in recovering the loss-aversion estimate λ . Also, Booij, van Praag, and van de Kuilen (2010) could not reject the hypothesis that the curvature of the value function was the same in

gain and loss domains. Constraining α to be the same in both domains restricts the value function to be either S- or inverse S-shaped, that is, global concavity or convexity is excluded. While some have argued that there exists significant heterogeneity in the value function parameters in both domains, we note that existing evidence is consistent with participants not exhibiting globally concave or convex value functions. Abdellaoui (2000, Table 3) non-parametrically categorized forty participants according to the shape of their value functions—only four participants were globally concave and none were globally convex.

Regarding the probability-weighting function similar arguments apply. The curvature parameter of the two-parameter function is usually not found to be significantly different in the loss and gain domains (see Table 1 in Booi et al., 2010, which collects the parameter estimates across many studies). On the other hand, the elevation parameter is often found to be larger in the loss domain, see Abdellaoui (2000), Abdellaoui et al. (2005) and Fehr-Duda, de Gennaro, and Schubert (2006). However, the studies that found differences in the probability-weighting function in the gain and loss domains did so using in-sample estimation, but not cross-validation. This renders their findings open to the possibility that these results were due to overfitting; however, settling this is ultimately an empirical question. In our minds, we are not yet convinced that such complex specifications are necessary. Glöckner and Pachur (2012) compared various specifications of CPT with parameter restrictions in terms of cross-validation performance (across two different points in time) and concluded that simpler CPT specifications are adequate—specifically they argue that separate estimates are generally not required for the gain/loss domains. Finally, were we to decide to implement CPT with different parameters in the gain and loss domains, then with the same reasoning we should also extend the MVS model in a similar direction. Instead, given the above evidence, we choose to implement each model with the least complex parameterization possible that retains the important theoretical properties of each model.

D.4. Specification of priors

We specify weakly-informative prior distributions of the parameters of each individual decision model: CPT (Eq. (2)), EUT (Eq. (3)), and MVS (Eq. (4)). The prior distribution of the best response parameter for an individual i classified as using decision model c_i , is denoted by τ_i . The prior mean and standard deviation of the normally-distributed variable $\log(\tau_i)$ are drawn from identical distributions for each decision model, c_i ; note, the posterior distribution of the best response parameter is free to differ across decision models.

$$\log(\tau_i) \sim N(\mu_\tau, \sigma_\tau) \quad \mu_\tau \sim U(-3, 1) \quad \sigma_\tau \sim U(0, 1)$$

The priors for CPT parameters were modeled as log-normal distributions to respect the lower support of zero for all these parameters. The bounds on the uniform distribution of the mean of these parameters was chosen so that the maximum mean value of $e^1 \approx 2.71$ is sufficiently large to encompass prior information from existing DfD and DfE studies. The upper bound on the prior of the standard deviation of the distributions was set to the same value as the maximum mean value; therefore the 95% range of the distribution is at least approximately four times the mean. In datasets without mixed lotteries (D_{HO} , E_{HH} , E_{FMH}), the loss-aversion parameter was set to $\lambda = 1$. Similarly, the same log-normal prior distribution was chosen for the risk-aversion parameter in EUT.

$$\log(\gamma_i) \sim N(\mu_\gamma, \sigma_\gamma) \quad \mu_\gamma \sim U(-3, 1) \quad \sigma_\gamma \sim U(0, 1) \tag{9}$$

$$\log(\delta_i) \sim N(\mu_\delta, \sigma_\delta) \quad \mu_\delta \sim U(-3, 1) \quad \sigma_\delta \sim U(0, 1) \tag{10}$$

$$\log(\rho_i) \sim N(\mu_\rho, \sigma_\rho) \quad \mu_\rho \sim U(-2, 1) \quad \sigma_\rho \sim U(0, 1) \tag{11}$$

$$\log(\lambda_i) \sim N(\mu_\lambda, \sigma_\lambda) \quad \mu_\lambda \sim U(-3, 1) \quad \sigma_\lambda \sim U(0, 1) \tag{12}$$

$$\log(\alpha_i) \sim N(\mu_\alpha, \sigma_\alpha) \quad \mu_\alpha \sim U(-2, 1) \quad \sigma_\alpha \sim U(0, 1) \tag{13}$$

Weakly informative priors for the MVS model were scaled according to the magnitude of the lottery moments. Since the mean value of a lottery has a (normalized) parameter coefficient of one in Eq. (4), and the mean and variance of the payoffs in Fig. A.4 are of approximately the same order of magnitude throughout, we choose a maximum absolute value of 1 for β_i^v . Using the same reasoning but observing that skewness values are approximately one order of magnitude larger, we appropriately scaled the prior values for β_i^s . Note, that the results confirm the appropriateness of the imposed scaling on the priors as the posterior distributions were concentrated well within the upper and lower bounds assumed in their priors.

$$\beta_i^v \sim N(\mu_{\beta^v}, \sigma_{\beta^v}) \quad \mu_{\beta^v} \sim U(-1.0, 1.0) \quad \sigma_{\beta^v} \sim U(0, 1.0) \tag{14}$$

$$\beta_i^s \sim N(\mu_{\beta^s}, \sigma_{\beta^s}) \quad \mu_{\beta^s} \sim U(-0.1, 0.1) \quad \sigma_{\beta^s} \sim U(0, 0.1) \tag{15}$$

D.5. Model selection criteria

We performed model selection using a cross-validation (CV) criterion (based on average likelihood) since other common techniques such as DIC (Spiegelhalter, Best, & Carlin, 2002) are problematic for selection between mixture models (Celeux et al., 2003). In contrast to information criteria that specify a penalty for the complexity of a model, CV focuses instead on the observable effect of increased complexity, that is, overfitting leading to poor predictive ability. As a result, CV automatically accounts not only for the degrees of freedom, but also the parametric flexibility resulting from the specific functional form and the resulting parameter interactions. We perform k -fold cross-validation ($k = 3$) so that all the data observations are used exactly once to determine the predictive ability of the models. The observations from each dataset were randomly assigned to one of three subsets. Each model was computed three times using two of these folds as the estimation sample and the remaining fold as the prediction sample. The CV criterion, \overline{lcV} was then computed over the resulting three prediction samples. Ultimately, in any empirical study involving model comparisons we are concerned with two types of significance with respect to difference between models: statistical and economic (or s -significance and e -significance). Although model comparisons using CV most often ignore the question of whether models are s -significantly different from each

