Cumulative Prospect Theory's Functional Menagerie

Dr. Henry P. Stott

Department Of Psychology

University Of Warwick

Coventry, CV4 7AL, England

h.stott@dectech.org

(44) 20 7289 9174

Abstract

Many different functional forms have been suggested for both the value function and probability weighting function of Cumulative Prospect Theory (Tversky & Kahneman, 1992). There are also many stochastic choice functions available. Since these three components only make predictions when considered in combination, this paper examines the complete pattern of 256 model variants that can be constructed from twenty functions. All these variants are fit to experimental data and their explanatory power assessed. Significant interaction effects are observed. The best model has a power value function, a risky weighting function due to Prelec (1998), and a Logit function.

Key Words: cumulative prospect theory, stochastic choice

JEL Subject Categories: C52, D81

Cumulative Prospect Theory (CPT) (Luce & Fishburn, 1991; Tversky & Kahneman, 1992) is a prominent deterministic theory in risky decision-making that incorporates two key transformations--one for outcome values and another for objective probabilities. Furthermore, because people's risky choices are stochastic, the core CPT model is often supplemented with a choice function. Whilst there is general consensus on the qualitative shapes of these three transformations, a variety of specific functional forms have been suggested. The objective of this paper is to evaluate which of these specific forms gives the best explanatory account of experimental data.

When fitting one transformation, it is clearly necessary to assume the other two. As such, it is only possible to test functions in combinations. Consequently, a methodical test of each individual function involves testing it in conjunction with all the other possible functional configurations. Such a combinatorial analysis of twenty transformations, including a non-parametric approach, forms the basis of this paper. Several authors (Birnbaum & Chavez, 1997; Camerer & Ho, 1994; Gonzalez & Wu, 1999) have attempted to identify better functional forms for CPT. However, no one seems to have undertaken an extensive and systematic test such as the one described here. Nevertheless, this combinatorial approach is similar to several examinations completed on the functional forms of other decision-making theories (Blondel, 2002; Buschena & Zilberman, 2000; Chechile & Cooke, 1997).

This empirical examination of CPT will be useful for determining which versions of the model practitioners should adopt when applying CPT to real world decision-making problems. Such applications are encountered in many practical situations including financial, medical, and legal contexts (Barberis & Huang, 2005; Doctor et al., 2004; Gutherie, 2003). As discussed, hitherto there has been relatively little empirical guidance on which functional form of CPT to use in such situations.

The work will also provide evidence concerning the axioms used to derive some of the transformations (Bell & Fishburn, 2000, 2001; Luce, 2001; Prelec, 1998; Wakker & Tversky, 1993). Such axiomatizations do not currently play a large role in choosing between functions, but if large performance differences were observed this could contribute to future theorizing.

The transformations tested included eight value functions, eight risky weighting functions, and four choice functions. The combinatorial pattern of these yielded 256 model variants, which can then be fit to individual participant data. In the current experiment this is choice data comprising 90 pairs of two-outcome prospects. Each variant can be fit using maximum likelihood estimation and then an appropriate statistical adjustment made for the varying complexity (i.e. degrees of freedom) of each model. On this basis, the explanatory power of each model can be assessed and the most effective functional forms identified. The results of this analysis lead to several conclusions.

Firstly, there are some notable performance differences between the functions. For example, a choice rule due to Luce (1959) that uses the ratio, rather than the absolute difference, of prospects performs well. It appears that only difference driven choice functions have been used in conjunction with CPT previously, suggesting that there are other choice functions worth exploring in future. Similarly the parabolic value function, which is associated with variance as a risk measure, performs badly. This suggests that much of modern finance is associated with an empirically weak assumption.

Secondly, there are significant interaction effects between the transformations. In other words, the function that performs best as one of CPT's transformations depends on what functions are being used for the other transformations. This is an interesting comment on the descriptive clarity of CPT itself. It also suggests the need to use less complicated transformations that have less latitude to interact, to incorporate stochastic choice in future theorizing, and to design any future tests of CPT against other risky decision-making theories in order to address the effects of such functional performance differences.

Finally, given these interactions, the most predictive version of CPT has a power value curve, a single parameter risky weighting function due to Prelec (1998), and a Logit stochastic process. As such, it is interesting to note that the use of a non-parametric approach does not appear to offer an explanatory advantage over this parametric form. Likewise, the commonly adopted combination of a power value curve, a two parameter risky weighting function due to Goldstein and Einhorn (1987), and a Logit stochastic process is also rejected.

The paper is laid out as follows. In the next section, a stochastic version of CPT is described. Then the second section reviews the different functional forms that have been suggested for CPT together with their normative considerations. Thirdly, prior empirical evidence concerning the different functions is examined. The fourth section describes the experimental procedure. Fifthly, the main findings are reported. Finally, in the last section, some of the main implications of these findings are discussed, along with potential areas for future research.

1. Theory

Cumulative Prospect Theory (CPT) (Tversky & Kahneman, 1992) was developed as an extension of the earlier Prospect Theory, which in turn was formulated as evidence accumulated against older theories such as Expected Utility (EU) (e.g. Allais, 1953; Starmer & Sugden, 1989a). CPT belongs to a broader class of Rank Dependent Utility models (RDU) and is accordingly part of a larger theoretical literature (e.g. Luce & Fishburn, 1991, 1995; Quiggin, 1982; Segal, 1989; Wakker & Tversky, 1993; Yaari, 1987).

There is evidence to support CPT (Camerer & Ho, 1994; Fennema & Wakker, 1997; Hey & Orme, 1994; Loomes, Moffatt, & Sugden, 2002) although clearly no one would claim that it is a definitive theory (Birnbaum, Patton, & Lott, 1999; Starmer, 2000). Nevertheless, it does predict many of the key features of risky decision-making behavior. The original theory is now described followed by a discussion on extending it to accommodate stochastic choice.

1.1 Cumulative Prospect Theory

CPT assigns a value to an *n*-outcome risky prospect $g = \left(p_1, x_1; \ldots; p_n, x_n\right)$ (where p_i is the chance of receiving outcome x_i, with the x_i's in descending order of attractiveness) given by the formula:

$$V(g) = \sum_{\mathrm{i=1}}^{\mathrm{n}} w_i \, v(x_i)$$

Here, $v(x_i)$ is a monotonic value or utility function and w_i is the subjective weighting derived from the outcome probabilities given by:

$$\begin{aligned} w_1 &= \pi\left(p_1\right) \\ w_i &= \pi\left(p_1 + \ldots + p_i\right) - \pi\left(p_1 + \ldots + p_{i-1}\right) \text{ for } 1<\mathrm{i}<\mathrm{n} \end{aligned}$$

Here $\pi(p)$ is a monotonic risky weighting function constrained by $0 \le \pi(p) \le 1$ that previous experimenters have typically characterized using an inverse S-shape (Abdellaoui, 2000; Gonzalez & Wu, 1999; Tversky & Kahneman, 1992). In other words people tend to overweight small probabilities and underweight large ones when they are associated with large outcomes.

Many different functional forms have been suggested for both the risky weighting and value functions. These are discussed in the next section.

1.2 Stochastic Choice

The original form of CPT is deterministic. However, people's choices are stochastic. This is illustrated by Figure 1, which is reproduced from Mosteller and Nogee (1951). In their experiment various risky choices were presented to participants on multiple occasions. Essentially, they gave participants a pairwise choice between $(x, p; -5, (1-p))$ and nothing for certain. Figure 1 shows how often a specific participant, B-I, selected the risky option for a fixed *p* of .33 and various *x*. As can be seen, when *x* was to +5 or +7 cents, the participant always chose nothing for certain, and conversely when *x* was +16 cents the participant always accepted the gamble.

[Figure 1 About Here]

Their result illustrates two points. Firstly, people's risky decision-making is stochastic--when asked the same question multiple times, people often change their minds. Secondly, this stochastic behavior is lawful. In this case, Mosteller and Nogee observed a smooth transition from risk aversive to risk prone behavior as they varied *x*.

[Table 1 About Here]

The tendency of participants to change preferences on repeat questions is widely documented. Table 1 summarizes eight papers where participants have been presented with repeated pairwise choices of risky prospects. These papers find that reversal rates tend to lie in the range .10 to .30. As noted in the table, these reversal rates also vary between participants and between questions. In other words, there are consistent people and clear-cut questions.

1.3 Stochastic Specification

To analyze pairwise choice data it is therefore necessary to supplement the deterministic version of CPT with a choice function. Previous authors have often done this by positing a transformation $P(\ )$ that yields $f(\ )$, the likelihood of picking prospect g_1 given an alternative choice g_2. Formally:

$$f(g_1 | g_2, \theta) = P(V(g_1), V(g_2))$$

where θ is a parameter vector.

Three restrictions are placed upon $P(\ )$. Firstly, for all $V(g_1), V(g_2) \in \mathbb{R}$ then $0 \le P(V(g^1), V(g^2)) \le 1$. Secondly, $P(V(g_1), V(g_2))$ must be weakly increasing in $V(g_1)$. Thirdly, where pairwise choices have been randomized, it should be symmetric[1] with $P(V(g_1), V(g_2)) = 1 - P(V(g_2), V(g_1))$. Different functional forms of $P(\ )$ are discussed in the next section.

---

[1] A trivial consequence of this symmetry condition is that $P(V(g_1), V(g_1)) = \frac{1}{2}$. Combining this with the second restriction means that if $V(g_1) \ge V(g_2)$ then $P(V(g_1), V(g_2)) \ge \frac{1}{2}$ which can be understood as $P(\ )$ preserving the relation "$V(g_1)$ is preferred to $V(g_2)$".

2. Functional Forms

This section reviews the various functional forms that have been suggested for CPT together with various choice functions. Different forms for the value function appear in Table 2 and the risky weighting function in Table 3. Inevitably there is not room to include every value function[2] and every risky weighting function[3], so these tables select the more notable forms. The choice functions are in Table 4. The next three sub-sections discuss each table in turn.

2.1 Value Functions

Of the eight value functions examined in this paper, seven appear in Table 2, starting with the linear approach that represents risk neutrality. The eighth value function is a non-parametric transformation which is discussed at the end of this sub-section.

