Journal of Mathematical Psychology 55 (2011) 84–93

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Hierarchical Bayesian parameter estimation for cumulative prospect theory

Håkan Nilsson [a,b,*], Jörg Rieskamp [a], Eric-Jan Wagenmakers [c]

[a] University of Basel, Basel, Switzerland
[b] Uppsala University, Uppsala, Sweden
[c] University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 17 February 2010
Received in revised form
26 July 2010
Available online 18 September 2010

Keywords:
Decision making
Loss aversion
Maximum likelihood
Shrinkage
WinBUGS

ABSTRACT

Cumulative prospect theory (CPT Tversky & Kahneman, 1992) has provided one of the most influential accounts of how people make decisions under risk. CPT is a formal model with parameters that quantify psychological processes such as loss aversion, subjective values of gains and losses, and subjective probabilities. In practical applications of CPT, the model's parameters are usually estimated using a single-participant maximum likelihood approach. The present study shows the advantages of an alternative, hierarchical Bayesian parameter estimation procedure. Performance of the procedure is illustrated with a parameter recovery study and application to a real data set. The work reveals that without particular constraints on the parameter space, CPT can produce loss aversion without the parameter that has traditionally been associated with loss aversion. In general, the results illustrate that inferences about people's decision processes can crucially depend on the method used to estimate model parameters.

1. Hierarchical Bayesian parameter estimation for cumulative prospect theory

Should I cross the road now or wait until that oncoming car has passed? Should I buy a house or rent an apartment? Should I sell my shares, await market developments, or perhaps buy more shares? The consequences of real-life decisions are often unpredictable and risky, and the study of how people make decisions under risk is a central topic in psychology and economics.

The standard approach for studying decision making under risk is to let people choose between gambles. Each gamble leads to different monetary outcomes x_i that occur with probability p_i. For instance, a participant might be given a choice between two options: gamble A, which yields $100 with probability .6 and −$100 with .4 (typically written as [$100, .6; −$100, .4]); and gamble B, which yields −$10 with probability .5 and $0 with probability .5 (i.e., [−$10, .5; $0, .5]). How do people make choices between such gambles? The economic view of rational decision making assumes that when choosing between risky options such as gambles, people simply prefer the option that maximizes the expected utility (e.g., Savage, 1954; Von Neumann & Morgenstern, 1947). In the above example, *Homo economicus* could calculate that the expected payoff equals $20 for gamble A and −$5 for gamble B. Hence, gamble A is clearly more attractive than gamble B.[1]

The *expected utility* perspective on human decision making has been challenged by a substantial body of psychological research. This research inspired a variety of alternative accounts, including regret and disappointment theory (Bell, 1982, 1985; Loomes & Sugden, 1982), the priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006), the transfer-of-attention exchange model (Birnbaum, 2008; Birnbaum & Chavez, 1997), the decision field theory (Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001), the weighted utility theory (e.g. Fishburn, 1983), the proportional difference model (González-Vallejo, 2002), the decision affect theory (Mellers, 2000), the dual system model of preference under risk (Mukherjee, 2010), and the rank-dependent expected utility theories (e.g., Quiggin, 1982; for a summary of alternatives to expected utility see Camerer, 1989; Machina, 1989; Rieskamp, Busemeyer, & Mellers, 2006; Starmer, 2000).

Here we focus on the by far the most influential alternative to expected utility: *prospect theory* (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Prospect theory suggests that people put subjective weights on values and probabilities and that people weight values and probabilities associated with positive outcomes (i.e., gains) differently from those associated with negative outcomes (i.e., losses). Prospect theory successfully accounts for a number of systematic deviations from expected utility, such as the

* Corresponding address: Department of Psychology, Uppsala University, Box 1225, SE-751 42 Uppsala, Sweden.
E-mail addresses: hakan.nilsson@psyk.uu.se, hakan.nilsson@unibas.ch (H. Nilsson).

[1] This example assumes that money has a linear utility. An arguably more realistic example will be provided later.

0022-2496/$ – see front matter
doi:10.1016/j.jmp.2010.08.006

finding that people are risk-averse for gains of high probability but risk-seeking for gains of low probability, and that people are risk-seeking for losses of high probability but risk-averse for losses of low probability (the so-called fourfold pattern of risk; e.g., Tversky & Kahneman, 1992).

Prospect theory has been instantiated as a formal mathematical model, featuring several parameters that isolate and quantify psychological concepts such as loss aversion, subjective value functions for gains and losses, and probability weighting functions. Despite the prominence of prospect theory as an account of how people make decisions under risk, little work has addressed how to estimate the model's parameters. The most common approach is single-subject maximum likelihood estimation (MLE; e.g., Harrison & Rutström, 2009; Harless & Camerer, 1994; Rieskamp, 2008; Stott, 2006; for a tutorial see Myung, 2003). In general, single-subject MLE is consistent and efficient as the number of observations per participant grows large. One drawback of single-subject MLE is that it treats participants as if they were completely independent. Often, though, it is more reasonable to assume that participants are similar to each other, such that their individual parameter values originate from a group-level distribution. This way, parameter estimation for a single participant could be assisted by information obtained from all other participants. Another drawback of single-subject MLE is that it is almost always used to obtain only point estimates for each subject; these point estimates are then entered in an analysis of variance to test for significant differences between groups or conditions. This procedure ignores the precision with which the participant parameters are estimated. As we will discuss later, some parameters in cumulative prospect theory tend to be estimated with relatively high uncertainty, and for these parameters single-subject maximum likelihood point estimates may be extreme and unrealistic.

In the present article we propose a hierarchical Bayesian parameter estimation procedure for cumulative prospect theory that addresses the above limitations of MLE. Our estimation procedure implements a natural compromise between the two extremes of complete independence (i.e., as in single-subject MLE) and complete pooling (i.e., fitting averaged data and implicitly assuming that all participants are identical). Moreover, the Bayesian hierarchical procedure prevents inference from being dominated by a few outlying point estimates — extreme results will "shrink" towards the group mean, a phenomenon that is more pronounced for parameters that are estimated with much uncertainty. Our Bayesian estimation routine is implemented in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; Lunn, Spiegelhalter, Thomas, & Best, 2009), making it easy for other researchers to apply and adjust.

The outline of this article is as follows. The first section briefly outlines the mathematical and conceptual basis that underlies cumulative prospect theory. The second section critically reviews the present approaches for estimating cumulative prospect theory parameters, and the third section introduces our hierarchical Bayesian estimation procedure for cumulative prospect theory. The fourth section contains a parameter recovery study that features both single-subject MLE and our Bayesian hierarchical method. The fifth section applies these two estimation techniques to experimental data drawn from Rieskamp (2008).

2. Cumulative prospect theory

Prospect theory (Kahneman & Tversky, 1979) was formulated as a reaction to the growing evidence that people do not follow the norms of economic expected utility maximizing (for reviews see Rieskamp et al., 2006; Schoemaker, 1982; Starmer, 2000). Classic economic theory states that when people choose between options with risky outcomes they will choose the option that maximizes their expected utility. The expected utility (EU) of an option O is defined by

$$EU(O) = \sum p_i u(x_i), \tag{1}$$

where $u(\cdot)$ is a utility function that defines the subjective utility of x_i. The subjective utility of x_i is weighted by the probability with which outcome i occurs.