Table D.8
Model performance per dataset

		D_{EER}	D_{HO}	D_{HR}	E_{EER}	E_{HH}	E_{FMH}
π [95% HPDI]	CPT	0.55 [0.50,0.60]	1	1	0.76 [0.70,0.82]	0	0
	MVS	0.45 [0.40,0.50]	0	0	0.24 [0.19,0.31]	1	1
	EUT	0	0	0	0	0	0
$\overline{lc\bar{v}}$	{CPT,MVS,EUT}	<i>0.675</i>	<i>0.712</i>	<i>0.643</i>	<i>0.758</i>	0.729	0.682
	{CPT,MVS}	<i>0.675^{*,†}</i>	<i>0.712</i>	<i>0.643[*]</i>	<i>0.759^{*,†}</i>	<i>0.730</i>	<i>0.683</i>
	{CPT,EUT}	0.665	<i>0.712</i>	<i>0.641</i>	<i>0.746</i>	0.709	0.678
	{MVS,EUT}	0.642	0.700	0.640	0.741	0.728	<i>0.683</i>
	{CPT}	0.665	<i>0.712^{*,†}</i>	<i>0.641[†]</i>	0.748	0.709	0.678
	{MVS}	0.641	0.657	0.626	0.744	<i>0.728^{*,†}</i>	<i>0.682^{*,†}</i>
	{EUT}	0.561	0.694	0.618	0.713	0.696	0.661
$s.e.(\overline{lc\bar{v}})$		0.001	0.0002	0.0013	0.004	0.006	0.0031
Hit rate	{CPT,MVS,EUT}	0.762	0.780	0.713	0.833	0.797	0.750
	{CPT,MVS}	0.762 [†]	0.78	0.714	0.832	0.795	0.752
	{CPT,EUT}	0.751	0.780	0.711	0.830	0.784	0.748
	{MVS,EUT}	0.739	0.768	0.712	0.833	0.800	0.750
	{CPT}	0.751	0.780 [†]	0.712 [†]	0.829	0.780	0.748
	{MVS}	0.740	0.732	0.699	0.834	0.798 [†]	0.755 [†]
	{EUT}	0.672	0.764	0.685	0.838 [†]	0.783	0.736

Models that did not converge according to the \hat{R} -statistic due to partial non-identifiability of EUT and CPT classes are reported in gray
Italics indicate the best performing model for each dataset.

* and † Indicate the model chosen according to s -significance and e -significance respectively.

other, we employ the following technique.

Since $\overline{lc\bar{v}}$ is a random variable we account for the variance that arises from different ways of partitioning the dataset into training and cross-validation sets. We use the *one standard error rule* (Hastie, Tibshirani, & Friedman, 2001) that selects the most parsimonious model within one standard error of the model with the highest $\overline{lc\bar{v}}$.¹¹ The choice of e -significance is often quite subjective when an external standard does not exist. We believe that a reasonable standard is to consider models whose $\overline{lc\bar{v}}$ differs by more than 0.01 as e -significantly different. This corresponds to at least a 1% point difference in the predictive ability of a model. Therefore, e -significance selects the most parsimonious model whose $\overline{lc\bar{v}}$ is within 0.01 units of the highest $\overline{lc\bar{v}}$ of all models. Note, that the standard error of $\overline{lc\bar{v}}$ for all models ranged from 0.0002 to 0.006, therefore in all cases the e -significance criterion was the stricter criterion of the two. Consequently, we focus on the latter but present both in the interest of transparency.

D.6. Model selection results

Table D.8 presents the statistical results for all six datasets and all latent-class combinations of the decision models including the Highest Posterior Density Intervals (HPDI) of relevant parameters. Numbers in italics correspond to the model with the highest (best) cross-validation criterion ($\overline{lc\bar{v}}$) for each dataset. We have also included the models' hit rates¹² in Table D.8 as an alternative performance measure of the predictive power of the models. Numbers superscripted with a * correspond to the best model as selected by the s -significance criterion, and numbers superscripted with a † correspond to the best model as selected by the e -significance criterion.

Some models did not converge according to the \hat{R} -statistic. We have determined that in *all* cases this occurred because of the existence of a subset of individuals in CPT populations that behaved similarly to EUT, rendering the two relatively non-identifiable.¹³ While measures of fit or cross-validation performance are still reliable in this case, the posterior distributions of the proportions of the different model classes are not. In *all* cases, using the same model but excluding the EUT latent class solved the problem with *no change* in the cross-validation performance. This is exactly what we would expect to find if the non-convergence was driven from the non-identifiability of a subset of the latent classes rather than other causes. We clearly identify the non-convergent models in the Results section and discuss extensively the possibility of a minority of EUT types existing in a CPT population.

¹¹ The standard error of $\overline{lc\bar{v}}$ for $k = 3$ is calculated as $\sqrt{\frac{\text{var}(lc\bar{v}_1, lc\bar{v}_2, lc\bar{v}_3)}{3}}$.

¹² The hit rate calculates the proportion of correct predictions made by assuming that the model *deterministically* predicts the most likely action. Consequently, the hit rates exclude information regarding the strength of preference as captured by the exact predicted choice probability. Therefore, we use the $\overline{lc\bar{v}}$ performance criterion for model selection as it is more informative.

¹³ Recall that CPT nests EUT for specific parameter values. Therefore, if the posterior distribution of the CPT parameters exhibit significant non-zero density around these parameter values, some participants will be approximately behaving as EUT types.

Table D.9
Proportion of participants classified at various levels of probability

Classification probability >	0.99	0.95	0.9	0.8	0.7	0.6
D_{EER}	0.48	0.58	0.70	0.85	0.93	0.95
E_{EER}	0.3	0.50	0.50	0.70	0.80	0.98

Table D.10
Spearman rank correlation coefficients for CPT parameters

D_{EER}					D_{HO}			
	δ	α	λ	τ		δ	α	τ
γ	0.10	0.11	0.09	0.22	γ	-0.05	0.04	0.03
δ	0.00	0.00	0.05	0.09	δ	0.00	-0.02	-0.11
α	0.00	0.00	-0.04	-0.15	α	0.00	0.00	-0.80
λ	0.00	0.00	0.00	-0.07				

D_{HR}					E_{EER}				
	δ	α	λ	τ		δ	α	λ	τ
γ	-0.02	0.02	-0.02	-0.02	γ	-0.02	0.05	0.05	0.10
δ	0.00	0.08	0.00	-0.08	δ	0.00	0.10	0.00	-0.01
α	0.00	0.00	-0.08	-0.46	α	0.00	0.00	-0.09	-0.01
λ	0.00	0.00	0.00	-0.09	λ	0.00	0.00	0.00	0.03