The second and third equations in Table 2 are the logarithmic and power laws. The logarithmic function is generally accepted as the first utility function, having been proposed by Bernouilli in the eighteenth century. Bernoulli based this on the argument that incremental utility

---

[2] Excluded value functions include $v(x)=\dfrac{\left(e^{kx^{\beta}}-1\right)}{C(+)}$ (Luce & Fishburn,1991), the Log-Quadratic $v(x)=\ln(x+1)+a\left(\ln(x+1)\right)^2$ (Camerer & Ho, 1994; Wu & Gonzalez, 1996), the Sumex $v(x)=ae^{bx}+ce^{dx}$ (Bell & Fishburn, 2000), and the Linear times Exponential $v(x)=(ax+b)e^{cx}$ (Bell & Fishburn, 2000).

[3] Excluded risky weighting functions include the Exponential-power $\pi(p)=\exp\left(-\dfrac{r}{s}\left(1-p^{s}\right)\right)$ (Luce, 2001; Prelec, 1998), the Hyperbolic-logarithm $\pi(p)=\left(1-r\ln p\right)^{-s/r}$ (Luce, 2001; Prelec, 1998), and a linear form with discontinuous end points (Loomes, Moffatt, & Sugden, 2002; Tversky & Fox, 1995).

ought to be proportional to incremental wealth when measured as a proportion of existing wealth (i.e. $\delta v \propto \delta x/x$) (von Neumann & Morgenstern, 1944, p. 629). By similar reasoning, the logarithmic function was also identified by Fechner in the nineteenth century as an implication of Weber's law.

Subsequently, Stevens (1957) modified Fechner's psychophysical model. Stevens cited experimental evidence across 14 different perceptual continua that demonstrated psychological magnitudes were better described by power functions. Whilst Stevens did not discuss wealth directly, many subsequent authors adopted the power law for describing utility. The power function also forms the basis of the Cobb-Douglas utility function, first introduced into economics in the early twentieth century. Several authors have described the normative attractions of a power value function for CPT (Luce, 1991; Tversky & Kahneman, 1992; Wakker & Tversky, 1993).

[Table 2 About Here]

The fourth function listed is the quadratic. This function has played a prominent role in finance. For example, Bell (1995a, 1995b) explored the compatibility of different utility functions with a financial risk-return interpretation and noted that one property of the quadratic transformation is that its evaluation of a prospect can be restated in terms of the prospect's statistical mean and variance. This therefore associates quadratic utility with the traditional measurement of risk in finance.

Another function in the class of curves that can be translated into a risk-return formulation is one that Bell refers to as "linear plus exponential". This is labeled Bell in Table 2. Bell and

Fishburn (2000, 2001) found further normative support for this function within EU and RDU frameworks. Note how the exponential function is a nested case of the Bell function.

The exponential form itself has been frequently cited in risk literature (e.g. Abdellaoui, Vossmann, & Weber, 2003; Camerer & Ho, 1994; Fishburn & Kochenberger, 1979). Earlier authors had noted some unattractive characteristics within an EU framework. Nevertheless, Luce and Fishburn (1995) demonstrated that within CPT, the exponential value function is a consequence of some reasonable conditions. Similarly, Wakker and Tversky (1993) note that CPT's value function takes an exponential form if preferences are invariant under the addition of a positive constant to outcomes.

Five of the value functions used in this analysis are part of the wider family of HARA utilities (Hyperbolic Absolute Risk Aversion). The HARA family is derived by constraining the value function's curvature measure $-v''(x)/v'(x)$ to be hyperbolic in x (see Ingersoll, 1987, p. 38 for details). The special cases presented here are linear ($a \to 1$ and b = 0), quadratic (a=2), power (b = 0), logarithmic ($a \to 0$), and exponential ($a \to -\infty$).

Finally, one transformation not included in Table 2, but included in the present analysis is the non-parametric. Essentially this means having as many free parameters as there are values that require translation. So, for an experiment that is designed to test the value function $v(x_i)$ for a finite number of x_i's, the non-parametric approach involves assigning $v(x_i) = a_i$ for all the i's and then treating the a_i's as free parameters.

The non-parametric approach has been applied to CPT by a number of authors (Abdellaoui, 2000; Bleichrodt & Pinto, 2000; Gonzalez & Wu, 1999; Hey & Orme, 1994; Mellers et al. 1992) who have noted its attractiveness because it assumes no specific functional form. However, it remains an interesting question as to whether the additional parameters of the non-parametric approach are warranted. In other words, are they extracting noise or conversely are the existing parametric forms of CPT missing some critical dimensionality that can so far only be expressed by a non-parametric function?

2.2 Risky Weighting Functions

The seven parametric risky weighting functions investigated in this paper are listed in Table 3 (again, the eighth function is the non-parametric approach). The first entry is the identity function where probabilities are not transformed. Using this means adopting EU.

[Table 3 About Here]

The second function is a power curve. Luce, Mellers, and Chang (1993) discuss the use of a power law to represent the risky weighting function in RDU. They show how this functional form characterizes the assumption of a condition referred to as *the reduction of compound gambles*. Further normative support for using a power curve comes from Prelec (1998) and Luce (2001). However, to the contrary, Camerer and Ho (1994) express doubt about the power curve as a risky weighting function since it cannot be used to fit what they refer to as *quasi-concave* and *quasi-convex* behavior.

The third equation in Table 3 is referred as the Goldstein-Einhorn function (GE). This function was originally conceived by Goldstein and Einhorn (1987), albeit not as a probability transformation. As they note, the function is an extension of Karmarkar's (1978, 1979) weighting function which he derived by assuming a linear relationship between the log weighting odds and the log probability odds that passes through the origin. GE is obtained by relaxing the second assumption and adding an intercept term:

$$\mathrm{Ln}\left(\frac{\pi(p)}{1-\pi(p)}\right)=r\mathrm{Ln}\left(\frac{p}{1-p}\right)+\mathrm{Ln}(s)$$

Consequently, the exponent parameter *r* mainly controls the curvature and the constant parameter *s* primarily controls the elevation of the risky weighting function. To see this second characteristic note that the value of the weighting function at $p = \frac{1}{2}$ is independent of r and given by $\pi(1/2) = s/(s+1)$.

Lattimore, Baker, and Witte (1992) used a general function that has GE as a special case for two outcome gambles. Their formulation highlights how GE is part of a larger family of contextually contingent weighting functions given by:

$$\pi(p_i) = \frac{sp_i^{\,r}}{sp_i^{\,r} + \sum_{k \neq i} p_k^{\,r}}$$

The fourth equation in Table 3 is referred to as Tversky-Kahneman (TK). This function first appeared in their original CPT paper (1992) and has subsequently been applied to various parametric analyses, including Wu and Gonzalez's (1996). In a footnote Wu and Gonzalez (p. 1686) point out that the TK function is a special case of a more general function which appears as the fifth equation in Table 3. No one appears to have made a normative case for adopting either of these functions.

Prelec (1998) developed an axiomatic family of weighting functions that was also explored by Luce (2001). The most general form is the last equation in Table 3, labeled PrelecII (Prl-II). This function has two nested cases--PrelecI and the power law. PrelecI (Prl-I) is obtained by setting the parameter *s* to unity. Similarly, the power law is obtained by setting the parameter *r* to unity. Qualitatively, the Prl-II function behaves similarly to the GE function.

2.3 Choice Functions

The current paper explores the explanatory performance of various stochastic choice functions. All the functions investigated here can be expressed as a function $P\left(V\left(g_1\right),V\left(g_2\right)\right)$ that can be added-on to a deterministic decision-making theory in order to accommodate stochastic behavior.

By contrast, Becker, DeGroot, and Marschak (1963b) have described an alternative class called "random utility models". They involve making the underlying decision theory stochastic. For example, CPT might be specified with variable, rather than fixed, parameters. In this case, when choosing between prospects, a person's preferences will depend on the parameter values in operation at that instance. The current analysis excludes this approach for two main reasons. Firstly, the add-on functions tested offer enough variation and therefore represent a good starting point. Secondly, there is evidence to suggest that the add-on approach might perform better than random utility and certainly not significantly worse (Loomes & Sugden, 1998).

The tested choice functions appear in Table 4. The first function, labeled Constant Error, was originally introduced into the risky decision literature by Harless and Camerer (1994) and Wakker, Erev, and Weber (1994). For this function, there is a fixed chance of expressing true preference and a fixed chance of not. This approach has also been dubbed the "trembling hand".

[Table 4 About Here]

The next two stochastic functions are the Probit and Logit transformations. Both of these have been used extensively in the analysis of risky decision-making data. For example Hey and

Orme (1994), Buschena and Zilberman (2000), and Carbone and Hey (2000) used the Probit approach whereas Birnbaum and Chavez (1997), Birnbaum, Patton, and Lott (1999), and Carbone and Hey (1995) used Logit.

Finally, Luce (1959) described the fourth function, which is a generalization of Herrnstein's (1997) matching law. The Luce function differs from the other functions in that it depends on the ratio of the prospect values rather than their absolute difference. As such, the Luce approach is consistent with Weber's law, making it the most psychologically motivated rule tested.

3. Previous Findings

This section reviews various prior results concerning the comparative performance of different functional forms of CPT. However, one immediate observation is that such comparative tests are sparse and there does not appear to have been any previous attempt to systematically investigate a full combinatorial pattern of functions.

3.1 Value Functions

As part of their meta-analysis, Camerer and Ho (1994) fitted CPT using three different value functions together with a TK risky weighting function. They found that the Pwr function provided the best fit followed by a function referred to as the Log-Quadratic (see earlier footnote). The Expo function was the weakest.

Likewise, Blondel (2002) examined the explanatory power of the Pwr, Expo, and Lin functions for various theories, including CPT with a TK risky weighting. The results showed that the Expo function fit the data marginally better than the Pwr function. The Lin function was the worst.