Imagine a choice between the two gambles introduced above, [\$100, .6; −\$100, .4] (gamble A) and [−\$10, .5; \$0, .5] (gamble B). Assume that $u(\cdot)$ is a power function so that $EU(O) = \sum p_i(x_i)^\alpha$ and that $\alpha = .5$, then the expected utility of gamble A equals 2 and the expected utility of gamble B equals −1.6. In this case, utility maximization implies choosing gamble A over gamble B.

Prospect theory follows EU theory by assuming that it is possible to assign to each option a subjective value that represents its desirability to the decision maker. However, prospect theory differs from EU theory in the way the subjective value is determined. First, prospect theory assumes that the outcomes of risky options are evaluated relative to a reference point, such that the outcomes can be framed in terms of losses and gains. When comparing losses against gains it is further assumed that the absolute value of a loss has a larger impact on a decision than the same value of a gain (e.g., "losses loom larger than gains", Kahneman & Tversky, 1979, p. 279), a phenomenon known as *loss aversion*. In other words, losing \$25 is worse than winning \$25 is good. Second, prospect theory assumes that people have a subjective representation of probabilities, such that small probabilities are overestimated and medium and large probabilities are underestimated. Cumulative prospect theory (Tversky & Kahneman, 1992) suggests a refined version of the original prospect theory (Kahneman & Tversky, 1979). Hereafter, we will write CPT when we discuss the model and cumulative prospect theory when we discuss the theory behind the model.

According to CPT, if prospect O has two possible outcomes, as all prospects discussed in this paper, then the subjective value V of O can be determined by

$$V(O) = \sum \pi(p_i) v(x_i), \tag{2}$$

where $\pi(\cdot)$ is a weighting function of the objective probabilities and $v(\cdot)$ is a function defining the subjective value of outcome i. It is assumed that both the probability weighting function and the value function differ for gains and losses.

The subjective value of payoff x is defined as

$$v(x) = \begin{cases} x^\alpha, & \text{if } x \geq 0 \\ -\lambda(-x)^\beta, & \text{if } x < 0, \end{cases} \tag{3}$$

where α and β are free parameters that vary between 0 and 1 and modulate the curvature of the subjective value functions (the weighting functions for gains and losses will be different as long as $\alpha \neq \beta$). The λ parameter specifies loss aversion, with larger values expressing larger loss aversion. Cumulative prospect theory assumes that losses carry more weight than gains so that one could restrict λ to be larger than 1. However, in order to test the loss aversion assumption of prospect theory one can allow values for λ to be smaller than 1 (but larger than 0). For computational simplicity λ should be further constrained to a maximum large value (e.g., $\lambda < 10$). A person with a value of $\lambda < 1$ will show the opposite of loss aversion. That is, the person will give larger weight to gains than to losses of the same absolute value.

The two most notable aspects of the value function defined in Eq. (3) is that, as long as α and β are neither 0 nor 1 and $\lambda > 1$, then (i) the difference between x and x plus \$1 (and between $-x$ and $-x$ minus \$1) will be perceived as greater when payoffs are close to 0 than when payoffs are distant from 0 and (ii) the difference between $-x$ and $-x$ minus \$1 will be perceived as larger than the

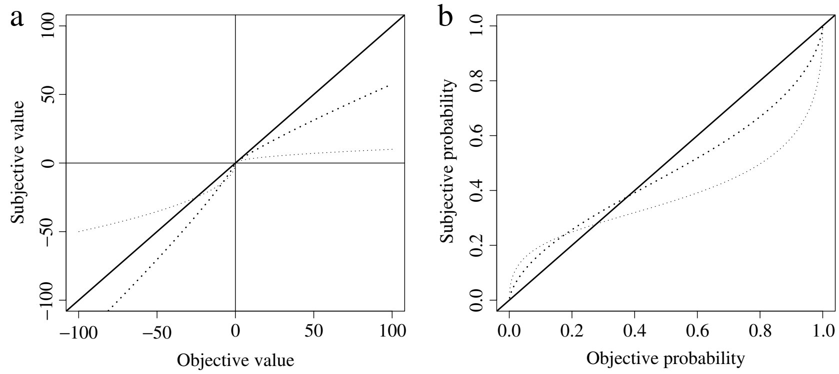

Fig. 1. (A) Two value functions (thick dotted line: $\alpha = .88$, $\beta = .88$, and $\lambda = 2.25$; thin dotted line: $\alpha = .5$, $\beta = .5$, and $\lambda = 5$) and (B) two probability functions (thick dotted line: $c = .69$; thin dotted line: $c = .5$) generated by CPT.

difference between x and x plus \$1. This is illustrated in Fig. 1(A), which displays two separate value functions. The thick dotted line shows a value function with $\alpha = .88$, $\beta = .88$, and $\lambda = 2.25$ and the thin dotted line shows a value function with $\alpha = .5$, $\beta = .5$, and $\lambda = 5$ (the former function uses the median values from the individual fits in Tversky & Kahneman, 1992). As can be seen, a lower value on α and β is associated with a faster decrease in marginal utility and a higher value on λ is associated with stronger loss aversion.

Probabilities are transformed by a weighting function. In the case of only two possible outcomes (as for all options considered in the present article) the weighting function can be reduced to

$$\pi(p_i) = \frac{p_i^c}{(p_i^c - [1 - p_i^c])^{1/c}} \tag{4}$$

with $c = \gamma$ for positive payoffs and $c = \delta$ for negative payoffs. Parameter c, which can take values between 0 and 1, specifies the inverse s-shaped transformation of the weighting function.

Fig. 1(B) shows two probability functions. The thick dotted line represents a case where $c = .69$ and the thin dotted line represents a case where $c = .5$. As long as $c < 1$, low probabilities are overestimated and high probabilities are underestimated and the subjective difference between p and p plus .01 is larger if p is an extreme probability than if p is a moderate probability. The magnitude of these effects increases as the value of c decreases.

Imagine, once again, a choice between gambles A ([\$100, .6; −\$100, .4]) and B ([−\$10, .5; \$0, .5]) presented above. Assume a CPT equipped with the median parameter values from the individual fits in Tversky and Kahneman (1992, i.e., $\alpha = .88$, $\beta = .88$, $\lambda = 2.25$, $\gamma = .61$, and $\delta = .69$). According to Eq. (3), the subjective value of \$100 equals 57.54 and the subjective value of −\$100 equals −129.47. According to Eq. (4), the subjective probability of .6 equals .47 and the subjective probability of .4 equals .39. Finally the subjective value of gamble A, derived from Eq. (2), equals −23.45. In the same way it can be calculated that the subjective value of gamble B equals −7.75. Hence, in contrast to the suggestion by the gambles' expected values and the EU model exemplified above, CPT equipped with the median parameter provided by Tversky and Kahneman (1992) states that gamble B should be chosen over gamble A. Note that it is essentially loss aversion that causes CPT to choose differently than the *EU* model.