Table D.11
Spearman rank correlation coefficients for MVS parameters

D_{EER}			E_{EER}			E_{HH}			E_{FMH}		
	β^s	τ		β^s	τ		β^s	τ		β^s	τ
β^v	-0.04	0.05	β^v	-0.18	0.01	β^v	-0.05	-0.03	β^v	0.11	-0.08
β^s	0.00	0.00	β^s	0.00	-0.04	β^s	0.00	-0.09	β^s	0.00	-0.16

D.7. Participant classification

If the postulated latent classes are well separated and the statistical technique is efficient at recovering them, then participants should be classified to each decision model with a high probability. Table D.9 presents the statistics on the proportion of participants that were classified as using a particular decision model for different levels of probability (i.e., the degree of certainty with which we assign participants to decision models). Overall, participants are classified with a high probability to one of the latent-classes; 48% [58%, 93%] and 30% [50%, 80%] of participants were classified with a probability greater than 0.99 [0.9, 0.7] in D_{EER} and E_{EER} respectively. We conclude that significant model heterogeneity existed in the EER experiments both for DfD and DfE and the heterogeneity is best described by a combination of CPT and MVS decision makers.

D.8. Parameter identification

We first check the identifiability of these models by calculating the Spearman rank correlations between each decision models’ parameters from their joint (bivariate) posterior distributions—these are presented in Tables D.10 and D.11. With the exception of two cases, the magnitude of all Spearman rank correlation coefficients were less than 0.25 (and usually less than 0.1) indicating strong identification. We find that the parameters α and τ are only weakly identified in D_{HO} and D_{HR} as they exhibited high correlations of -0.8 and -0.46 respectively. Note, Broomell and Bhatia (2014) find that the best response parameter τ is often recovered poorly in CPT models.

D.9. Posterior marginal distributions of CPT and MVS parameters

See Figs. D.5 and D.6.

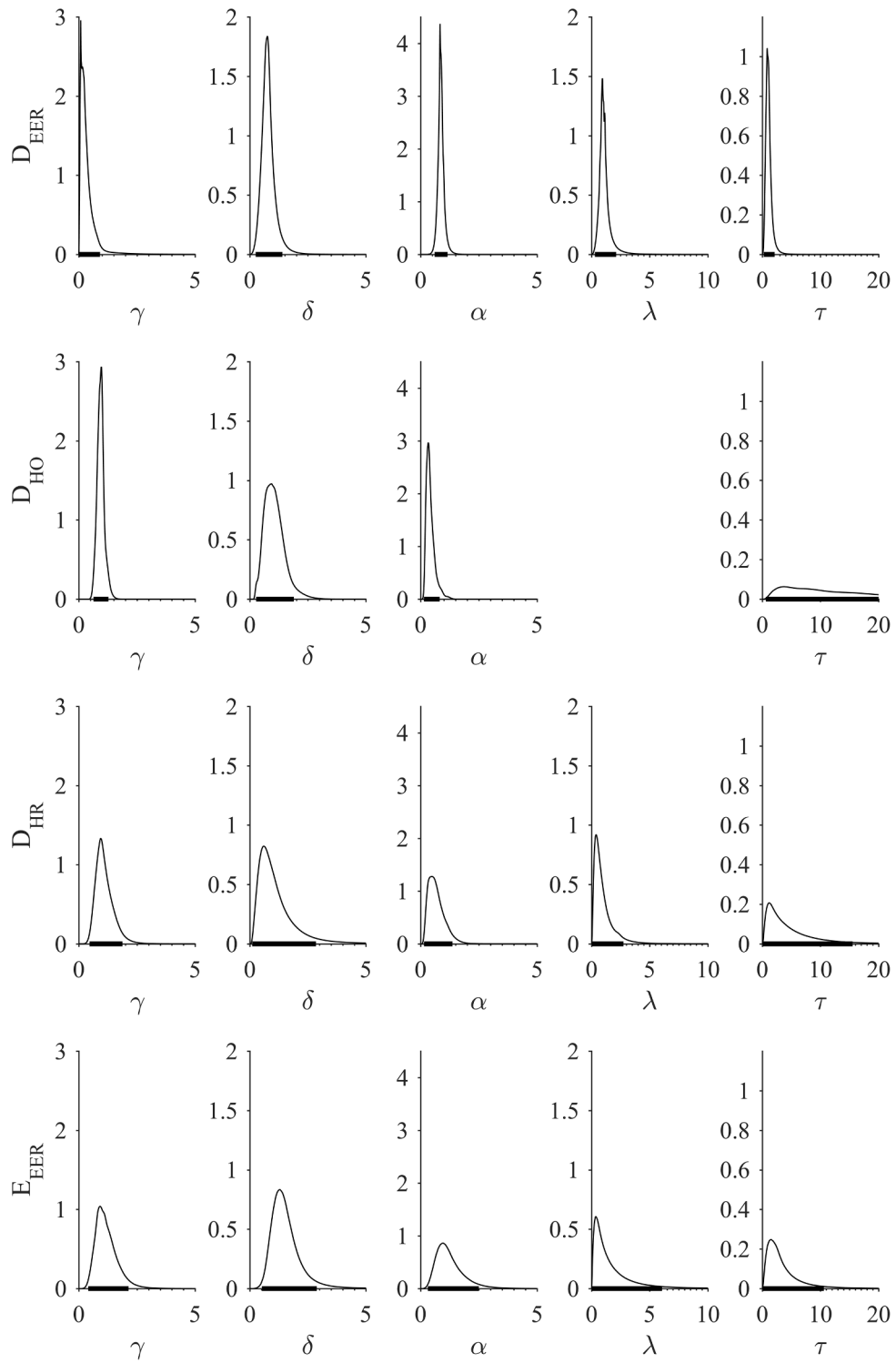


Fig. D.5. Marginal posterior distributions for CPT. Solid bars indicate the 95% HPDI range.