Finally, Birnbaum and Chavez (1997) tested several parametric forms of CPT. As part of those tests, the Pwr and Expo forms of the value function were tested using a Pwr curve for the risky weighting function and a Logit transformation for the choice function. They found that the Pwr function fit the data better than Expo.

In summary, there have been three limited tests of different value functions for CPT that span four of the seven parametric forms tested here. It is perhaps surprising that no one appears to have tested the descriptive accuracy of the Quad function given its role in finance. The available tests seem to favor Pwr.

3.2 Risky Weighting Functions

Gonzalez and Wu (1999) fitted non-parametric curves to participants and then tested different functional forms on these estimates. They concluded that the GE and Prl-II functions could not be rejected, but did reject the single parameter TK and Prl-I. Likewise, Bleichrodt and Pinto (2000) fitted the same four functions and concluded that the two-parameter models outperform the single. Thirdly, Sneddon and Luce (2001) fitted three parametric forms to non-parametric estimates and concluded that the best to worst fitting risky weighting functions were Prl-II (.61), Pwr (.22), and TK (.09) (percentage is the proportion of participants for which the function provided the best fit).

In contrast, Wu and Gonzalez (1996) report the opposite result, although they fitted functions to pooled participant data. The ranking of the risky weighting functions in order of best unadjusted fit were WG, TK, GE, and Prl-I. However, adjusting for degrees of freedom, it is the TK single parameter model that outperforms the others. Hence, it may be that whilst a single parameter function is adequate for aggregate population data, the two-parameter functions are preferable when fitting individual participants.

Finally, Buschena and Zilberman (2000) fit TK and Pwr risky weighting functions in conjunction with a non-parametric value transformation and several different stochastic processes. They concluded that the Pwr function marginally outperformed the TK function.

In summary, prior results concerning the best risky weighting function are patchy and equivocal. Whilst the majority of studies favor two-parameter versions for fitting individual participant data, these results are quite marginal and no specific two-parameter version has been singled out.

3.3 Choice Functions

In 1997, Ballinger and Wilcox presented evidence against Constant Error by examining several predictions of the hypothesis and rejecting them. Likewise, Loomes and Sugden (1995, 1998) reported several tests that rejected the Constant Error approach.

The Constant Error approach was also tested within an EU framework and rejected by Carbone (1997). However, in a follow up paper, Carbone and Hey (2000) reported a similar analysis across several different risky decision-making theories and found that Constant Error outperformed the Probit approach under CPT. Since other authors have found that reversal rates do change across different pairwise questions, contrary to what Constant Error predicts, then one possibility is that Carbone and Hey's experimental stimuli did not have sufficient range to refute the function.

For example, consider "a 40% chance of £40 or a 50% chance of £30" and "a 5% chance of £40 or a 95% chance of £30". It seems clear that these are difficult and easy questions respectively that will be characterized by high and low reversal rates, refuting Constant Error. However, if a question set contained a relatively homogenous mix of all difficult or all easy questions, then the Constant Error approach may fit the data.

Interestingly, Loomes, Moffatt, and Sugden (2002) reused Loomes and Sugden's (1998) data to test a Probit model with a Constant Error term added and found that using this hybrid model improved on both. Since their hybrid incorporated a decay parameter, it indicated that the Constant Error process was generating .111 preference reversals at the start of the experiment and .014 at the end. It was apparent that during the course of the experiment participants were learning, which led to a decrease in trembling hand.

In summary, several authors have tested different approaches for adapting CPT in order to account for stochastic choice and have generally found evidence against Constant Error.

4. Experiment

The objective of this paper is to compare the explanatory performance of the 256 different functional forms of CPT generated by a combinatorial pattern of eight value functions, eight risky weighting functions, and four choice functions. Explanatory performance was assessed by fitting each of these models to individual participant data collected by asking them to choose between two prospects across a set of ninety such pairwise questions. The data collection and selection method is detailed in the next two sub-sections followed by an account of the model fitting process. The section then ends with an examination of the experiment's reliability.

4.1 Method

The question set consisted of 96 stimuli with 6 practice and 90 actual questions. Each question consisted of a pair of two outcome prospects, with participants indicating which they would prefer to own. A list of these stimuli is provided in the appendix together with the aggregate participant preferences. There are two main reasons why pairwise choice data was used, as opposed to indifference data[4]. Firstly, people appear to have problems judging equivalence (Lichtenstein & Slovic, 1971) and its elicitation can be influenced by the presence of other options and reference levels (Stewart et al. 2002). Secondly, indifference data cannot be used to examine choice functions.

For each question, the prospects were displayed side by side in words (e.g. “a 90% chance of £10,000 or a 10% chance of £0”). The question order and prospect sides were randomized. Participants were run individually in soundproof cubicles. Stimuli were presented on a 17 inch computer screen using a purpose built application written in Java. Participants expressed their choice by mouse click and were not permitted to express indifference. The practice questions were run first, followed by the experimental questions. The experiment was conducted on 96 participants (46 male) who were recruited on campus during term time at the University of Warwick.

The number of questions in the experiment was a balance between what was practicable for one session and what was needed to fit the models. Participants took 17.6 (*SD* = 5.9) minutes on average, equivalent to 11 seconds per question. No data was discarded from obvious failure to

[4] Indifference data is where participants are asked to compare two prospects and then adjust one feature of one of the prospects until they judge the two prospects to be equally attractive.

undertake the task, though data was missing on 7 questions for one participant due to a computer problem. This data was included in the analysis, but no findings depend on it.

4.2 Question Set and Incentives

Three aspects of the stimuli used in the experiment are now discussed in more detail.

Firstly, an important aim of this research is to identify the best functional form of CPT for use in real world applications. As such, consider that the students in this experiment were soon going to graduate and face employment choices. Since average UK income is about £25,000 and university leavers can expect to work at their first employer for several years, these forthcoming employment decisions involved stakes in the tens of thousands of pounds. Likewise, people often have to make choices such as whether to buy or rent a home (the average UK home now costs about £200,000), where to place their retirement savings, what car to buy, how to educate their children, and so forth.

Accordingly, the amounts used in the current experiment were chosen to represent the important financial decisions an average person can expect to encounter in their lives. Hence, the choices offered to participants consisted of gambles constructed from £0, £2,500, £5,000, £10,000, £20,000 and £40,000. As a result, the average expected value of each question was approximately £9,000 (*SD* = £6,500). Some of the implications of selecting this magnitude of outcomes are discussed presently.

The second issue concerns the actual design of the question set. By combining the different outcome amounts chosen with a range of different probabilities (running from .1 to .9 in .1 increments), every possible permutation of pairwise choice question can be generated. Of these, in order to reduce the scope for participant editing, only non-dominated questions and those that included at least one £0 outcome between the two gamble choices were considered for inclusion in the current experiment. This left 1,889 questions from which 90 questions were extracted.

Two criteria were used for this extraction process. Firstly, questions were selected that could divide the population and minimize the parameter estimation errors on individual participants. Intuitively, a good question set needs to contain a range of questions that can separate the relatively risk seeking from the relatively risk averse participants. Conversely, the question set needs to avoid superfluous questions where all the participants are likely to make the same choices. The second criterion concerns multi-collinearity. In one of their footnotes, Gonzalez and Wu (1999, p. 157) report that there were substantial interactions between the CPT parameters they estimated. For example, risk taking behavior could be equally well explained by having a risky weighing function with greater elevation combined with a lower value function exponent and vice versa. Accordingly, the current question set was also designed to minimize these parameter interaction problems (further details on the question selection procedure and its effectiveness are available from the author on request).

The final experimental detail discussed in this sub-section relates to participant motivation and financial incentives. Psychologists generally hold that the intrinsic motivation of participants is enough to conduct experiments without the need for further incentives (Hertwig & Ortmann, 2001) and this view has been substantiated by a number of empirical studies. For example, Camerer and Hogarth (1999, JRU) reviewed 74 experiments and observed that financial incentives have no mean performance effect across a variety of judgement tasks. However, they also note that whilst this is the case on average, there is a pattern of evidence suggesting that incentives are helpful for certain tasks (e.g. those that are repetitive and dull) and vice versa. On risky decision-making tasks they draw three conclusions. Firstly, incentives often decrease the amount of noise in responses. Secondly, increased incentives do not change average behavior substantively. Thirdly, when incentives do influence behavior, participants have tended to exhibit greater risk aversion. This summary is also consistent with an earlier review of 31 studies carried out by Smith and Walker (1993).

Because the current study is concerned with testing different functional forms of CPT, the potential noise reduction engendered by using financial incentives is an important benefit. Clearly, noisier data would make differentiating between the competing functional forms more difficult and thereby weaken the discriminatory power of the experiment. For this reason, participants were paid £3 plus an incentive payment that was determined at the end of the experiment by selecting one of the questions at random and playing the participant's chosen prospect. Given the amounts involved in the gambles, it was impractical to offer participants the chance of playing their choices for real. Hence, for the incentive payment, all outcomes were re-scaled so that the maximum outcome was £5. On this basis, the average incentive payment was £2.13.

Whilst it is hoped that this incentive will have improved the quality of the experimental data, it may also have had an effect on the way that the stimuli were perceived by the participants. For example, Kachelmeier and Shehata (1992) obtained certainty equivalents from two sets of Canadian students. The first group was given a set of stimuli and a financial incentive. The second were given the same stimuli with the outcomes multiplied by 100 and financial incentives paid out at a conversion rate of 100:1 (i.e. they had nominally larger stimuli, but effectively the same incentives as the first group). They concluded that there was no systematic difference between the two groups.

So, whilst the sizes of the outcomes were still set at a level that was representative of the important financial decisions people encounter, it is possible that this objective was then negated by the adopted incentive structure. Nevertheless, the advantage of retaining the financial incentives to reduce experimental noise was judged worthwhile.