The original version of CPT is deterministic. That is, the decision maker should always choose the options with the larger subjective value. However, to account for the probabilistic nature of human choice behavior it is essential to add an error theory to the model. By doing so, the model predicts the choice of an option with a specific probability. There are various ways that such an error theory can be specified. The simplest way is to use a choice rule where the choice probabilities are assumed to be a monotonic function of the differences of the gambles' subjective values. A range of probabilistic choice rules have been used (for a summary see Stott, 2006). We will use an exponential Luce choice rule, which states that the probability of choosing option A over option B equals

$$p(A, B) = \frac{e^{\varphi \cdot V(A)}}{e^{\varphi \cdot V(A)} + e^{\varphi \cdot V(B)}}, \tag{5}$$

where $\varphi > 0$ is a sensitivity parameter that quantifies the extent to which the model's choice is determined by the difference in subjective values for options A and B. For computational stability it is often useful to rewrite Eq. (5) as a logistic choice rule,

$$p(A, B) = \frac{1}{1 + e^{\varphi[V(B) - V(A)]}}. \tag{6}$$

When the sensitivity parameter φ equals 0 choice behavior is random, that is, $p(A, B) = .5$. As φ increases, choice behavior is determined more and more by the difference in subjective value for the two choice options. In the limit, as φ grows large, even a very small difference in subjective values will result in a consistent preference, such that the probabilistic version becomes virtually identical to the deterministic version. Due to its prominence we have selected the exponential choice rule for implementing an error theory. However, it should be noticed that the choice rule cannot explain violations of stochastic transitivity or of the independence from irrelevant alternatives principle (c.f. Rieskamp et al., 2006). Thus, we take the exponential choice rules as a simplifying approximation. In sum, probabilistic CPT comprises the following six free parameters:

1. Parameter α quantifies the curvature of the subjective value function for gains;
2. Parameter β quantifies the curvature of the subjective value function for losses;
3. Parameter λ quantifies loss aversion;
4. Parameter γ quantifies the shape of the probability weighting function for gains;
5. Parameter δ quantifies the shape of the probability weighting function for losses;
6. Parameter φ quantifies the extent to which choice behavior is guided by subjective values.

3. Parameter estimation for cumulative prospect theory: Present approaches

The parameters of CPT quantify the various cognitive processes that are thought to drive decision making under risk. This means that researchers can use CPT to assess the influence of categorical predictors (e.g., different experimental conditions) or continuous predictors (e.g., age or income) not just on the level of observed decisions, but also on the deeper level of unobserved psychological processes. In order to do so, however, one needs to estimate the parameters of CPT. Here we briefly discuss three popular approaches to do so (see also Busemeyer & Diederich, 2009; Cohen, Sanborn, & Shiffrin, 2008).

3.1. Approach 1: Obtain parameter estimates from previous studies

The simplest solution to the parameter estimation problem is not to engage in parameter estimation at all, but instead rely on estimates from previous studies. For instance, Benartzi and Thaler (1995) explained the equity premium puzzle by assuming that people are loss-averse. That is, people prefer bonds over stocks because the potential short-term losses are larger for stocks than for bonds. The authors illustrated the effect of loss aversion on investments using the CPT parameter estimates from Tversky and Kahneman (1992). In a different study, Brandstätter et al. (2006) used the same set of parameters when they compared CPT to a simple model of decision making, the (parameter free) "priority heuristic".

One disadvantage of this approach is that the conclusions are based on a specific set of parameter estimates — estimates obtained in a specific decision task, with specific participants, at a specific time, and in a specific context. The conclusions do not necessarily generalize to different tasks and different participants. For example, though the ability to explain behavior in the Allais paradox is one of the key trademarks of the model, CPT produces a choice behavior that is inconsistent with the behavior in the Allais paradox if it is equipped with the parameter estimates from Tversky and Kahneman (1992), see Nielson and Stowe (2002). A second disadvantage of this approach is that it does not allow researchers to use CPT as a measurement tool to quantify changes in psychological processes as a function of categorical or continuous covariates.

3.2. Approach 2: Complete pooling: Participants are identical

As its name suggests, the complete pooling approach first averages the data across all participants, and then estimates the model's parameters for the averaged data. In other words, the researcher effectively analyzes the data as if they were generated by a single participant. The main advantage of the complete pooling approach is that the aggregation procedure might sometimes reveal underlying patterns that are otherwise obscured by noise in individual data. Therefore, the approach can be useful in situations where the number of data per participant is low, that is when individual data can be assumed to be relatively noisy (Cohen et al., 2008). A second advantage is that it is computationally straightforward to use maximum likelihood to estimate parameters based on the aggregate behavior over participants.

Unfortunately, the approach comes with a serious drawback. The complete pooling approach neglects the fact that participants may differ. When participants differ, conclusions drawn from the complete pooling approach are known to be potentially misleading. In particular, Estes and others have shown that individual differences, when ignored, can induce averaging artifacts in which the data that are averaged over participants are no longer representative for any of them (Estes, 1956, 2002; Heathcote, Brown, & Mewhort, 2000). For example, consider a situation in which one half of participants is risk-seeking whereas the other half is risk-averse. When CPT is fitted to the average data it may support the conclusion that the participants are risk- neutral, a conclusion that is correct for none of the individual participants.

3.3. Approach 3: Complete independence: Participants are unique

The assumption that all participants are identical is clearly unrealistic, and this is why many researchers now estimate model parameters for each participant separately (e.g. Brown & Heathcote, 2003; Estes & Maddox, 2005; Haider & Frensch, 2002). This complete independence approach implicitly assumes that each participant is unique. By considering each participant as a separate unit of analysis, the complete independence approach avoids the averaging artifacts that plague the complete pooling approach, and allows for statistical inferences both for the entire group and for individual participants.

The main drawback of the complete independence approach is that in standard experimental setups individual participants contribute relatively few data. Consequently, individual parameter estimates are relatively noisy and unreliable. Below we will illustrate how single-subject maximum likelihood, one of the most popular estimation methods for CPT (e.g. Harrison & Rutström, 2009; Harless & Camerer, 1994; Stott, 2006), can produce extreme, implausible point estimates for parameters estimated with high uncertainty.

In sum, fundamental problems are associated with all current methods to estimate parameters in CPT. The complete pooling model, although robust, may lead to averaging artifacts — participants are not identical. The complete independence model, although it avoids averaging artifact, may lead to noisy and extreme parameter estimates — the price that has to be paid for assuming that each participant is unique.

4. Hierarchical Bayesian parameter estimation

Hierarchical modeling constitutes an attractive compromise between the extremes of complete pooling and complete independence (Gelman & Hill, 2007; Shiffrin, Lee, Kim, & Wagenmakers, 2008). In a hierarchical model, individual parameter estimates are assumed to come from a group-level distribution, often a normal distribution with estimated mean and standard deviation. When the group-level standard deviation is estimated to be very small, this indicates that the individual participants behave similarly, consistent with the assumption of the complete pooling approach; but when the group-level standard deviation is estimated to be very large, this indicates that the individual participants behave differently, consistent with the assumption of the complete independence model.