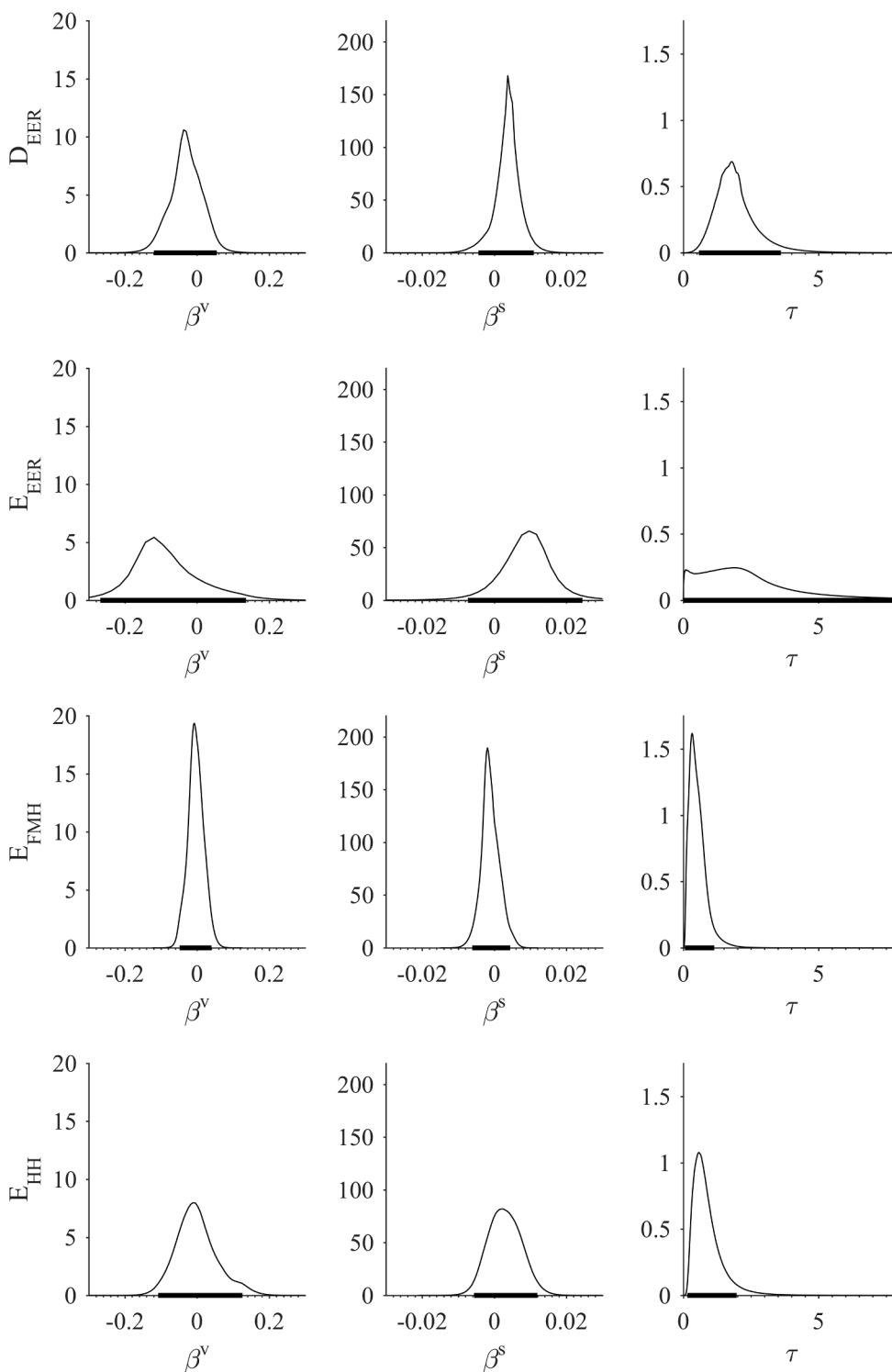


Fig. D.6. Marginal posterior distributions for MVS. Solid bars indicate the 95% HPDI range.

Appendix E. Comparison of model performance in E_{EER} with the Erev et al. (2010) tournament models

We compare the best performing 2-class model in E_{EER} with the best model from the Erev et al. (2010) tournament in the relevant E-sampling condition. In the tournament, the models only predicted the proportion of choices (across all participants) for each lottery pair; that is, in contrast to our models they did not predict at the individual (participant and lottery pair) level. However, we can aggregate our individual estimates to construct the predicted choice proportions for each lottery pair. We note that there are still differences in how we estimate the proportions. For example, we used k -fold cross validation and therefore have estimates for the whole dataset, whereas Erev et al. (2010) separated the dataset into estimation and prediction sets. However, since the assignment was done randomly, we would expect on average the same results. Our comparison is meant to show that the models we have employed here have approximately the same predictive power as the models in the tournament, which included sampling models.

We find that our 2-class CPT and MVS model in E_{BER} performs on par with the best model found in Erev et al. (2010) (see their Table 3b). For the first criterion, *Pagree*, the best tournament model (the Ensemble model) achieved a rate of 83% compared to our model’s performance of 82%. In terms of the second criterion, *r*, the Ensemble model achieved a correlation of 0.80 compared to a correlation of 0.76 for our model. Finally, in terms of the third criterion, *MSD*, our model (0.007) significantly outperforms the Ensemble model (0.019) significantly.¹⁴ We conclude that our models, whilst not explicitly process- or sampling-based, perform on par with the most competitive models in the Erev et al. (2010) tournament.

Appendix F. The E_{HH} experiment

F.1. Tasks

A total of 30 decision problems were presented in a randomized order to each participant. Each problem consists of two lotteries with three outcomes each, summarized in Table F.12. There are 10 prototypes of problems, half in the gain domain (#1–5), half in the loss domain (#6–10). From these prototypes, additional problems were created by adding a constant to all outcomes. In problems 4, 5, 9 and 10, one of the gambles contains a rare event (10% chance of occurrence). The rare event is desirable in problems 4 and 9, and undesirable in problems 5 and 10. The lotteries vary significantly in the magnitude of the first-, second-, and third-order moments, and the sign of the odd-moments; the latter is useful for testing differences in skewness preference. Recall that these are the *objective* characteristics of the lotteries. However, due to sampling from experience the *experienced* characteristics will differ, inducing even greater variability in the subjective characteristics, particularly for higher-order moments.