4.3 Model Fitting

Maximum Likelihood Estimation (MLE) was used to fit the models to the individual participant data. MLE has been used extensively in modeling risky decision-making (e.g. Birnbaum, Patton, & Lott, 1999; Camerer & Ho, 1994; Carbone & Hey, 2000; Hey & Orme 1994; Loomes, Moffatt, & Sugden, 2002). The approach can be summarized in three steps. Firstly, assuming a given model and parameter values, the probability of how a participant will choose, $f\left(g^1 \middle| g^2, \theta\right)$, is predicted. The probability of observing the participant's actual data is then given by $\mathcal{L} = \prod_s f\left(g^s \middle| g^{\bar{s}}, \theta\right)$. For various reasons, including computational convenience, this likelihood is usually transformed by a natural logarithm. Finally, the model parameters are adjusted to find the parameter value $\tilde{\theta}$ that maximizes $\ln \mathcal{L}$.

The maximized parameter values are therefore the most likely values to have generated the observed data within the architectural constraints of the given theory. Furthermore, the maximized likelihood provides a measure of the theory's fit. For example, assuming no information is captured by the theory, it would assign a .5 chance to each pairwise choice and the log likelihood across the ninety questions would sum to $90 \ln(.5) = -62.4$. Conversely, if the theory fit perfectly, the log likelihood would be $90 \ln(1.0) = 0$.

The different forms of CPT were fit to each participant individually to avoid the hazards of making a single agent assumption (Myung, Kim & Pitt, 2000). However, this did mean fitting 256 forms of CPT to 96 participants. Performing these 24,576 separate MLE hillclimbs was accomplished using an iterative fixed step approach implemented in a purpose built software application. Whilst more sophisticated searches, such as gradient sensitive hillclimbing, random restart techniques, and simulated annealing are possible (Rich & Knight, 1991), this simpler method was used because it was robust enough to be evenly applied across the range of models and data qualities presented. Clearly, this consistency of analysis was important for ensuring the comparability of the results. A combination of algorithm testing, output cross-checking, and process inspection indicated that the hillclimb was working effectively.

Finally, the unadjusted maximized log likelihood can be a misleading indicator of which model provides the best explanatory performance when models have different numbers of free parameters (Grünwald, 2000; Myung & Pitt, 1997). Essentially, the more complex a model, the better it fits data. For this paper, an adjustment technique developed by Akaike (1973) was adopted. The Akaike Information Criterion (AIC) is given by AIC = $-2\mathrm{Ln}\mathcal{L} + 2k$, where k is the number of free parameters in the model. Hence, the better the explanatory power of a model, the lower its AIC.

There are three reasons why AIC was used. Firstly, the criterion was derived for identifying superior descriptive models rather than the likelihood of a given model being true (Burnham & Anderson, 2002; Myung, 2000). The former is more in-line with this paper's purpose. Secondly, AIC has been successfully used in other empirical studies of risky decision-making (Carbone & Hey, 1994, 1995; Hey & Orme, 1994). Finally, a Monte-Carlo simulation performed for this paper on the effectiveness of various model selection approaches favored AIC.

4.4 Reliability

Having fit the different forms of CPT to each participant's data, the reliability of the output was assessed based on the maximized log likelihoods found by the hillclimb. Figure 2 plots the distribution. The log likelihoods range from –64.2 to 0.0 or, equivalently from $e^{-64.2/90} = 0.49$ through to 1.00, in terms of the geometric mean probability predicted for actual responses.

[Figure 2 About Here]

This lower bound shows where the model has completely failed to describe the data. Conversely, the upper bound indicates where the model has fit exactly. Reassuringly, few of the log likelihoods lie at either extreme. The inter-quartile range is –46.8 to –25.3, or equivalently .59 to .76 in mean probabilities and the median is –37.2 or .66.

From Table 1 it can be seen that previous authors have found that participants changed their preferences about .25 of the time on repeat questions. Hence, a good test is whether a model can approach assigning a .75 chance to actual outcomes. Using this benchmark, the median probability of .66 and quartiles that range from .59 to .76 appear reasonable. In other words, these predicted probabilities could have been much lower if some part of the experimental process had not worked. As it is, these results indicate that the participant data contains lawful preferences that the MLE fit of CPT has managed to capture.

5. Results

This section details the main experimental findings. In the second sub-section, the explanatory powers of the individual functions are reported. In the third, the interaction effects between these individual functions are examined. However before these two main results are reported, the experiment's aggregate parameter estimates are compared with those found by previous authors.

5.1 Parameter Estimates

A comparison of the current parameter estimates with those reported by other authors' appears in Table 5. The values shown in this table are medians unless otherwise noted. The distributions of the parameter estimates for the Pwr value function and GE risky weighting function across participants have also been plotted in Figures 3, 4, and 5. These figures have tick marks showing the other author values from Table 5. All the estimates reported from the current experiment use the Logit transformation.

[Table 5 About Here]

[Figure 3 About Here]

[Figure 4 About Here]

[Figure 5 About Here]

As can be seen in Table 5 and Figure 3, the current parameter estimates for the Pwr function are at the lower end of previous findings. So, given the spread of prior findings, does the current estimate represent a statistically significant departure? Using the average median and variance of the prior results, the current median of 0.19 has a z-score of $z = -1.31$. Using a two-tailed normal distribution test, the current findings can therefore be accepted as drawn from the same experimental population. In other words, whilst the current experiment has returned a low figure, the result is not outlandish.

A similar picture emerges for the risky weighting function. For example, the current median of the *s* elevation parameter is 1.40. Compared to the eight prior results this is high, meaning that the current participants had relatively optimistic risky weighting. The prior medians average 0.65 ($SD = 0.25$) so the current median, at $z = +2.98$, is a significant departure from these previous results (two-tailed $p < .01$). Conversely, for the curvature parameter, *r*, shown in Figure 5, the majority of studies have found the exponent to be below unity ($M = 0.79$, $SD = 0.36$). In this case the current median is 0.96, an almost straight function. This result is not a significant departure ($z = +0.47$, *ns*). So although the median risky weighting function is not distinctly inverse S-shaped, the current parameter estimate appears credible. Against this backdrop, three potential sources for these observed parameter differences are now discussed nevertheless.

Firstly, as described earlier, the current stimuli set was designed to minimize the parameter interaction effects reported by other authors. Analysis of some of the current parameter errors suggests that these multi-collinearity problems were indeed reduced in the current experiment. Exactly how this will have influenced the current aggregate parameter levels is unclear. Nevertheless, the use of these novel stimuli will doubtless have had some positive implications for the findings whilst potentially also causing some divergence from prior results.

Secondly, the size of the outcomes in the stimuli set are large compared with previous studies. This was to make the experimental findings more applicable to real world situations, though there are reasons to believe that participants might have been unable or unwilling to accurately express their preferences over such large outcomes, particularly given the experimental incentive employed. Nonetheless, since previous studies have shown that changes in outcome scale can affect choice behavior (simplistically making choices more risk averse) it is possible that this contributed to the present findings.

Finally, even though the financial incentive required a conversion rate, prior evidence also suggests that including an incentive can induce participants to behave differently. Here the aggregate value function exponent has been compared to seven other studies. Only one of those employed an incentive. Similarly, for the risky weighting function, the current estimates were compared with eight other studies, of which only two used incentives. On that basis, this may be another factor that has contributed to the slightly different aggregate parameter levels observed.

Overall there are many reasons why participants may behave differently in one study versus another. Indeed, Weber and Kirsner (1997) managed to observe such differences in risky decision-making simply by changing the font size of their stimuli. Differences between the aggregate parameter estimates in the current paper and prior studies therefore appear acceptable. Firstly, the differences are not too large. Secondly, there are several possible sources for the differences and these arise from consciously adopted features of the experimental design. Finally, it is worth noting that the specific parameter values observed in the current paper are the byproduct of testing the explanatory performance of different versions of CPT. As such, their levels are not critical to the paper's main findings.

5.2 Function Performance

The next two sub-sections address the main topic of this paper: what functional forms of CPT provide the best descriptive account of experimental data? Individual performance is considered first and then interactions are discussed in the next sub-section.

Table 6 provides an initial perspective by listing pairwise comparisons for functions that are nested. For example, the model formed by a Pwr value function, TK risky weighting function, and Logit function (the notation *Pwr/TK/Logit* is adopted hereafter) is nested within Non-Para/TK/Logit. In other words, you can create the former by constraining the latter. In these nested cases, which model has the greater explanatory power can be examined using Likelihood Ratio testing[5]. Table 6 gives each of the nested relationships and the proportion of Likelihood Ratio tests (across participants and ancillary functional forms) for which the larger function yields a significant improvement over the smaller one. For example, in 32 percent of tests, the Bell value function explains the data significantly better than the Expo function and conversely, in 68 percent of cases the more complicated function is unwarranted. Where the larger function is appropriate in more than half the tests, the cell is shaded.

[Table 6 About Here]

[5] For this test, under the null hypothesis that both models provide an equally good description of a given participant's data, twice the difference between the two models' maximized log likelihoods has a χ^2_k distribution, where k is the number of parameter restrictions applied to obtain the nesting. In the example in the text, the parametric model Pwr/TK/Logit uses 3 parameters and its non-parametric equivalent uses 11, so $k = 8$

Table 6 therefore highlights the first finding of the study--there is not much shading. In fact, the more complicated functions are only better in three out of twenty comparisons. One implication of this is that the non-parametric approach offers a generally inferior account of participant data compared to the parametric approach, with the only exception being Quad. Note that since every parametric function is nested in the non-parametric, this comparison is comprehensive. Likewise, the table shows that the Prl-I risky weighting function is preferable to the Prl-II function and that the TK function is preferable to WG. This result in favor of the one parameter versions is interesting given the inconclusive nature of the prior evidence.

A more systematic comparison of the functions is provided in Table 7. Here, all the fitted models were ranked within participant using the AIC measure of explanatory performance (1 = best, 256 = worst). These AICs and their within participant ranks form the basis of the remaining results[6]. The table lists the functions in order of ascending average rank (the "Full" column). Since there are 256 ranked models, a rank below 128.5 is worse than average. The "Subset" column will be discussed presently. The table also shows average AIC.