Note that as in the complete independence approach, hierarchical models also estimate parameters for each individual participant; these parameters, however, are constrained by the higher-level group distribution. This group-level constraint allows the potentially unreliable estimation of a particular individual's parameter to borrow strength from the information that is available about the other individuals.

Thus, hierarchical models simultaneously account for both differences and similarities between people (e.g. Morey, Pratte, & Rouder, 2008; Morey, Rouder, & Speckman, 2008; Navarro, Griffiths, Steyvers, & Lee, 2006; Wetzels, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2010). Hierarchical models avoid the averaging artifacts that come with the complete pooling approach, and also avoid, at least to the extent possible, the unreliability that comes with the estimation of parameters for individual participants. Naturally, the benefits of hierarchical modeling depend on the extent to which the hierarchical structure is appropriate for the data under scrutiny; for example, it would be ill-advised to assume a group-level normal distribution when individual participants cluster in two or three separate subgroups (e.g., people who learn versus people who do not, e.g., Lee & Webb, 2005; Lee & Wetzels, 2010). Therefore, the kind of hierarchical structure that is imposed on the individual units should be informed by prior knowledge and by exploratory data analysis.

In the following we present a hierarchical Bayesian parameter estimation procedure for CPT. Although hierarchical analyses can be carried out using orthodox methodology (i.e. Farrell & Ludwig, 2008; Hoffman & Rovine, 2007), there are philosophical and practical reasons to prefer the Bayesian methodology (e.g. Gelman & Hill, 2007; Lindley, 2000, respectively).

4.1. Application to cumulative prospect theory

We implemented a Bayesian hierarchical estimation procedure for CPT, as follows. First, recall that CPT has six parameters: $\alpha \in [0, 1]$, $\beta \in [0, 1]$, $\lambda \in (0, \infty)$, $\gamma \in [0, 1]$, $\delta \in [0, 1]$, $\varphi \in (0, \infty)$ (as described below, we constrained the upper boundary of λ and φ for the analyses in this paper). Parameter $\alpha(\beta)$ quantifies the curvature of the subjective value function for gains (losses). As long as $\alpha(\beta)$ is neither 0 nor 1, the difference between x and x plus \$1 ($-x$ and $-x$ minus \$1) will be perceived as greater when payoffs are close to 0 than when payoffs are distant from 0 (see Fig. 1(A)). Parameter λ captures whether gains and losses carry an equal amount of weight ($\lambda = 1$) or if relatively more weight is put on either gains ($\lambda < 1$) or losses ($\lambda > 1$) (the latter producing loss aversion, a key assumption of CPT). Parameter $\gamma(\delta)$ quantifies the curvature of the probability weighting function for gains (losses). As long as $\gamma(\delta) < 1$, low probabilities are overestimated, high probabilities are underestimated, and the subjective difference between p and p plus .01 is larger if p is an extreme probability than if p is a moderate probability (see Fig. 1(B)). Parameter φ captures whether the choice behavior is random (low values on φ) or guided by subjective values (high values on φ). Together, these six parameters determine the probability that a decision maker will prefer a certain gamble A over another gamble B (Eqs. (2)–(6)). In the hierarchical Bayesian framework, we assumed that each decision maker or participant i has their own parameters: α_i, β_i, γ_i, δ_i, λ_i, and φ_i.

Consider first all parameters that are constrained to be between 0 and 1, that is, α_i, β_i, γ_i, δ_i. In order to facilitate hierarchical modeling, we followed Rouder and Lu (2005) and first transformed these parameters to the probit scale, that is, $\alpha_i^{\varphi} = \Theta^{-1}(\alpha_i)$, $\beta_i^{\varphi} = \Theta^{-1}(\beta_i)$, $\gamma_i^{\varphi} = \Theta^{-1}(\gamma_i)$, $\delta_i^{\varphi} = \Theta^{-1}(\delta_i)$, where Θ^{-1} denotes the inverse cumulative distribution function of the standard normal distribution (see also Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). On the probit scale the parameters cover the entire real line, and we assumed that the individual probitized parameters come from independent group-level normal distributions, that is, $\alpha_i^{\varphi} \sim N(\mu^{\alpha}, \sigma^{\alpha})$, $\beta_i^{\varphi} \sim N(\mu^{\beta}, \sigma^{\beta})$, $\gamma_i^{\varphi} \sim N(\mu^{\gamma}, \sigma^{\gamma})$, and $\delta_i^{\varphi} \sim N(\mu^{\delta}, \sigma^{\delta})$. Note that the assumption of prior independence does not mean that the parameters remain uncorrelated a posteriori. Alternatively, one can specify dependent group-level distributions using a prior on the variance matrix (e.g. Klauer, 2010). In Bayesian statistics, both approaches are common practice (e.g. Ntzoufras, 2009; Rouder et al., 2007). Here we prefer to use the independence method because it is simpler to implement and easier to communicate.

Finally, we assigned priors to the group-level parameters. For the group means we used standard normal priors, as these correspond to uniform priors on the rate scale: $\mu^{\alpha} \sim N(0, 1)$, $\mu^{\beta} \sim N(0, 1)$, $\mu^{\gamma} \sim N(0, 1)$, and $\mu^{\delta} \sim N(0, 1)$. For the group standard deviations we used uninformative uniform priors (cf. Gelman & Hill, 2007): $\sigma^{\alpha} \sim U(0, 10)$, $\sigma^{\beta} \sim U(0, 10)$, $\sigma^{\gamma} \sim U(0, 10)$, and $\sigma^{\delta} \sim U(0, 10)$. Note that when the group standard deviation is >3, the individual parameters are almost certainly bimodal on the rate scale, something that we deem a priori implausible.

The two remaining parameters are λ_i and φ_i. In principle, these parameters can take on any positive value. We therefore assumed that these parameters come from a lognormal distribution, $\lambda_i \sim LN(\mu^{\lambda}, \sigma^{\lambda})$ and $\varphi_i \sim LN(\mu^{\varphi}, \sigma^{\varphi})$. Based on our experience with CPT, we decided that for both parameters, the group mean was certain to lie in an interval that ranges from 0.1 to 5. Assuming an uninformative uniform prior distribution for the lognormal group means, this translates to the following priors: $\mu^{\lambda} \sim U(-2.30, 1.61)$ and $\mu^{\varphi} \sim U(-2.30, 1.61)$. Finally, we assigned uninformative uniform priors for the lognormal standard deviations σ^{λ} and σ^{φ}, ranging from 0 to 1.13; this latter number (i.e., $3.91/\sqrt{12}$) is the standard deviation of a uniform distribution that ranges from -2.30 to 1.61. Given our prior knowledge that plausible values are restricted to the interval from 0.1 to 5 (i.e., -2.30 to 1.61 on the log scale), and that the group-level distribution is unlikely to be bimodal, the value 1.13 represents a reasonable upper bound. Hence, the model specification is completed by $\sigma^{\lambda} \sim U(0, 1.13)$ and $\sigma^{\varphi} \sim U(0, 1.13)$.