Table F.12
Characteristics of the lotteries in the E_{HH} experiment

Nr.	Lottery A	Lottery B	A_{mean}	A_{var}	A_{skew}	B_{mean}	B_{var}	B_{skew}	$p(A)$
1a	(3, .25; 8, .50; 11, .25)	(0, .20; 1, .60; 15, .20)	7.50	8.25	-12.00	3.60	32.64	276.43	85.0%
1b	(4, .25; 9, .50; 12, .25)	(1, .20; 2, .60; 16, .20)	8.50	8.25	-12.00	4.60	32.64	276.43	77.5%
1c	(5, .25; 10, .50; 13, .25)	(2, .20; 3, .60; 17, .20)	9.50	8.25	-12.00	5.60	32.64	276.43	80.0%
1d	(6, .25; 11, .50; 14, .25)	(3, .20; 4, .60; 18, .20)	10.50	8.25	-12.00	6.60	32.64	276.43	78.8%
2a	(3, .25; 4, .50; 20, .25)	(0, .20; 6, .60; 8, .20)	7.75	50.19	406.41	5.20	7.36	-23.42	73.8%
2b	(4, .25; 5, .50; 21, .25)	(1, .20; 7, .60; 9, .20)	8.75	50.19	406.41	6.20	7.36	-23.42	66.3%
2c	(5, .25; 6, .50; 22, .25)	(2, .20; 8, .60; 10, .20)	9.75	50.19	406.41	7.20	7.36	-23.42	70.0%
2d	(6, .25; 7, .50; 23, .25)	(3, .20; 9, .60; 11, .20)	10.75	50.19	406.41	8.20	7.36	-23.42	70.0%
3a	(7, .25; 8, .25; 10, .50)	(0, .25; 1, .25; 12, .50)	8.75	1.69	-0.47	6.25	33.19	-2.16	77.5%
3b	(8, .25; 9, .25; 11, .50)	(1, .25; 2, .25; 13, .50)	9.75	1.69	-0.47	7.25	33.19	-2.16	67.5%
3c	(9, .25; 10, .25; 12, .50)	(2, .25; 3, .25; 14, .50)	10.75	1.69	-0.47	8.25	33.19	-2.16	72.5%
3d	(10, .25; 11, .25; 13, .50)	(3, .25; 4, .25; 15, .50)	11.75	1.69	-0.47	9.25	33.19	-2.16	77.5%
4a	(0, .50; 5, .40; 20, .10)	(0, .30; 4, .40; 8, .30)	4.00	34.00	378.00	4.00	9.60	0.00	50.0%
4b	(1, .50; 6, .40; 21, .10)	(1, .30; 5, .40; 9, .30)	5.00	34.00	378.00	5.00	9.60	0.00	42.5%
4c	(2, .50; 7, .40; 22, .10)	(2, .30; 6, .40; 10, .30)	6.00	34.00	378.00	6.00	9.60	0.00	60.0%
4d	(3, .50; 8, .40; 23, .10)	(3, .30; 7, .40; 11, .30)	7.00	34.00	378.00	7.00	9.60	0.00	57.5%
5a	(0, .10; 3, .60; 9, .30)	(2, .30; 5, .30; 6, .40)	4.50	9.45	16.20	4.50	2.85	-3.30	45.0%
5b	(1, .10; 4, .60; 10, .30)	(3, .30; 6, .30; 7, .40)	5.50	9.45	16.20	5.50	2.85	-3.30	60.0%
5c	(2, .10; 5, .60; 11, .30)	(4, .30; 7, .30; 8, .40)	6.50	9.45	16.20	6.50	2.85	-3.30	53.8%
5d	(3, .10; 6, .60; 12, .30)	(5, .30; 8, .30; 9, .40)	7.50	9.45	16.20	7.50	2.85	-3.30	47.5%
6a	(-4, .25; -5, .50; -6, .25)	(-1, .25; -7, .50; -21, .25)	-5.00	0.50	0.00	-9.00	54.00	-300.00	83.8%
6b	(-5, .25; -6, .50; -7, .25)	(-2, .25; -8, .50; -22, .25)	-6.00	0.50	0.00	-10.00	54.00	-300.00	87.5%
7a	(0, .25; -7, .50; -8, .25)	(-4, .25; -5, .50; -16, .25)	-5.50	10.25	36.00	-7.50	24.25	-135.00	80.0%
7b	(-1, .25; -8, .50; -9, .25)	(-5, .25; -6, .50; -17, .25)	-6.50	10.25	36.00	-8.50	24.25	-135.00	66.3%
8a	(-2, .50; -6, .25; -8, .25)	(0, .50; -12, .25; -16, .25)	-4.50	6.75	-3.75	-7.00	51.00	-42.00	72.5%
8b	(-3, .50; -7, .25; -9, .25)	(-1, .50; -13, .25; -17, .25)	-5.50	6.75	-3.75	-8.00	51.00	-42.00	72.5%
9a	(0, .10; -3, .60; -9, .30)	(-2, .30; -5, .30; -6, .40)	-4.50	9.45	-16.20	-4.50	2.85	3.30	55.0%
9b	(-1, .10; -4, .60; -10, .30)	(-3, .30; -6, .30; -7, .40)	-5.50	9.45	-16.20	-5.50	2.85	3.30	48.8%
10a	(0, .50; -5, .40; -20, .10)	(0, .30; -4, .40; -8, .30)	-4.00	34.00	-378.00	-4.00	9.60	0.00	41.3%
10b	(-1, .50; -6, .40; -21, .10)	(-1, .30; -5, .40; -9, .30)	-5.00	34.00	-378.00	-5.00	9.60	0.00	46.3%

F.2. Participants

A total of 80 respondents (67 female) from different disciplines at the University of Basel participated in the experiment. Their mean age was 24.2 years (s.d. = 6.0 years). The experiment was comprised of two distinct parts, the data we analyze here corresponds to the first part of the experiment. The second part of the experiment involved continuous distributions in DfE; we did not include this as a similar experiment in DfD was not available to permit comparisons. Participants received either 15 Swiss Francs as a show-up fee or—in the case of psychology students—course credits, *plus* the earnings that resulted from all of their choices. Given participants’ choices, the mean of the distribution of the expected performance payments to participants (for the first part of the experiment) was 9.56 Swiss Francs with a minimum of 6.69 Francs and a maximum of 10.8 Francs.¹⁵ No participant in the study incurred more losses than gains from the incentivized choices. Total payoffs from the experiment were

¹⁴ However, the poorer performance of the Ensemble model will be in part due to the fact that it modeled proportions of choices rather than individual behavior. Consequently, our modeling allowed for more accurate inference by accounting for participant heterogeneity, which is important for the *MSD* criterion, but not as important for *Pagree* and *r* since they ignore some information. The former is a directional or qualitative measure which uses a cutoff proportion of 0.5 to determine the agreement, and the latter uses rank order information.

¹⁵ Due to a program failure, the lottery realizations performed at the end of the experiment and the documentation of the actual payment per person were not recorded. We therefore present the mean expectation given each participant’s observed choices as an indicator of the magnitude of the experimental incentives. Note, the dispersion of payments presented above underestimates the true dispersion, as the latter would also include the variability induced by the lottery realizations.

approximately equal to or greater than the minimum wage in Swiss francs. In the DfE literature, it is common to incentivize all of a participant's choices rather than a single choice to encourage participants to sample sufficiently from the distributions. However, this introduces the possibility of wealth effects as—on average—participants' expected accumulated wealth should be increasing. Note, that participants were not aware of the outcome of the lottery after each trial and therefore could at best only estimate their expected, not their realized, change in wealth. We believe that this should serve to mediate any wealth effects, but to confirm this we specifically test the hypothesis that the order of presentation did not affect choices. We perform logit regressions of the binary choice variable on two independent variables, presentation order and its square, separately for each problem. The null hypothesis of no order effects corresponds to a joint $\chi^2(2)$ test that the coefficients of the independent variables are both zero. Out of the thirty regressions we rejected the null hypothesis of no order effects at the 5% significance level only once—this is within the expected number of chance rejections ($30 \times 0.05 = 1.5$). Note, that these tests have adequate power as for each problem we have 80 (participant) observations. We conclude that wealth effects did not significantly affect choices.