[Table 7 About Here]

For the value functions, the Quad and Lin forms are notably weak. This reinforces the earlier assessment of Quad and questions the descriptive credibility of mean-variance based financial portfolio analysis. Conversely, Pwr and Log perform well. Also of interest is the

---

[6] Using ranks is consistent with the work of previous authors (e.g. Buschena & Zilberman, 2000; Hey, 1995; Hey & Orme, 1994) and avoids distortion by outliers. For example, this might be generated from unusually skewed results caused by one-off problems with the hillclimb. When 24,576 individual model fits are being undertaken, the odd misfire seems inevitable.

comparative performance of Pwr and Expo. In the prior literature these functions have been difficult to distinguish. In the current experiment Pwr is better. A paired t-test of the average ranks is highly significant at t (3071) = 24.84, $p < .0001$. For the risky weighting functions, in contrast to the pairwise nested results, the two parameter functions Prl-II and GE perform best. This reversal against the single parameter functions is discussed presently. Meanwhile, Pwr, Lin, and Non Para all fall below average. Lastly, two choice function findings are noteworthy. Firstly, in line with the prior literature, Constant Error seems weak. Secondly, the Luce function offers the marginally best fit.

The contrasting results between Tables 6 and 7 for the risky weighting function is explored further by Figure 6. It shows the average ranks for the different combinations of parametric value and risky weighting functions. The value functions appear in performance descending order and the risky weighting functions are plotted as separate lines. For clarity only Prl-I and Prl-II are labeled. The figure shows how Prl-I produces the highest average ranking when it is paired with value functions that have good explanatory power. By contrast, for the two functions that produce below average rankings (Quad and Lin) the Prl-II function yields the best performance and the Prl-I's rankings are almost the worst. Indeed, these Prl-I rankings are so poor that the overall averages shown in Table 7 actually favor Prl-II in spite of Prl-I's better performance for six out of eight of the value functions. TK and WG exhibit a similar, though less extreme pattern.

[Figure 6 About Here]

Hence the performance of single and double parameter risky weighting functions depends on what they are paired with. Presumably where the surrounding functions have a worse fit, the extra risky weighting parameter can play a compensating role and thereby arrest a bigger fall in model performance. In the light of this effect, the column labeled Subset provides average rankings for each function only when combined with other functions that exhibit above average performance (excluded functions are identifiable from the blanks in the table). This restricted test confirms the superiority of the single parameter risky weighting functions under these conditions. However, the effect on the value and stochastic functions is negligible.

5.3 Function Interactions

In the prior sub-section, the explanatory power of individual functions was reported. However, evidence was also presented that suggests there are interaction effects between them. In other words, combinations of functions yield better and worse descriptive accounts of participant data than is implied by their individual performance. In this sub-section, the existence of interaction effects is tested directly and then the best performing model combinations are reported.

The presence of interaction effects was examined using a three factor ANOVA. Table 8 shows the details of this test for the full set of functions and the subset described in the prior sub-section. The sphericity of the various factors can be rejected everywhere (e.g. on the full set of functions, the value function sphericity is rejected with $\chi^2_{27} = 531.36$, $p < .0001$ using Mauchly's test). On that basis, the main effects and their interactions are all tested using the conservative Lower Bound approach.

[Table 8 About Here]

Where the full set of functions are tested, all the interaction terms are significant. Once the poorly performing functions are removed, the scores are lower but still mostly significant. This finding has two implications. Firstly, the average performance of individual functions described in the prior sub-section provides an incomplete picture of which versions of CPT perform best. Secondly, it is interesting to note that the choice function participates in these interaction effects. In other words, it does not make sense to hypothesize different forms of CPT without specifying how stochastic choice is being handled, since this latter specification will alter the descriptive adequacy of the former.

[Table 9 About Here]

Table 9 lists the top ten complete models that have the highest average rankings across participants. The variation in model performance was tested by comparing each model's average rank with that of the best model, Pwr/Prl-I/Logit, using a pairwise t-test. As shown, only 7 models cannot be eliminated at a .95 confidence level. The model with the highest average rank that is rejected by the t-test is Pwr/Prl-I/Probit, which is so similar to the top model that its performance would have to be nearly identical in order to avoid rejection. It is also interesting to note that Pwr/GE/Logit, a widely used version of CPT, is rejected.

6. Discussion

In this last section, some of the implications of the experimental findings will be reviewed. The first sub-section discusses four aspects of the results in more detail. Then some potential future research ideas are described. Finally, the section concludes with a brief summary of which functional form of CPT is recommended under which circumstance.

6.1 Implications of the Results

This study of the comparative explanatory power of 256 different functional forms of CPT has highlighted several conclusions.

Firstly, the individual function results are interesting. For example, the good performance of Luce, a choice rule that has not been previously tested on CPT, shows that there is considerable scope for investigating new stochastic functions. Observations on what might be examined in future are detailed in the next sub-section. Likewise, the poor performance of the Quad value function questions the descriptive accuracy of mean-variance portfolio analysis. In other words, despite its undoubted algebraic convenience, variance does not seem to capture people's risk aversion as well as other risk measures. It is perhaps difficult to overstate the potential implications of this, given the current widespread use of variance as a risk measure in both theoretical and applied finance.

Secondly, the contention that non-parametric models are somehow preferable has been tempered. Whilst this parametric freedom may be necessary where the shape of the function is itself under investigation, this reasoning does not apply to other situations. In these cases it is explanatory power that counts. The current results show that parametric forms of CPT generally fit risky decision-making data better than non-parametric ones. This reinforces a comment made by Luce (2000, p. 28) where he noted "…much of the data on risk aversion can be explained either in terms of the form of the utility function or in terms of the weighting function or both. In my opinion, the existence of very unspecified functions simply means that the theory is not adequately characterized..."

Thirdly, a related finding is that the two parameter versions of the risky weighting function are less attractive than the one parameter versions (i.e. Prl I is preferable to Prl II). While the more complicated functions can play a role in compensating for inadequacies in other parts of the model, as Luce noted, this is not an attractive property. It is surely better to specify a theory where the various components are not called upon to substitute for each other. The results presented here highlight how one route for improving specification is through using simpler functions that will act more independently.

The fourth observation is that even when the set of functions used to generate CPT is restricted to the more suitable ones, interaction effects remain. As mentioned above, this is a potential criticism of CPT itself. Alternatively, it could be construed as a criticism of available experimental techniques. In either case, it is useful to note that these interaction effects extend to the choice function (Buschena & Zilberman, 2000). As such, these findings imply that no risky decision-making theory is complete until it includes a stochastic choice mechanism. As Loomes and Sugden (1995, p. 648) put it, "…future theoretical and empirical work should not regard the stochastic specification as an 'optional add-on,' but rather as an integral part of every theory which seeks to make predictions about decision making under risk and uncertainty".

6.2 Future Research on Choice Functions

The findings presented here suggest several areas for future research. This sub-section focuses on one of these--the opportunity to extend and expand the empirical testing of choice functions in risky decision-making. Several avenues for developing new choice functions are suggested.

The Logit function plotted in Figure 1 belongs to a sub-class of choice functions referred to as "Strong Utility" in Ballinger and Wilcox (1997) and "Fechner" by Becker, DeGroot, and Marschak (1963b). This sub-class was originally axiomatized by Debreu (1958) and takes the form:

$$P\left(V\left(g_1\right),V\left(g_2\right)\right)=F\left(V\left(g_1\right)-V\left(g_2\right)\right)$$

where $F(\ )$ is some appropriately behaved S-curve. This form can therefore be re-written:

$$P\left(V\left(g_1\right),V\left(g_2\right)\right)=\text{Prob}\left(V\left(g_1\right)-V\left(g_2\right)+\eta\geq 0\right)$$

where η is a random variable symmetric about zero. Accordingly Fechner models have also been referred to as white noise models. The first three functions in Table 4 are Fechnerian. The class of Fechnerian functions contains as many members as there are noise distributions and to-date only a handful of these have been examined. Hence, one interesting development would be to evaluate some of these other Fechnerians, such as the Laplace and Cauchy versions described by Ballinger and Wilcox (1997).

Next, as Debreu (1958) has pointed out, one characteristic of Fechner models is that their iso-preference curves are parallel lines given by:

$$V\left(g_1\right)-V\left(g_2\right)=F^{-1}\left(a\right) \text{ where } a\in\left[0,1\right]$$

By contrast, the Luce function has radial iso-preference curves, consistent with its dependence on the ratio, rather than the difference, of prospect values. It appears that to-date, only Fechnerian choice functions have been applied to CPT. The current analysis is therefore not only the first time that the Luce function has been applied to CPT, but also the first time that a non-Fechnerian function has been used.

Since the Luce rule performed marginally better than the Logit or Probit approaches in this analysis, it might also be useful to investigate other non-Fechnerian choice functions. For example, the Luce function belongs to a sub-class of choice rules that Becker, De Groot, and Marschak (1963a) refer to as the “strict utility model for wagers”, given by:

$$P\left(V\left(g_1\right),V\left(g_2\right)\right)=\frac{F\left(V\left(g_1\right)\right)}{\sum_i F\left(V\left(g_i\right)\right)}$$

where $F(\ )$ is some increasing function. The Luce choice surface can be obtained when $F(\ )$ is a power function and Herrnstein's matching law is obtained by using the identity[7]. This sub-class might provide some further interesting functions, such as when $F(\ )$ is logarithmic and the iso-preference lines are power curves.

Similarly, another non-Fechnerian sub-class of choice functions can be defined using a noise term η such that:

$$P\left(V\left(g_1\right),V\left(g_2\right)\right)=\text{Prob}\left(\frac{V\left(g_1\right)}{V\left(g_1\right)+V\left(g_2\right)}+\eta\geq 0.5\right)$$

Since this definition is analogous to the definition of the Fechner sub-class, it raises the possibility of performing a paired test between the two sub-classes. Essentially the question would be "for a given set of noise processes, does the absolute difference or ratio provide a better explanatory fit?"

Finally, another potential area for choice function development concerns the use of other independent variables beyond prospect features. This question has already been examined by others (Buschena & Zilberman, 2000; Hey, 1995) and it has been reported that, for example, the certainty of a participant's preferences increases with the amount of time taken to answer. Nevertheless, there remains a host of other variables that might be incorporated into stochastic

[7] Note that this sub-class of choice surfaces overlaps with the Fechnerian sub-class because the Logit choice surface results when $F(\ )$ is exponential. It can easily be shown that this is the only differentiable Fechnerian member of the strict utility sub-class

modeling, including number of prospect outcomes, question number, display format, time of day, and so on.