We implemented the hierarchical Bayesian model in WinBUGS (Lunn et al., 2000, 2009), a free software program that comes with preprogrammed distributions, functions, and MCMC algorithms. Although the MCMC algorithms in WinBUGS may not be the most efficient for any particular nonstandard application, a model implemented in WinBUGS is easy for other researchers to understand, apply, and adjust. The WinBUGS code for our model is available online.[2]

5. Study 1: Parameter recovery

The main goal of Study 1 was to confirm that the Bayesian hierarchical estimation procedure is able to accurately recover parameter values from data simulated under CPT. A secondary goal was to compare the performance of the Bayesian hierarchical procedure with that of standard single-subject MLE (e.g. Rieskamp, 2008).

To assess parameter recovery performance we used CPT to generate three sets of choice data, each set containing 30 synthetic subjects that completed an experiment featuring 60 pairs of mixed gambles. Mixed gambles – gambles in which one possible outcome is a loss and the other a gain – are required to estimate the loss aversion parameter λ. The set of mixed gambles was taken from a real experiment (Rieskamp, 2008, Study 2), in which the gambles were randomly generated under the constraint that one gamble never stochastically dominated the other gamble and that the expected values of the two gambles were relatively similar, so that the choices were not too easy.

To simulate the choice behavior using CPT, the generative parameter values for each synthetic subject were set to the values provided by Tversky and Kahneman (1992): $\alpha = .88$, $\beta = .88$, $\gamma = .61$, $\delta = .69$, and $\lambda = 2.25$. To examine the effect of noise on parameter recovery we used three different values for the sensitivity parameter φ and used each value to generate a separate data set. The three levels of φ were $\varphi = .04$ (high noise), $\varphi = .14$ (medium noise), and $\varphi = .40$ (low noise); these values were chosen so that average simulated choice behavior would correspond to the deterministic prediction of CPT in 60%, 75%, and 90% of all choices, respectively. We refer to these three data sets as D60, D75, and D90.

After the three data sets had been generated, the parameters of CPT were estimated using both MLE and our hierarchical Bayesian

[2] The model can be found under the heading *Hierarchical Bayesian Parameter Estimation for CPT* at http://psycho.unibas.ch/datensaetze/abteilungen/economic-psychology/supementary-materials/abteilung/economic-psychology/.

method. For the latter method, posterior distributions were approximated by a total of 30,000 MCMC samples obtained from three chains, after a burn-in of 1000 samples and after subsampling such that only every tenth sample was recorded (for an introduction to Bayesian modeling using WinBugs see, Ntzoufras, 2009). Convergence of the MCMC chains was confirmed by visual inspection and by computing the $\hat{R}$ statistic (Gelman & Rubin, 1992).

5.1. Results and discussion

Table 1 summarizes the main results. For the maximum likelihood method, Table 1 shows the median of the 30 MLE point estimates across the synthetic participants. The standard deviation (SD) is given in brackets and serves as an indication of the variability of the median. For the hierarchical Bayesian method, Table 1 shows the median of 30 individual posterior modes (for each synthetic subject, we considered the individual-level joint posterior and estimated its joint mode). By reporting the median of the modes of the individual-level posteriors (instead of, say, the mode of the posterior for the group-level mean) we facilitate a comparison between MLE and the hierarchical Bayesian method.

The upper rows of Table 1 – labeled "Maximum Likelihood" or "Hierarchical Bayesian" – show the parameter recovery results for the full version of CPT, that is, the version in which all parameters are free to vary. In general, the data-generating parameters were recovered better as the noise in the choice process decreased. However, even with large noise both estimation methods recovered parameters α,β, γ, and δ with some tolerable error. As expected, the individual point estimates from MLE tend to have a larger variability (i.e., larger SD) than those from the hierarchical Bayesian approach. The Bayesian approach also seems to have an advantage in estimating the sensitivity parameter φ, a parameter that MLE tends to overestimate.

There were, however, two unexpected results. First, λ was significantly underestimated at all levels of noise, a tendency that was more pronounced for MLE than for the Bayesian approach. This is particularly troublesome because λ represents a key psychological concept in prospect theory, namely, loss aversion. The second unexpected result is that α was systematically estimated to be lower than β.

It is likely that these results are caused by a peculiarity of CPT, that is, its ability to account for loss aversion in multiple ways. The most obvious way for CPT to account for loss aversion is by parameter λ (after all, the purpose of λ is to measure loss aversion). A second way, however, is to decrease the marginal utility at a faster pace for gains than for losses. This occurs when α is smaller than β. Based on this reasoning, we hypothesized that the parameter estimation routines compensate for the underestimation of λ by assigning lower values to α than to β; in this way, CPT accounts for the existing loss aversion indirectly in a manner that we had not anticipated.

To test our hypothesis we carried out a new set of analyses. In this new set, we constrained CPT such that α and β were forced to take on the same value. These results are presented in the lower row of each cell in Table 1, labeled "$\alpha = \beta$". As can be seen from the table, adding this restriction to the model greatly improved the recovery of λ for both estimation methods (without affecting the recovery of the other parameters). Most notably, in the case of low noise (i.e., D90), the hierarchical Bayesian method recovered parameter λ perfectly.

The improvement in recovery performance gained by constraining CPT is illustrated further by Fig. 2. Fig. 2 presents the posterior distributions for the means of the group-level distributions from the hierarchical Bayesian method (when CPT is fitted to D90). Solid lines represent the posteriors from the full CPT, dotted lines represent the posteriors from the restricted CPT, and the vertical

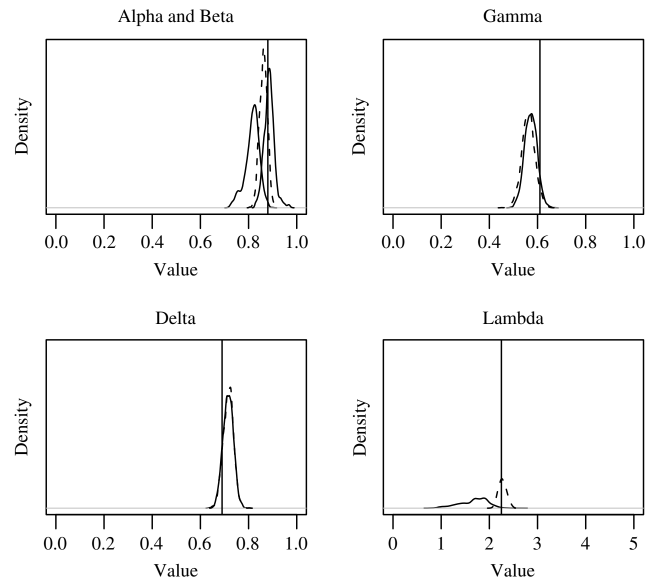

Fig. 2. Posterior distributions for group-level means for the full version of CPT (solid lines) and for the restricted version of CPT (dotted lines), applied to simulated data (i.e., Study 1). Vertical lines represent the true values used to generate the data.

lines indicate the true data-generating parameter values. As can be seen from the figure, the distributions for parameters α,β, γ, and δ are located near the true values for both the full CPT and the restricted CPT. The reason the peaks of the posterior distributions do not fall exactly on the target value is that there are some systematic patterns of the data generated by the error that the model cannot separate. However, these deviations would have most likely disappeared if we had used a larger data set.