F.3. Instructions

Welcome and many thanks for participating in this experiment!

In this experiment, you will make money-related decisions. Please do your best to make the *best decisions possible*. To make the situation realistic, the payment you will receive will depend on the outcomes of your decisions; you will be paid at a ratio of 10 to 1 Swiss francs.

You can considerably influence the *payment* you receive through the *decisions* you make.

The experiment consists of *two parts*, separated by a short break.

During the experiment, you will *not receive any feedback* on the outcomes of your decisions, as this might influence your further decisions. You will get this information when you are paid at the end of the experiment. Note that you will not make a loss. If you lose more money than you win, you will simply not receive a payment. However, the likelihood of that happening is very low.

Press ENTER to move on to the instructions for the first part of the experiment.

Part 1

In part 1 you will see two computer-generated stacks of cards. The cards show sums of money (wins or losses *in Swiss Francs*; losses are indicated by a minus sign). The various cards are not necessarily equally represented in the two stacks. You have to decide from which of the two stacks to draw a card at random. The money amount on the card will determine your payment.

Each trial begins with a *learning phase*, during which you can gather information about the two stacks. You can draw and look at cards one by one, from either stack. These draws have no effect on your winnings. The stack does *not* change through your drawing cards. In other words, the cards in a stack are *always the same*, irrespective of what and how often you have already drawn from it. Please draw as many cards as you need to in order to be confident of making a good decision.

Press ENTER to continue.

When you have completed the learning phase, you come to the *decision*. Here, you decide from which stack to draw a card at random. If the card shows a win, we credit that amount to your account. If the card shows a loss (indicated by a minus sign), we deduct that amount from your winnings. As mentioned earlier, you will not immediately be told which card you have drawn. You will not find out about the outcomes of your decisions until the end of the experiment.

Procedure

The experiment begins with a *warm-up trial* that has no effect on your winnings. You will then make **30 decisions** in which you have a chance to win money. Each trial begins with a learning phase that lasts until you are ready to make a decision. After this decision, you will move on to two new stacks of cards.

Please ask the test administrator if you have any questions, otherwise press ENTER to continue.

Appendix G. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cognition.2018.10.023>.

References

- Ågren, M. (2006). Prospect theory and higher moments. *Working Paper, Department of Economics, Uppsala University* (pp. 1–37).
- Åstebro, T., Mata, J., & Santos-Pinto, L. (2015). Skewness seeking: Risk loving, optimism or overweighting of small probabilities? *Theory and Decision*, 78(2), 189–208. <https://doi.org/10.1007/s11238-014-9417-4>.
- Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11), 1497–1512. <https://doi.org/10.1287/mnsc.46.11.1497.12080>.
- Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: do we need two prospect theory specifications? *Management Science*, 57(10), 1879–1895. <https://doi.org/10.1287/mnsc.1110.1368>.
- Abdellaoui, M., Vossman, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science*, 51(9), 1384–1399. <https://doi.org/10.1287/mnsc.1050.0388>.
- Albuquerque, R. (2012). Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. *Review of Financial Studies*, 25(5), 1630–1673. <https://doi.org/10.1093/rfs/hhr144>.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine [rational man's behavior in face of risk: critique of the American school's postulates and axioms]. *Econometrica*, 21(4), 503–546. <https://doi.org/10.2307/1907921>.
- Arditti, F. D. (1967). Risk and the required return on equity. *The Journal of Finance*, 22(1), 19–36. <https://doi.org/10.1111/j.1540-6261.1967.tb01651.x>.
- Barberis, N., & Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, 98(5), 2066–2100. <https://doi.org/10.1257/aer.98.5.2066>.
- Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3), 215–233. <https://doi.org/10.1002/bdm.443>.
- Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making*, 4, 447–460. Retrieved from <<http://search.proquest.com/openview/43fcd118011e530645ac143e5f9cec8/1?pq-origsite=gscholar&amc=696407>>.
- Blavatskyy, P. (2014a). A probability weighting function for cumulative prospect theory and mean-gini approach to optimal portfolio investment. *Mimeo* (pp. 1–32). Retrieved from <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2380484>.
- Blavatskyy, P. (2014b). A theory of decision-making under risk as a tradeoff between expected utility, expected utility deviation and expected utility skewness (pp. 1–34). doi:<https://doi.org/10.2139/ssrn.2505828>. <<http://ssrn.com/abstract=2401325>>.
- Blavatskyy, P. R. (2014c). Stronger utility. *Theory and Decision*, 76(2), 265–286. <https://doi.org/10.1007/s11238-013-9366-3>.
- Booij, A. S., van Praag, B. M. S., & van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, 68(1–2), 115–148. <https://doi.org/10.1007/s11238-009-9144-4>.
- Borch, K. (1969). A note on uncertainty and indifference curves. *The Review of Economic Studies*, 36(1), 1–4. <https://doi.org/10.2307/2296336?ref=no-x-route:83e59889e7c7ed6c8b4b6671bf537d3d>.