6.3 Conclusion - Which Functions To Use?

The main purpose of this paper has been to examine the explanatory performance of various functional forms of CPT. The paper therefore concludes with some suggestions regarding which functional forms are appropriate under which circumstances.

A variety of authors have published studies where CPT's shape or some shape property has been under investigation (Bleichrodt & Pinto, 2000; Wu & Gonzalez, 1996). In this case, the non-parametric form of CPT is clearly appropriate. Whilst it may not represent the most predictive means for analyzing participant data, it still represents the only unbiased way of conducting a descriptive investigation.

Next, some previous authors have been interested in comparing different risky decision-making theories based on their explanatory powers (Hey & Orme, 1994). An implication of the current results is that the form Pwr/Prl-I/Logit should be used for CPT. However, it remains an open question as to whether the explanatory variation across different risk theories is greater or smaller than the variation across the different functional forms of those theories. This raises the possibility that future testing of different theories should extend to testing different functional forms too using the better performing functions listed in Table 7. This would also address a related and interesting question "does a functional form that works well in one theory typically work well in another?"

Finally, there may be occasions when a practical application in risky decision-making is being explored and the experimenter selects CPT as the framework for extracting a person's underlying preferences. Two examples are considered. Firstly, several authors have sought to measure the transportability of risk preferences between different decision-making paradigms (Isaac & James, 2000; Slovic, 1972). In these cases, the degree of transportability observed will in part rest upon how effectively the risk preference signal is being extracted from people's experimental responses. Hence, where CPT is being used, it is suggested that Pwr/Prl-I/Logit be adopted in order to maximize the information extraction from participant data and thereby increase the chances of detecting a relationship with other measures of behavior.

The second instance concerns applied decision-making. This is similar to the previous application except that the experimenter is not constrained to one functional form, or indeed one decision-making theory. For example, this might involve helping a patient make a medical decision, evaluating different legal options or assisting an investor with their asset allocation. In such cases, the researcher could ask the person to make a variety of relevant risky decisions. The analysis would then help the researcher estimate the person's dominant preference for the question at hand. This approach might be better than simply asking the person for their view on the question directly because their single response would contain stochastic error. Under these circumstances the researcher should select several of the CPT versions in Table 9 and then combine the forecasts using a technique such as Bayesian model averaging (Hoeting et al., 1999).

Appendix

The appendix contains the set of 90 pairwise questions used as stimuli in the experiment. The questions are numbered 7 to 96 since there were six practice questions at the start of the experiment that are not reported here. Each question comprises two prospects, denoted G1 and G2 in the tables.

Finally, the questions have been broken into three tables of 30 questions each. These tables relate to different question types. Respectively these are of the form "$(\pounds x,\ p;\ \pounds 0,\ 1-p)$ vs. $(\pounds x,\ q;\ \pounds y,\ 1-q)$", "$(\pounds x,\ p;\ \pounds 0,\ 1-p)$ vs. $(\pounds y,\ q;\ \pounds z,\ 1-q)$", and "$(\pounds x,\ p;\ \pounds 0,\ 1-p)$ vs. $(\pounds y,\ q;\ \pounds 0,\ 1-q)$".

Stimuli Question Set

Question	G1A1	G1P1	G1A2	G1P2	G2A1	G2P1	G2A2	G2P2	G1%
7	2,500	40	5,000	60	0	20	5,000	80	.75
8	10,000	90	40,000	10	0	60	40,000	40	.92
9	2,500	90	20,000	10	0	70	20,000	30	.94
10	2,500	30	5,000	70	0	10	5,000	90	.58
11	2,500	40	5,000	60	0	10	5,000	90	.55
12	2,500	50	5,000	50	0	20	5,000	80	.72
13	2,500	60	5,000	40	0	20	5,000	80	.67
14	2,500	70	5,000	30	0	30	5,000	70	.77
15	5,000	90	40,000	10	0	70	40,000	30	.95
16	2,500	60	10,000	40	0	10	10,000	90	.38
17	2,500	70	10,000	30	0	10	10,000	90	.33
18	2,500	90	10,000	10	0	60	10,000	40	.91
19	2,500	50	20,000	50	0	10	20,000	90	.38
20	2,500	90	20,000	10	0	50	20,000	50	.68
21	2,500	90	20,000	10	0	60	20,000	40	.83
22	2,500	80	40,000	20	0	50	40,000	50	.68
23	2,500	80	40,000	20	0	60	40,000	40	.84
24	2,500	80	40,000	20	0	70	40,000	30	.97
25	2,500	90	40,000	10	0	60	40,000	40	.84
26	2,500	90	40,000	10	0	70	40,000	30	.86
27	2,500	50	5,000	50	0	10	5,000	90	.48
28	2,500	60	5,000	40	0	10	5,000	90	.49
29	2,500	70	5,000	30	0	10	5,000	90	.47
30	2,500	70	5,000	30	0	20	5,000	80	.60
31	2,500	80	5,000	20	0	10	5,000	90	.49
32	5,000	50	10,000	50	0	20	10,000	80	.68
33	5,000	90	20,000	10	0	60	20,000	40	.86
34	5,000	90	40,000	10	0	40	40,000	60	.67
35	5,000	90	40,000	10	0	50	40,000	50	.72
36	5,000	90	40,000	10	0	60	40,000	40	.86

Note. Each line of the table details a question from the experiment. Each question involved participants choosing between two prospects--a "G1P1 chance of G1A1 and a G1P2 chance of G1A2" versus a "G2P1 chance of G2A1 and a G2P2 chance of G2A2". Here A1 and A2 refer to the lower and higher sums that can be won for a given gamble and P1 and P2 are the chances of winning them. The final column shows the proportion of participants that chose G1.

Stimuli Question Set (Cont'd)

Question	G1A1	G1P1	G1A2	G1P2	G2A1	G2P1	G2A2	G2P2	G1%
37	0	10	10,000	90	2,500	60	40,000	40	.25
38	0	30	10,000	70	5,000	80	20,000	20	.09
39	0	50	10,000	50	2,500	90	40,000	10	.16
40	0	10	20,000	90	2,500	40	40,000	60	.33
41	0	90	40,000	10	10,000	10	20,000	90	.01
42	0	90	40,000	10	10,000	70	20,000	30	.01
43	10,000	30	20,000	70	0	50	40,000	50	.82
44	10,000	40	20,000	60	0	50	40,000	50	.81
45	10,000	60	20,000	40	0	50	40,000	50	.78
46	10,000	70	20,000	30	0	50	40,000	50	.79
47	10,000	70	20,000	30	0	60	40,000	40	.92
48	10,000	80	20,000	20	0	60	40,000	40	.92
49	10,000	90	20,000	10	0	50	40,000	50	.70
50	10,000	90	20,000	10	0	60	40,000	40	.85
51	10,000	90	20,000	10	0	70	40,000	30	.94
52	10,000	90	40,000	10	0	20	20,000	80	.78
53	2,500	50	10,000	50	0	60	40,000	40	.76
54	2,500	90	10,000	10	0	30	5,000	70	.78
55	2,500	80	20,000	20	0	40	10,000	60	.80
56	2,500	90	20,000	10	0	40	10,000	60	.81
57	2,500	90	40,000	10	0	10	20,000	90	.29
58	2,500	90	40,000	10	0	20	20,000	80	.48
59	2,500	90	40,000	10	0	30	20,000	70	.59
60	2,500	90	40,000	10	0	40	20,000	60	.77
61	2,500	90	5,000	10	0	70	40,000	30	.69
62	5,000	90	20,000	10	0	20	10,000	80	.74
63	5,000	80	40,000	20	0	40	20,000	60	.91
64	5,000	90	40,000	10	0	10	20,000	90	.39
65	5,000	90	40,000	10	0	20	20,000	80	.54
66	5,000	90	40,000	10	0	30	20,000	70	.68

Stimuli Question Set (Cont'd)

Question	G1A1	G1P1	G1A2	G1P2	G2A1	G2P1	G2A2	G2P2	G1%
67	0	10	10,000	90	0	30	20,000	70	.58
68	0	10	10,000	90	0	40	20,000	60	.82
69	0	10	10,000	90	0	40	40,000	60	.60
70	0	10	10,000	90	0	50	40,000	50	.70
71	0	10	10,000	90	0	60	40,000	40	.79
72	0	10	10,000	90	0	70	40,000	30	.91
73	0	20	10,000	80	0	50	20,000	50	.79
74	0	20	10,000	80	0	70	40,000	30	.88
75	0	90	10,000	10	0	10	5,000	90	.02
76	0	10	2,500	90	0	40	5,000	60	.75
77	0	30	2,500	70	0	60	5,000	40	.76
78	0	40	2,500	60	0	60	5,000	40	.63
79	0	40	2,500	60	0	70	5,000	30	.84
80	0	50	2,500	50	0	80	10,000	20	.69
81	0	50	2,500	50	0	70	5,000	30	.64
82	0	70	2,500	30	0	90	10,000	10	.57
83	0	70	2,500	30	0	80	5,000	20	.32
84	0	10	20,000	90	0	30	40,000	70	.63
85	0	10	20,000	90	0	40	40,000	60	.77
86	0	20	20,000	80	0	40	40,000	60	.65
87	0	20	20,000	80	0	50	40,000	50	.76
88	0	30	20,000	70	0	50	40,000	50	.66
89	0	10	5,000	90	0	40	10,000	60	.79
90	0	70	5,000	30	0	30	2,500	70	.09
91	0	80	5,000	20	0	10	2,500	90	.04
92	0	80	5,000	20	0	40	2,500	60	.09
93	0	80	5,000	20	0	50	2,500	50	.15
94	0	80	5,000	20	0	60	2,500	40	.19
95	0	90	5,000	10	0	10	2,500	90	.03
96	0	90	5,000	10	0	20	2,500	80	.07

References

Abdellaoui, Mohammed (2000). Parameter free elicitation of utilities and probability weighting functions. *Management Science, 46,* 1497-1512.