The important aspect of Fig. 2 is the difference between full and restricted CPT that is evident in the panel for the loss aversion parameter λ. For the full CPT, the posterior for λ is relatively wide and located almost entirely below the true data-generating value; in contrast, for the restricted CPT the posterior for λ is narrowly peaked and centered on the true data-generating value. This supports our hypothesis that there is information in the data that can be used to estimate λ, but that much of this information is absorbed by α and β if these parameters are allowed to vary freely.[3]

From this simulation study, we conclude that the hierarchical Bayesian method recovers the data-generating parameters values somewhat more accurately than MLE, and with less variability. This is particular evident for the sensitivity parameter φ. Recovery performance for both the hierarchical Bayesian method and MLE increases as the level of noise decreases. Finally, the initial analysis revealed that the unrestricted, full CPT dramatically underestimates loss aversion, for both the Bayesian method and MLE. Instead, the full CPT can account for loss aversion by assigning lower values to α than to β. This hypothesis was confirmed by a second analysis in which a restricted (i.e., $\alpha = \beta$) version of CPT was able to accurately recover the loss aversion parameter λ. This finding suggests that when researchers use CPT to analyze real data, they should be careful to quantify loss aversion by λ from the full model. Instead, it would be prudent to also fit the $\alpha = \beta$-restricted CPT, and, in case the restricted model still provides an acceptable fit to the data, to quantify loss aversion by λ from the restricted model.

[3] CPT could also produce loss aversion by creating an asymmetry between γ and δ. We fitted a CPT where α and β were allowed to vary freely but where $\gamma = \delta$. This constrained version of CPT was just as poor as the full version at recovering λ.

Table 1
Median point estimate (maximum likelihood estimation) and median of the posterior mode (hierarchical Bayesian method) for each parameter for each generated data set. Standard deviation enclosed in brackets.

Method	Parameters					
	α (.88)	β (.88)	γ (.61)	δ (.69)	λ (2.25)	φ
	D60 — high noise, $\varphi = .04$					
Maximum likelihood	.88 (.36)	1.0 (.16)	.49 (.36)	.60 (.28)	.31 (2.05)	.23 (3.43)
$\alpha = \beta$	.93 (.23)		.43 (.37)	.60 (.24)	.84 (1.33)	.15 (3.65)
Hierarchical Bayesian	.63 (.02)	.99 (.00)	.50 (.09)	.61 (.07)	.55 (.10)	.12 (.00)
$\alpha = \beta$	.74 (.01)		.50 (.36)	.58 (.01)	1.50 (.12)	.10 (.00)
	D75 — medium noise, $\varphi = .14$					
Maximum likelihood	.90 (.23)	.97 (.13)	.58 (.26)	.67 (.14)	1.48 (2.83)	.20 (1.90)
$\alpha = \beta$	.86 (.11)		.52 (.29)	.68 (.14)	1.73 (1.25)	.22 (2.44)
Hierarchical Bayesian	.84 (.01)	.99 (.00)	.56 (.00)	.71 (.01)	.57 (.00)	.18 (.01)
$\alpha = \beta$	.91 (.01)		.59 (.01)	.72 (.02)	2.42 (.05)	.11 (.00)
	D90 — low noise, $\varphi = .40$					
Maximum likelihood	.79 (.17)	.87 (.09)	.56 (.19)	.66 (.09)	1.07 (2.41)	1.38 (2.56)
$\alpha = \beta$	.85 (.07)		.57 (.21)	.69 (.09)	1.72 (.81)	.91 (2.43)
Hierarchical Bayesian	.82 (.01)	.88 (.00)	.58 (.02)	.73 (.01)	1.83 (.11)	.60 (.06)
$\alpha = \beta$	.86 (.00)		.55 (.04)	.73 (.00)	2.25 (.06)	.54 (.05)

6. Study 2: Application to behavioral data

The goal of Study 2 was to explore whether the conclusions from simulated data (i.e., Study 1) generalize to real empirical data. Here we analyze the data from Study 2 in Rieskamp (2008). As explained above, the Rieskamp gambles were generated by semi-randomly sampling values and probabilities, thereby addressing a variety of decision-making situations.

The Rieskamp study featured 30 participants, each of which was confronted with a series of 180 pairs of gambles; 60 pairs had only positive outcomes (i.e., gains), 60 pairs had only negative outcomes (i.e., losses), and 60 pairs had mixed outcomes (i.e., one negative and one positive outcome).

The full and the $\alpha = \beta$ restricted CPT were fit to the Rieskamp data with both the maximum likelihood method and with the hierarchical Bayesian method.[4] For the hierarchical Bayesian method we estimated posterior distributions based on a total of 50,000 MCMC samples from three chains, after a burn-in of 1000 samples and after subsampling such that only every 10th sample was recorded. Convergence of the MCMC chains was confirmed by visual inspection and by computing the $\hat{R}$ statistic (Gelman & Rubin, 1992).

6.1. Results and discussion

The main results are summarized in Table 2. Table 2 show the median of the 30 point estimates for the maximum likelihood method and the median of the 30 individual posterior modes for the hierarchical Bayesian method (standard deviations are given within brackets). Rows 1 and 3 presents estimates for the full version of CPT and rows 2 and 4 presents estimates for the constrained version of CPT (constrained so that $\alpha = \beta$).

There are three notable aspects of Table 2. First, while there is a high correspondence between the median point estimate and the median posterior mode for all parameters, standard deviations are systematically larger for point estimates than for posterior modes. That is, as expected, the main difference between the two methods is that the MLE method indicates larger individual differences than the the hierarchical Bayesian method.

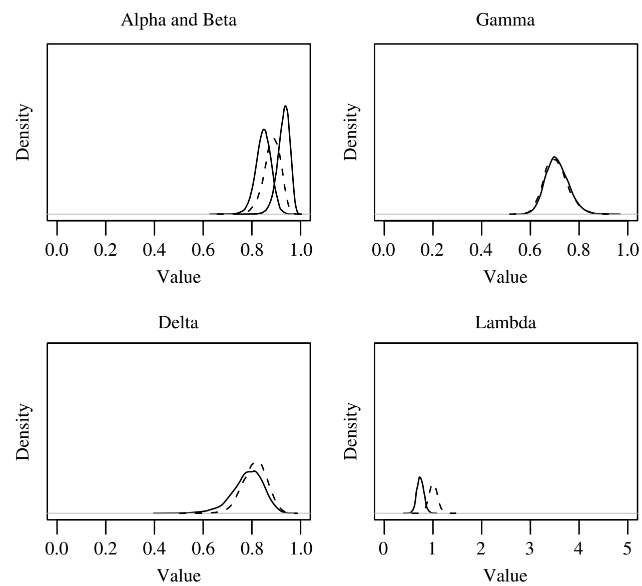

Fig. 3. Posterior distributions for group-level means for the full version of CPT (solid lines) and the constrained version of CPT (dotted lines).