- Brockett, P. L., & Kahane, Y. (1992). Risk, return, skewness and preference. *Management Science*, 38(6), 851–866. <https://doi.org/10.1287/mnsc.38.6.851>.
- Broomell, S. B., & Bhatia, S. (2014). Parameter recovery for decision modeling using choice data. *Decision*, 1(4), 252–274. <https://doi.org/10.1037/dec000020>.
- Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4), 1375–1412. <https://doi.org/10.3982/ECTA7139>.
- Burke, C. J., & Tobler, P. N. (2011). Reward skewness coding in the insula independent of probability and loss. *Journal of Neurophysiology*, 106(5), 2415–2422. <https://doi.org/10.1152/jn.00471.2011>.
- Camerer, C. F., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2), 167–196. <https://doi.org/10.1007/BF01065371>.
- Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review*, 18(2), 377–384. <https://doi.org/10.3758/s13423-010-0040-2>.
- Celeux, G., Forbes, F., Robert, C. P., & Titterton, D. M. (2003). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4), 651–674. <http://books.google.gr/books?id=TIkkPwAACAAJ&dq=intitle:Deviance+Information+Criteria+for+Missing+Data+Modelshl=&cd=1&source=gbs_api>.
- Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, 15(4), 692–712. <https://doi.org/10.3758/PBR.15.4.692>.
- Conte, A., Hey, J. D., & Moffatt, P. G. (2011). Mixture models of choice under risk. *Journal of Econometrics*, 162(1), 79–88. <https://doi.org/10.1016/j.jeconom.2009.10.011>.
- Coricelli, G., Diecidue, E., & Zaffuto, F. D. (2016). Aspiration levels and preference for skewness in choice under risk. INSEAD. *Working Paper Series* (pp. 1–34). Retrieved from <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2767382>.
- Costa-Gomes, M. A., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: an experimental study. *Econometrica*, 69(5), 1193–1235. <https://doi.org/10.1111/1468-0262.00239>.
- Cox, J. C., & Sadiraj, V. (2006). Small- and large-stakes risk aversion: implications of concavity calibration for decision theory. *Games and Economic Behavior*, 56(1), 45–60. <https://doi.org/10.1016/j.geb.2005.08.001>.
- d'Acremont, M., & Bossaerts, P. (2008). Neurobiological studies of risk assessment: A comparison of expected utility and mean-variance approaches. *Cognitive, Affective, Behavioral Neuroscience*, 8(4), 363–374. <https://doi.org/10.3758/CABN.8.4.363>.
- Deck, C., & Schlesinger, H. (2010). Exploring higher order risk effects. *The Review of Economic Studies*, 77(4), 1403–1420. <https://doi.org/10.1111/j.1467-937X.2010.00605.x>.
- Ebert, S. (2015). On skewed risks in economic models and experiments. *Journal of Economic Behavior & Organization*, 112, 85–97. <https://doi.org/10.1016/j.jebo.2015.01.003>.
- Ebert, S., & Wiesen, D. (2011). Testing for prudence and skewness seeking. *Management Science*, 57(7), 1334–1349. <https://doi.org/10.1287/mnsc.1110.1354>.
- Ebert, S., & Wiesen, D. (2014). Joint measurement of risk aversion, prudence, and temperance. *Journal of Risk and Uncertainty*, 48(3), 231–252. <https://doi.org/10.1007/s11166-014-9193-0>.
- Eeckhoudt, L., & Schlesinger, H. (2006). Putting risk in its proper place. *American Economic Review*, 96(1), 280–289.
- Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., ... Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1), 15–47. <https://doi.org/10.1002/bdm.683>.
- Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, 111(3), 10818–10825. <https://doi.org/10.1073/pnas.1402846111>.
- Erner, C., Klos, A., & Langer, T. (2013). Can prospect theory be used to predict an investor's willingness to pay? *Journal of Banking & Finance*, 37(6), 1960–1973. <https://doi.org/10.1016/j.jbankfin.2012.12.008>.
- Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, 12(3), 403–408. <https://doi.org/10.3758/BF03193784>.
- Fehr-Duda, H., de Gennaro, M., & Schubert, R. (2006). Gender, financial risk, and probability weights. *Theory and Decision*, 60, 283–313. <https://doi.org/10.1007/s11238-005-4590-0>.
- Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: age differences emerge as a function of choice set size. *Cognition*, 142, 60–80. <https://doi.org/10.1016/j.cognition.2015.05.004>.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511.
- Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General*, 145(4), 486–508. <https://doi.org/10.1037/a0040103>.
- Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1), 21–32. <https://doi.org/10.1016/j.cognition.2011.12.002>.
- Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2), 236–254. <https://doi.org/10.1037/0033-295X.94.2.236>.
- Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, 118(4), 523–551. <https://doi.org/10.1037/a0024558>.
- Gonzalez, C., Lerch, J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4), 591–635. [https://doi.org/10.1016/S0364-0213\(03\)00031-4](https://doi.org/10.1016/S0364-0213(03)00031-4).
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129–166.
- Grossman, P. J., & Eckel, C. C. (2015). Loving the long shot: Risk taking with skewed lotteries. *Journal of Risk and Uncertainty*, 51(3), 195–217. <https://doi.org/10.1007/s11166-015-9228-1>.
- Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2), 433–458. <https://doi.org/10.1111/j.1468-0262.2006.00753.x>.
- Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, 12, 133–158.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., & Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5), 469–485. <https://doi.org/10.1080/14697681003756877>.
- Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3), 1263–1295. <https://doi.org/10.1111/0022-1082.00247>.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Data Mining, Inference, and Prediction, 2nd ed. <https://doi.org/10.1007/978-0-387-84858-7>.
- Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, 23(1), 48–68. <https://doi.org/10.1002/bdm.665>.
- Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21, 493–518. <https://doi.org/10.1002/bdm.598>.
- Heinemann, F. (2008). Measuring risk aversion and the wealth effect. In J. C. Cox, & G. W. Harrison (Eds.), *Risk aversion in experiments*. Bingley: Emerald. Retrieved from <<http://www.worldcat.org/title/risk-aversion-in-experiments/oclc/680038705>>.
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539. Retrieved from <<http://ps.sagepub.com/content/15/8/534.short>>.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12), 517–523. <https://doi.org/10.1016/j.tics.2009.09.004>.
- Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6), 1291–1326. <https://doi.org/10.2307/2951750>.
- Johnstone, D., & Lindley, D. (2013). Mean-variance and expected utility: The borch paradox. *Statistical Science*, 28(2), 223–237. <https://doi.org/10.1214/12-STS408>.
- Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston, MA: Hart, Schaffner, Marx; Houghton Mifflin Co. Retrieved from <http://scholar.google.com/scholar?q=related:6yZ5o_SXBCj:scholar.google.com/&hl=en&as_sdt=1,5&as_ylo=1921&as_yhi=1921>.
- Koszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4), 1133–1165. Retrieved from <<http://qje.oxfordjournals.org/content/121/4/1133.short>>.
- Kraus, A., & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance*, 31(4), 1085–1100. <https://doi.org/10.2307/2326275>.
- L.A.T., Cox, Jr. (2008). Why risk is not variance: An expository note. *Risk Analysis*, 28(4), 925–928. <https://doi.org/10.1111/j.1539-6924.2008.01062.x>.
- Law, D., & Peel, D. A. (2009). Skewness as an explanation of gambling in cumulative prospect theory. *Applied Economics*, 41, 685–689. <https://doi.org/10.1080/00036840601007476>.
- Lejarraga, T., Schulte-Mecklenbeck, M., Pachur, T., Hertwig, R. (2018). Beneath loss aversion lies invariant loss vigilance. Manuscript submitted for publication.
- Link, W. A., & Eaton, M. J. (2012). On thinning of chains in mcmc. *Methods in Ecology and Evolution*, 3(1), 112–115. <https://doi.org/10.1111/j.2041-210X.2011.00131.x>.
- Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*, 43(2), 286–313. <https://doi.org/10.1006/jmps.1999.1259>.
- Maier, J., & Rieger, M. (2011). Experimental evidence on higher order risk preferences with real monetary losses. *Chapman University (ESI), Working Paper 15-10* (pp. 1–37).
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.2307/2975974>.
- Neilson, W., & Stowe, J. (2002). A further examination of cumulative prospect theory parameterizations. *The Journal of Risk and Uncertainty*, 24(1), 31–46.
- Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1), 84–93. <https://doi.org/10.1016/j.jmp.2010.08.006>.
- Nursimulu, A. D., & Bossaerts, P. (2014). Risk and reward preferences under time pressure. *Review of Finance*, 18(3), 999–1022. <https://doi.org/10.1093/rof/rft013>.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29–46. Retrieved from <<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=6046307&retmode=ref&cmd=prlinks>>.
- Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, 143(5), 2000–2019. <https://doi.org/10.1037/xge0000013>.
- Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, 122(4), 621–647. <https://doi.org/10.1037/a0039413>.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4), 323–343. [https://doi.org/10.1016/0167-2681\(82\)90008-7](https://doi.org/10.1016/0167-2681(82)90008-7).
- Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5), 1281–1292. <https://doi.org/10.1111/1468-0262.00158>.
- Rakow, T., & Demes, K. A. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, 106(2), 168–179. Retrieved