Abdellaoui, Mohammed, Frank Vossmann, and Martin Weber (2003). *Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty*. Manuscript obtained from authors.

Akaike, Hirotugu (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

Allais, Maurice (1953). Le comportement de l'homme rationel devant le risque, critique des postulates et axiomes de l'ecole americaine. *Econometrica, 21,* 503-546. [In French].

Ballinger, T. Parker and Nathaniel T. Wilcox (1997). Decisions, error and heterogeneity. *Economic Journal, 107,* 1090-1105.

Barberis, Nicholas and Ming Huang (2005). *Stocks as lotteries: The implications of probability weighting for securities prices*. Manuscript obtained from authors.

Becker, Gordon M., Maurice H. De Groot, and Jacob Marschak (1963a). An experimental study of some stochastic models for wagers. *Behavioural Science, 3,* 199-202.

Becker, Gordon M., Maurice H. De Groot, and Jacob Marschak (1963b). Stochastic models of choice behaviour. *Behavioural Science, 8,* 41-55.

Bell, David E. (1995a). Risk, return and utility. *Management Science, 41(1),* 23-30.

Bell, David E. (1995b). A contextual uncertainty condition for behavior under risk. *Management Science, 41(7),* 1145-1150.

Bell, David E. and Peter C. Fishburn (2000). Utility functions for wealth. *Journal of Risk and*

Uncertainty, 20, 5-44.

Bell, David E. and Peter C. Fishburn (2001). Strong one-switch utility. *Management Science, 47(4),* 601-604.

Birnbaum, Michael H. and Alfredo Chavez (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71,* 161-194.

Birnbaum, Michael H., Jamie N. Patton, and Melissa K. Lott (1999). Evidence against Rank Dependent utility theories: Tests of Cumulative independence, Interval dependence, Stochastic dominance and Transitivity. *Organizational Behavior and Human Decision Processes, 77,* 44-83.

Bleichrodt, Hans and Jose Luis Pinto (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science, 46(11),* 1485-1496.

Blondel, Serge (2002). Testing theories of choice under risk: Estimation of individual functionals. *Journal of Risk and Uncertainty, 24(3),* 251-265.

Brandstätter, Eduard, Anton Kühberger and Friedrich Schneider (2002). A cognitive-emotional account of the shape of the probability weighting function. *Journal of Behavioral Decision Making, 15*, 79-100

Burnham, Kenneth P. and David R. Anderson (2002). *Model selection and multimodel inference*. New York: Springer-Verlag.

Buschena, David and David Zilberman (2000). Generalized expected utility, heteroscedastic error, and path dependence in risky choice. *Journal of Risk and Uncertainty, 20,* 67-88.

Camerer, Colin F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty, 2,* 61-104.

Camerer, Colin F. and Robin M. Hogarth (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19,* 7-42.

Camerer, Colin F. and Teck-Hua Ho (1994). Violations of the betweeness axiom and nonlineraity in probability. *Journal of Risk and Uncertainty, 8,* 167-196.

Carbone, Enrica (1997). Investigation of stochastic preference theory using experimental data. *Economic Letters, 57(3),* 305-311.

Carbone, Enrica and John D. Hey (1994). Discriminating between preference functionals: A preliminary monte carlo study. *Journal of Risk and Uncertainty, 8,* 223-242.

Carbone, Enrica and John D. Hey (1995). A comparison of the estimates of EU and non-EU preference functionals using data from pairwise choice and complete ranking experiments. *Geneva Papers on Risk and Insurance Theory, 20,* 111-133.

Carbone, Enrica and John D. Hey (2000). Which error story is best. *Journal of Risk and Uncertainty, 20,* 161-176.

Chechile, Richard A. and Alan D. J. Cooke (1997). An experimental test of a general class of utility models: Evidence for context dependency. *Journal of Risk and Uncertainty, 14,* 75-93.

Debreu, Gerard (1958). Stochastic choice and cardinal utility. *Econometrica, 26(3),* 440-444.

Doctor, Jason, N. et al. (2004). A new and more robust test of QALYs. *Journal of Health Economics, 23,* 353-367.

Fennema, Hein and Marcel van Assen (1998). Measuring the utility of losses by means of the trade-off method. *Journal of Risk and Uncertainty, 17,* 277-295.

Fennema, Hein and Peter Wakker (1997). Original and cumulative prospect theory: A discussion of empirical differences. *Journal of Behavioral Decision Making, 10,* 53-64.

Fishburn, Peter C. and Gary A. Kochenberger (1979). Two piece Von Neumann-Morgernstern utility functions. *Decision Sciences, 10(4),* 503-518.

Goldstein, William M. and Hillel J. Einhorn (1987). Expression theory and the preference reversal phenomena. *Psychological Review, 94,* 236-254.

Gonzalez, Richard and George Wu (1999). On the shape of probability weighting function. *Cognitive Psychology, 38,* 129-166.

Guthrie, Chris (2003). Prospect theory, risk preference, and the law. *Northwestern University Law Review, 97(3),* 1115-1163.

Grünwald, Peter (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology, 44,* 133-152.

Harless, David W. and Colin F. Camerer (1994). The predictive utility of generalized expected utility theories. *Econometrica, 62(6),* 1251-1290.

Herrnstein, Richard J. (1997). *The matching law: Papers in psychology and economics.* H. Rachlin & D. I. Laibson (Eds.). New York: Russell Sage Foundation

Hertwig, Raph and Andreas Ortmann (2001). Experimental practices in economics: A methodological challenge to psychologists? *Behavioral and Brain Sciences, 24,* 383-451.

Hey, John D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review, 39,* 633-640.

Hey, John D. and Chris Orme (1994). Investigating generalizations of the expected utility theory using experimental data. *Econometrica, 62(6),* 1291-1326.

Hoeting, Jennifer A. et al. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14,* 382-401.

Ingersoll, Jonathan E., Jr . (1987). *Theory of financial decision making*. Savage, MD: Rowman & Littlefield.

Isaac, R. Mark and Duncan James (2000). Just who are you calling risk averse? *Journal of Risk* *and Uncertainty, 20(2),* 177-187.

Kachelmeier, Steven J. and Mohamed Shehata (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review, 82(5),* 1120-1141.

Karmarkar, Uday (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance, 21,* 61-72.

Karmarkar, Uday (1979). Subjectively weighted utility and Allais paradox. *Organizational Behavior and Human Performance, 24,* 67-72.

Lattimore, Pamela K., Joanna R. Baker, and Ann D. Witte (1992). The influence of probability on risky choice. *Journal of Economic Behaviour and Organization, 17,* 377-400.

Lichtenstein, Sarah and Paul Slovic (1971). Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89,* 46-55.

Loomes, Graham and Robert Sugden (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39,* 641-648.

Loomes, Graham and Robert Sugden (1998). Testing different stochastic specifications of risky choice. *Economica, 65,* 581-598.

Loomes, Graham, Peter Moffat, and Robert Sugden (2002). A microeconometric test of alternative stochastic theories of risk choice. *Journal of Risk and Uncertainty, 24,* 103-130.

Luce, R. Duncan (1959). *Individual choice behavior*. New York: John Wiley & Sons.

Luce, R. Duncan (1991). Rank and sign dependent linear utility models for binary gambles. *Journal of Economic Theory, 53,* 75-100.

Luce, R. Duncan (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Erlbaum.

Luce, R. Duncan (2001). Reduction invariance and Prelec's weighting functions. *Journal of*

Mathematical Psychology, 45(1), 167-179.

Luce, R. Duncan and Peter C. Fishburn (1991). Rank and sign dependent linear utility models of finite first-order gambles. *Journal of Risk and Uncertainty, 4,* 29-59.

Luce, R. Duncan and Peter C. Fishburn (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty, 11,* 5-16.

Luce, R. Duncan, Barbara A. Mellers and Shi-jie Chang (1993). Is choice the correct primitive? On using certainty equivalents and reference levels to predict choices amoung gambles. *Journal of Risk and Uncertainty, 6,* 115-143.

Mellers, Barbara A. et al. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 347-361.

Mosteller, Frederick and Philip Nogee (1951). An experimental measurement of utility. *Journal of Political Economy, 59,* 371-404.

Myung, In Jae (2000). Importance of complexity in model selection. *Journal of Mathematical Psychology, 44(1),* 190-204.

Myung, In Jae and Mark A. Pitt (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review, 4(1),* 79-95.

Myung, In Jae, Cheongtag Kim, and Mark A. Pitt (2000). Towards an explanation of the power law artifact: Insights from response surface analysis. *Memory and Cognition, 28(5)*, 832-840.

Prelec, Drazen (1998). The probability weighting function. *Econometrica, 66(3),* 497-528.

Quiggin, John (1982). A theory of anticipated utility. *Journal of Economic Behaviour and Organization, 3,* 323-343.

Rich, Elaine and Kevin Knight (1991). *Artificial Intelligence*. New York, NY: McGraw-Hill.

Segal, Uzi (1989). Anticipated utility: A measure representation approach. *Annals of*

Operations Research, 19, 359-373.

Slovic, Paul (1972). Psychological study of human judgement: Implications for investment decision making. *Journal of Finance, 27,* 779-799.

Smith, Vernon L. and James M. Walker (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry, 31(2),* 245-261.

Sneddon, Robert and R. Duncan Luce (2001). Empirical comparisons of bilinear and nonbilinear utility theories. *Organizational Behavior and Human Decision Processes, 84(1),* 71-94.

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

Starmer, Chris and Robert Sugden (1989a). Violations of the independence axiom in common ratio problems: An experimental test. *Annals of Operations Research, 19,* 79-102.

Starmer, Chris and Robert Sugden (1989b). Probability and juxtaposition effects: An experimental investigation of the common ratio effect. *Journal of Risk and Uncertainty, 2,* 159-178.

Stevens, S. Smitty (1957). On the psychophysical law. *Psychological Review, 64,* 153-181.