Second, when the full version of CPT is fitted to data, both MLE and the hierarchical Bayesian method suggest that the median participant in the Rieskamp study is loss-seeking (i.e., $\lambda < 1.0$). When the constrained version of CPT is fitted to data, however, both methods suggest that the median participant in the Rieskamp study is neither loss-seeking nor loss-averse. This is also illustrated by that the posterior distribution for the group-level mean of λ is almost perfectly symmetric around 1.0 when the constrained version of CPT is fitted (see Fig. 3). The results presented in Table 2 and Fig. 3 confirm the pattern that was evident in the parameter recovery study: the full version of CPT tend to underestimate λ if α and β are allowed to take different values. This possibility compromises the psychological interpretation of the model parameters.

Third, while the MLE method indicates very large individual differences for λ, the hierarchical Bayesian method indicates small individual differences for λ. Most notably, while the mode in the posterior distribution for λ is smaller than 1.0 for all participants, the point estimates generated by the MLE method is greater than 1.0 for eleven participants and greater than 2.0 for six of these eleven participants. This discrepancy reflects the fact that for some participants, the choice data hold no or very little information that

4 To avoid unreliable estimates of CPT's parameters in Rieskamp (2008) the following restrictions were set on the parameters: $0 < \alpha, \beta \leq 1$, $0.40 \leq \gamma, \delta \leq 1$, and $0 < \varphi \leq 10$. However, in order to test the central assumption of prospect theory, which holds that people are in general loss-averse, we allowed λ to vary between 0 and 10.

Table 2
Median point estimate (maximum likelihood estimation) and median of the posterior mode (hierarchical Bayesian method) for each parameter for the data of Rieskamp (2008, Study 2). Standard deviation is enclosed in brackets.

Method	Parameters					
	α	β	γ	δ	λ	φ
Maximum likelihood	.91 (.29)	1.0 (.28)	.78 (.25)	.77 (.24)	.81 (2.67)	.25 (0.73)
$\alpha = \beta$	.90 (.28)		.72 (.26)	.75 (.26)	.97 (1.93)	.21 (1.03)
Hierarchical Bayesian	.87 (.10)	.98 (.06)	.71 (.10)	.82 (.26)	.75 (.04)	.19 (.08)
$\alpha = \beta$	.91 (.16)		.68 (.11)	.89 (.19)	1.02 (.26)	.18 (.09)

Note.

can be used to estimate λ. Thus, single-subject MLE is confronted with the challenge of trying to find a maximum in a very flat parameter landscape. The result of this search is likely to yield almost any random value in the allowed range. In contrast, the hierarchical Bayesian method uses the information from other participants to pull extreme estimates in toward the group mean. This pull is stronger to the extent that the single-subject parameter is unreliable. The effect is clear: according to MLE, six participants had point estimates for λ that were higher than 2. According to the hierarchical Bayesian method, the median of the posterior modes for these six participants is .74, exactly equal to the posterior mode of the group-level mean for λ. This suggests that the participants with high MLE values for λ were not extremely loss-averse; instead, their data were relatively uninformative with respect to λ, and in the hierarchical Bayesian method this caused the estimates to be determined to a relatively large extent by the information available from the other participants.

Finally, we examined the descriptive accuracy of the hierarchical CPT (full version) by considering a posterior predictive measure. In order to obtain this measure, we randomly sampled parameter values from the individual posterior distributions, used these values to generate predicted choices for each of the 180 items for each of the 30 participants, and finally compare the predicted choices with the actual choices. In more detail, the analysis was conducted as follows (remember, 5000 samples from each of 3 chains were recorded for each of the six parameters for each of the 30 participants). (a) By randomly sampling value v_1 from a uniform distribution ranging from 1 to 3 and value v_2 from a uniform distribution ranging from 1 to 5000, one sampling round was determined for participant *i* (i.e., iteration v_2 from chain v_1 for participant *i*). (b) The parameter values that in the main simulation had been sampled for participant *i* on the targeted round were entered into CPT. (c) The 180 gamble pairs from the Rieskamp study were presented to the model and and a choice between gamble A and gamble *B* was made for each gamble pair. (d) Steps (a)–(c) were repeated once for each of the 30 participants. (e) The proportion of A choices, across all participants, was calculated for each of the 180 gamble pairs. (f) Steps (a)–(e) were repeated 1000 times thereby creating a distribution of predicted proportions of A choices for each gamble pair.

Results of the posterior predictive check showed a clear correspondence between the medians from the 180 distributions and the proportion of A choices from the behavioral data (r = .88 and mean absolute deviation = .11). The range between the 2.5th percentile and the 97.5th percentile in the distribution of predicted A choices covered the actual proportion of A choices for 131 out of 180 gamble pairs (73%). If the median in the distribution of A choices is taken as an indication of which gamble CPT predicts as most likely to be chosen, the prediction by CPT corresponded with the modal response in behavioral data for 151 out of 180 gamble pairs (84%; 87% if pairs where 50% of the participants choose A and 50% choose B are excluded). This analysis confirms that hierarchical CPT provides at least a decent fit to behavioral data. However, note that Rieskamp (2008) compared CPT to alternative choice models, including decision field theory (Busemeyer & Townsend, 1993); when CPT was rigorously tested against decision field theory by constructing choice situations in which the two models made different predictions, decision field theory performed much better than CPT. The main point of the reanalysis of the Rieskamp data for the current manuscript was to illustrate the difficulty of estimating the parameters of a complex model such as CPT and to highlight the advantage of using a hierarchical Bayesian method for parameter estimation.

7. General discussion

This article presented a hierarchical Bayesian method to estimate the parameters of cumulative prospect theory, the most influential psychological theory for decision making under risk. The hierarchical Bayesian method, used in several papers in this special issue (see for example Merkle, Smithson, & Verkuilen, 2011 and Ravenzwaaij, Dutilh, & Wagenmakers, 2011, who also considers individual differences in decision-making models), strikes a compromise between the extremes of complete pooling (i.e., participants are identical) and complete independence (i.e., participants are unique) and thereby avoids the risks of both approaches.

The hierarchical Bayesian method has the advantage that it provides robust estimates of a model's free parameters without ignoring or over-weighting individual differences. It does this by pulling individual estimates toward the group mean, an effect that becomes stronger when the individual estimate is less reliable. This so-called shrinkage effect prevents unreliable information from having an extreme and disproportionate influence. The advantages of the hierarchical Bayesian method were borne out in a parameter recovery study and in an application to real data. Compared to single-subject MLE, we found that the hierarchical Bayesian method leads to estimates that are less variable and more robust (this finding mirrors the finding by Ahn, Krawitz, Kim, Busemeyer, & Brown, in press, who in a similar simulation study explored the prospect valence learning model).