- from <<http://www.sciencedirect.com/science/article/pii/S0749597808000125>> .
- Regenwetter, M., & Robinson, M. M. (2017). The construct-behavior gap in behavioral decision research: a challenge beyond replicability. *Psychological Review*, 124(5), 533–550. <https://doi.org/10.1037/rev0000067>.
- Rubinstein, M. E. (1973). The fundamental theorem of parameter-preference security valuation. *The Journal of Financial and Quantitative Analysis*, 8(1), 61–69. <https://doi.org/10.2307/2329748>.
- Scheibehenne, B., & Pachur, T. (2014). Using bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, 22(2), 391–407. <https://doi.org/10.3758/s13423-014-0684-4>.
- Shafir, S., Bechar, A., & Weber, E. U. (2003). Cognition-mediated coevolution- context-dependent evaluations and sensitivity of pollinators to variability in nectar rewards. *Plant Systematics and Evolution*, 238, 195–209. <https://doi.org/10.1007/s00606-003-0280-y>.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.2307/2977928>.
- Spiegelhalter, D. J., Best, N. G., & Carlin, B. P. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64(4), 583–639. <https://doi.org/10.1002/9780470035948>.
- Spiliopoulos, L. (2012). Pattern recognition and subjective belief learning in a repeated constant-sum game. *Games and Economic Behavior*, 75(2), 921–935. <https://doi.org/10.1016/j.geb.2012.01.005>.
- Stahl, D. O. (2014). A re-examination of experiment data on lottery choices. *Mimeo*. Retrieved from <https://laits.utexas.edu/~stahl/Reexam_of_EUT_Data.pdf> .
- Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2), 101–130. <https://doi.org/10.1007/s11166-006-8289-6>.
- Strait, C. E., & Hayden, B. Y. (2013). Preference patterns for skewed gambles in rhesus monkeys. *Biology Letters*, 9(6), 20130902. <https://doi.org/10.1098/rsbl.2013.0902>.
- Symmonds, M., Bossaerts, P., & Dolan, R. J. (2010). A behavioral and neural evaluation of prospective decision-making under risk. *Journal of Neuro science*, 30(43), 14380–14389. <https://doi.org/10.1523/JNEUROSCI.1459-10.2010>.
- Symmonds, M., Moran, R. J., Wright, N. D., Bossaerts, P., Barnes, G., & Dolan, R. J. (2013). The chronometry of risk processing in the human cortex. *Frontiers in Neuroscience*, 7, 1–11. <https://doi.org/10.3389/fnins.2013.00146>.
- Symmonds, M., Wright, N. D., Bach, D. R., & Dolan, R. J. (2011). Deconstructing risk: Separable encoding of variance and skewness in the brain. *NeuroImage*, 58(4), 1139–1149. <https://doi.org/10.1016/j.neuroimage.2011.06.087>.
- Tarazona-Gomez, M. (2005). Are individuals prudent? an experimental approach using lottery choices. *Lerna-Ehess Working Paper* (pp. 1–52).
- Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25(2), 65–86. <https://doi.org/10.2307/2296205>.
- Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2), 269–283. <https://doi.org/10.1037//0033-295X.102.2.269>.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323. <https://doi.org/10.1007/BF00122574>.
- Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, 20(4), 473–479. Retrieved from <<http://pss.sagepub.com/content/20/4/473.short>> .
- van de Kuilen, G., & Wakker, P. P. (2011). The midweight method to measure attitudes toward risk and ambiguity. *Management Science*, 57(3), 582–598. <https://doi.org/10.1287/mnsc.1100.1282>.
- von Gaudecker, H.-M., van Soest, A., & Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, 101(2), 664–694. <https://doi.org/10.1257/aer.101.2.664>.
- Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge. Retrieved from <<http://k2s.cc/file/ac3aeff614848>> .
- Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, 17, 423–429. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0959438807000840>.
- Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, 111(2), 430–445. <https://doi.org/10.1037/0033-295X.111.2.430>.
- Wilcox, N. T. (2007). Predicting risky choices out-of-context: A monte carlo study. *Working Paper, Department of Economics, University of Houston*.
- Wilcox, N. T. (2015). Unusual estimates of probability weighting functions. *Chapman University (ESI), Working Paper 15–10* (pp. 1–56). doi:<https://doi.org/10.13140/RG.2.1.4014.0008>.
- Wilcox, N. T., & Wilcox, N. T. (2011). 'stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162, 89–104. <https://doi.org/10.1016/j.jeconom.2009.10.012>.
- Wu, C. C., Bossaerts, P., & Knutson, B. (2011). The affective impact of financial skewness on neural activity and choice. *PLoS ONE*, 6(2), 1–7. <https://doi.org/10.1371/journal.pone.0016838.g001>.
- Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42(12), 1676–1690. Retrieved from <<http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=SrcAuth=mekentosj&SrcApp=Papers&DestLinkType=FullRecord&DestApp=WOS&KeyUT=A1996WA59100004>> .
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2), 140–176. <https://doi.org/10.1037/bul0000115>.
- Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, 139(2), 497–518. <https://doi.org/10.1037/a0029383>.