Apart from demonstrating the benefits of the hierarchical Bayesian method, this article also provided theoretical insight about CPT. The parameter recovery study showed that one core component of prospect theory, namely loss aversion, is very hard to estimate: the loss aversion parameter λ tends to be dramatically underestimated if the value function parameters α and β are both allowed to vary freely. Apparently CPT is able to predict very similar behavior with different sets of parameters: loss-averse behavior can also be predicted by having lower values for α than for β. In order to interpret the parameters of CPT in terms of psychological processes, we recommend that researchers constrain CPT such that α and β have to take on the same value.

It should be noted, however, that even the restricted CPT did not provide any evidence in favor of loss aversion (i.e., evidence that $\lambda > 1.0$) for the Rieskamp (2008) data. This result parallels findings in several recent studies (e.g. Ert & Erev, 2008; Kermer, Driver-Linn, Wilson, & Gilbert, 2006; Koritzky & Yechiam, 2010). For example, in a study by Erev, Ert, and Yechiam (2008) the key problem included a choice between $0 for sure or a gamble that provided a fifty-fifty chance to either win or lose $1000. If people are loss averse they will choose the safe option because the gamble will be perceived as having a negative expected value (the $1000 loss will loom larger than the $1000 gain). Erev et al. (2008) found that 23 of 45 participants preferred the sure option while the remaining 22 participants preferred the gamble. That is, they found no systematic loss aversion. Despite the fact that Rieskamp (2008) used a paradigm where probabilities and values were explicitly given while Erev et al. (2008) used a paradigm where probabilities and values had to be learned by experience, data from studies such as these indicate that the claim that people are in general loss averse might not be as universal as proposed by Tversky and Kahneman (1992). Notably, it appears as if data patterns that have been taken as indicators of loss aversion are heavily dependent on task-specific factors such as presentation format or response mode (Ert & Erev, 2008), factors that a model such as CPT does not naturally account for.

The re-analysis of the data by Rieskamp (2008) showed that extreme high maximum likelihood estimates for λ might sometimes be caused by uninformative data rather than by participants being loss-averse. Hence, in several respects, our results demonstrate that obtaining clean and reliable estimates for loss aversion requires considerable care. Moreover, in line with Kahneman and Tversky (1979) we have restricted the parameters of CPT to a reasonable range, so that we allowed only for risk aversion in the gain domain and risk seeking in the loss domain. If we had not used these restrictions it might have made it even more complicated to estimate the parameters reliably. In general, one of the advantages of the Bayesian paradigm is to use prior distributions to meaningfully restrict the parameter space of the model under consideration. Finally, the goal of this article was to illustrate the advantages of a hierarchical Bayesian method for parameter estimation in comparison to standard maximum likelihood methods. To illustrate the advantages of the Bayesian method we used prospect theory as a prominent theory for decision making under risk. However, the goal of the present article was not to examine the empirical validity of prospect theory. For this purpose rigorous model comparison tests are necessary, which might illustrate the advantages of alternative models as compared to prospect theory in predicting decisions under risk (see Birnbaum, 2008; Harrison & Rutström, 2009; Rieskamp, 2008).

In general, when dealing with a complex model, such as prospect theory, which has several free parameters it is important to recognize that the model often might have different ways to produce almost the same predictions. When the functional relationships between the parameters of a model are unclear, the interpretation of a single parameter of a model can often be misleading. The present article shows that the hierarchical Bayesian method can lead to more robust estimates of a model's free parameter and a better understanding of the associated theory.

Acknowledgments

The research was sponsored by a research grant from the Swiss National Science Foundation to the first and second author (SNF-100014_126721), and a Vidi grant from the Dutch Organization for Scientific Research (NWO) to the third author. We thank Mary Louise Grossman for editing a draft of this paper.

References

Ahn, W. -Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. B. (in press). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics*.

Bell, D. E. (1982). Regret in decision-making under uncertainty. *Operations Research*, *30*, 961–981.

Bell, D. E. (1985). Disappointment in decision-making under uncertainty. *Operations Research*, *33*, 1–27.

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equilibrium puzzle. *Quarterly Journal of Economics*, *110*, 73–92.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*, 161–194.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. *Psychological Review*, *113*, 409–432.

Brown, S. D., & Heathcote, A. J. (2003). Averaging learning curves across and within participants. *Behaviour Research Methods, Instruments and Computers*, *35*, 11–21.

Busemeyer, J. R., & Diederich, A. (2009). *Cognitive modeling*. Sage.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field-theory: a dynamic cognitive approach to decision-making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk & Uncertainty*, *2*, 61–104.

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*, 692–712.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21*, 575–597.

Ert, E., & Erev, I. (2008). The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic. *Journal of Economic Psychology*, *29*, 715–723.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, *9*, 3–25.

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*, 403–408.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*, 1209–1217.

Fishburn, P. C. (1983). Transitive measurable utility. *Journal of Economic Theory*, *31*, 293–317.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, *7*, 457–472.

González-Vallejo, C. (2002). Making trade-offs: a probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*, 137–155.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 392–406.

Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics*, *12*, 133–158.

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*, 1251–1289.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science*, *17*, 649–653.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika*, *75*, 70–98.

Koritzky, G., & Yechiam, E. (2010). On the robustness of description and experience based decision tasks to social desirability. *Journal of Behavioral Decision Making*, *23*, 89–99.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, *12*, 605–621.

Lee, M. D., & Wetzels, R. (2010). Individual differences in attention during category learning. In R. Catrambone, & S. Ohlsson (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Lindley, D. V. (2000). The philosophy of statistics. *The Statistician*, *49*, 293–337.

Loomes, G., & Sugden, R. (1982). Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–824.

Lunn, D. J., Thomas, A, Best, N., & Spiegelhalter, D. (2000). WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Lunn, D. J., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choices under uncertainty. *Journal of Economic Literature*, *27*, 1622–1668.

Mellers, B. A. (2000). Choice and relative pleasure of consequences. *Psychological Bulletin*, *126*, 910–924.

Merkle, E. C., Smithson, M., & Verkuilen, J. (2011). Hierarchical models of simple mechanisms underlying confidence in decision making. *Journal of Mathematical Psychology*, *55*(1), 57–67.

Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology*, *52*, 376–388.

Morey, R. D., Rouder, J. N., & Speckman, P. L. (2008). A statistical model for discriminating between subliminal and near-liminal performance. *Journal of Mathematical Psychology*, *52*, 21–36.

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review*, *177*, 243–255.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*, 101–122.

Nielson, W., & Stowe, J. (2002). A further examination of cumulative prospect theory parameterization. *Journal of Risk and Uncertainty*, *24*, 31–46.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: Wiley.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*, 323–343.

Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2011). Cognitive model decomposition of the BART: Assessment and application. *Journal of Mathematical Psychology*, *55*(1), 94–105.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1446–1465.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. (2006). Expanding the bounds of rationality: a review of research on preferential choice. *Journal of Economic Literature*, *44*, 631–661.

Roe, R. M., Busemeyer, J. T., & Townsend, J. T. (2001). Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573–604.

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, *72*, 621–642.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schoemaker, P. J. H. (1982). The expected utility model: its variants, purposes, evidence and limitations. *Journal of Economic Literature*, *20*, 529–563.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Starmer, C. (2000). Development in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.

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

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

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010). Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, *54*, 14–27.