Stewart, Neil et al. (2002). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General, 132(1),* 23-46.

Tversky, Amos and Craig R. Fox (1995). Weighing risk and uncertainty. *Psychological Review, 102(2),* 269-283.

Tversky, Amos and Daniel Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Wakker, Peter and Amos Tversky (1993). An axiomatization of cumulative prospect theory.

Journal of Risk and Uncertainty, 7, 147-176.

Wakker, Peter, Ido Erev, and Elke Weber (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty, 9,* 195-230.

Weber, Elke U. and Britt Kirsner (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty, 14,* 41-61.

Wu, George and Richard Gonzalez (1996). Curvature of the probability weighting function. *Management Science, 42(12),* 1676-1690.

Yaari, Menahem E. (1987). The dual theory of choice under risk. *Econometrica, 55(1),* 95-115.

Tables

Table 1.
Choice Reversal Rates for Repeated Questions

Study	Ss	Qs	Reversal Rate[a]	Comments
Camerer (1989)	348	1	.32	
Starmer and Sugden (1989b)	283	1	.26	Stable if payment incentive reduced (.27).
Hey and Orme (1994)	80	100	≈ .25	7-10 day separation of repeat questions. Rate varies from .00 to .50 across participants.
Wakker, Erev and Weber (1994)	84	24	.33	Three outcome prospects rather than typical binary ones. Participant variation reported with 90% of them having rates over .25.
Carbone and Hey (1995)	40	42	.12	Rate varies from .00 to .40 across questions.
Ballinger and Wilcox (1997)	120	25	.21	Rate varies from .11 to .30 across questions.
Loomes and Sugden (1998)	92	40	.18	Only includes non-dominant pairs. Rate varies from .00 to .35 across participants.
Abdellaoui (2000)	40	24	.19	

[a] In each study, participants (number in Ss column) were asked to choose between the same two prospects on two different occasions (number of repeated questions in Qs column). The column reports how often an average participant changed their preference on the second presentation.

Table 2.
Summary of Functional Forms for Value Function

Name	Abbreviation	Equation[a]
Linear	Lin	$v(x) = x$
Logarithmic	Log	$v(x) = \ln(a + x)$
Power	Pwr	$v(x) = x^a$
Quadratic	Quad	$v(x) = ax - x^2$
Exponential	Expo	$v(x) = 1 - e^{-ax}$
Bell	Bell	$v(x) = bx - e^{-ax}$
HARA	Hara	$v(x) = -(b + x)^a$

[a] Each function is oriented so that it is increasing for $x > 0$. Note that some parameters require constraining (e.g. if b is less than zero for HARA then there are problems with £0).

Table 3.
Summary of Functional Forms for Risky Weighting Function

Name[a]	Abbreviation	Equation[b]
Linear	Lin	$\pi(p) = p$
Power	Pwr	$\pi(p) = p^r$
Goldstein-Einhorn	GE	$\pi(p) = \dfrac{sp^r}{sp^r + (1-p)^r}$
Tversky-Kahneman	TK	$\pi(p) = \dfrac{p^r}{\left(p^r + (1-p)^r\right)^{(1/r)}}$
Wu-Gonzalez	WG	$\pi(p) = \dfrac{p^r}{\left(p^r + (1-p)^r\right)^s}$
PrelecI	Prl-I	$\pi(p) = e^{-(-\ln p)^r}$
PrelecII	Prl-II	$\pi(p) = e^{-s(-\ln p)^r}$

[a] Most of the labels relate to the authors that seem to have first reported them

[b] Some of the free parameters need to be constrained. For example, r and s in Wu-Gonzalez cannot both be greater than 1.0 or the constraint $\pi(p) \leq 1$ is violated.

Table 4.
Summary of Functional Forms for Choice Function

Name	Equation
Constant Error	$P\left(V(g_1),V(g_2)\right) = \begin{cases} V(g_1) < V(g_2) \text{ then } \varepsilon \\ V(g_1) = V(g_2) \text{ then } \frac{1}{2} \\ V(g_1) > V(g_2) \text{ then } (1\text{-}\varepsilon) \end{cases}$
Probit	$P\left(V(g_1),V(g_2)\right) = \Phi\left[V(g_1) - V(g_2), 0, \varepsilon\right]$ where $\Phi[x,\mu,\varepsilon]$ is the cumulative normal distribution with mean and SD μ and ε, at point x
Logit	$P\left(V(g_1),V(g_2)\right) = \dfrac{1}{1 + e^{-\varepsilon\left(V(g_1) - V(g_2)\right)}}$
Luce	$P\left(V(g_1),V(g_2)\right) = \dfrac{V(g_1)^{\varepsilon}}{V(g_1)^{\varepsilon} + V(g_2)^{\varepsilon}}$

Table 5.
Parameter Estimates over Gains for CPT

Function	Name[b]	Parameter Values		Study[a]
Value Function	Pwr	$a = 0.88$		Tversky and Kahneman (1992)
		$a = 0.225$		Camerer and Ho (1994)
		$a = 0.50$ (w/ TK) $a = 0.48$ (w/ Prl-I)		Wu and Gonzalez (1996)
		$a = 0.82$		Birnbaum and Chavez (1997)
		$a = 0.389$ (outwards) $a = 0.210$ (inwards) $a = 0.364$ (cert. Equiv.)		Fennema and Van Assen (1999)
		$a = 0.49$		Gonzalez and Wu (1999)
		$a = 0.89$		Abdellaoui (2000)
		$a = 0.19$		Current Paper
Risky Weighting Function	GE	$r = 0.77$	$S = 0.69$	Tversky and Fox (1995)
		$r = 0.68$	$S = 0.84$	Wu and Gonzalez (1996)
		$r = 1.59$	$S = 0.31$	Birnbaum and Chavez (1997)
		$r = 0.44$	$S = 0.77$	Gonzalez and Wu (1999)
		$r = 0.962$	$S = 0.207$	Birnbaum, Patton and Lott (1999)
		$r = 0.60$	$S = 0.65$	Abdellaoui (2000)
		$r = 0.550$	$S = 0.816$	Bleichrodt and Pinto (2000)
		$r = 0.71$	$S = 0.88$	Brandstätter, Kühberger, and Schneider (2002)
		$r = 0.96$	$S = 1.40$	Current Paper
	TK	$r = 0.61$		Tversky and Kahneman (1992)
		$r = 0.56$		Camerer and Ho (1994)
		$r = 0.71$ $r = 0.50$ (w/ Exp)		Wu and Gonzalez (1996)
		$r = 0.60$		Abdellaoui (2000)
		$r = 0.674$		Bleichrodt and Pinto (2000)
		$r = 0.96$		Current Paper
	WG	$r = 0.721$	$S = 1.565$	Wu and Gonzalez (1996)
		$r = 0.75$	$S = 1.4$	Brandstätter, Kühberger, and Schneider (2002)
		$r = 0.93$	$S = 0.89$	Current Paper
	Prl-I	$r = 0.74$		Wu and Gonzalez (1996)
		$r = 0.533$		Bleichrodt and Pinto (2000)
		$r = 0.94$		Current Paper
	Prl-II	$r = 0.534$	$s = 1.083$	Bleichrodt and Pinto (2000)
		$r = 1.00$	$s = 1.00$	Current paper

[a] Studies were included if they reported fit parameter values for one of the functions listed in Tables 2 or 3 within a CPT framework. Only results over risk (rather than uncertainty) are included. Where a risky weighting function was required to fit the value function, they are either noted or can be inferred. Where necessary, the risky weighting functions have been fit using a Power value function, except in one noted case.

[b] The key to the labels and parameter notation are provided in Tables 2 and 3.

Table 6.
Proportion of Times Bigger Model is Significantly Better

	Nesting Function	Nested Function	Proportion ($p < .05$)
Value Functions	Non-Para	Quad	.80
		Expo	.46
		Bell	.32
		Pwr	.31
		Log	.29
		Hara	.26
	Hara	Quad	.81
		Expo	.43
		Pwr	.18
		Log	.16
	Bell	Expo	.32
Risky Weighting Functions	Non-Para	Pwr	.50
		TK	.43
		Prl-I	.42
		WG	.40
		GE	.36
		Prl-II	.32
	Prl-II	Pwr	.43
		Prl-I	.32
	WG	TK	.18

Table 7.
Average Within Participant Rank of Different Functional Forms

		Average Rank[a]		Average AIC
		Full	Subset	
Value Function	Pwr	96.2	63.1	74.5
	Log	97.8	68.1	74.5
	Hara	103.6	77.2	74.4
	Bell	109.6	69.1	76.4
	Non Para	115.2	93.0	75.0
	Expo	123.8	90.2	78.7
	Quad	188.7		96.6
	Lin	191.3		99.4
Risky Weighting Function	Prl-II	106.2	75.5	75.5
	GE	114.4	79.3	77.6
	TK	120.4	74.8	81.7
	Prl-I	122.5	66.0	85.0
	WG	123.5	88.3	78.9
	Pwr	132.3		80.0
	Lin	148.4		89.0
	Non Para	158.4		81.8
Stochastic Process	Luce	107.4	74.0	77.7
	Logit	116.9	76.0	79.1
	Probit	119.8	80.3	79.4
	Cont. Err.	169.1		88.6

[a] All 256 combinations of the functions were fitted to each participant. Within participant ranks (1= best) were assigned based on the Akaike Information Criterion of each fit. The Subset column is based on a similar analysis, but for a restricted set of functions.

Table 8.
Results of Three Factor Within Participant ANOVA on Model Rank

n = 96	Full		Subset[a]	
	F(1,95)	Sig	F(1,95)	Sig
Value	204.92	0.0000	15.43	0.0002
Risk	49.40	0.0000	17.02	0.0001
Stoch	106.30	0.0000	1.34	ns
Value x Risk	70.71	0.0000	2.19	0.1420
Value x Stoch	78.75	0.0000	12.79	0.0006
Risk x Stoch	16.82	0.0001	5.16	0.0254
Value x Risk x Stoch	18.65	0.0000	3.91	0.0508

[a] The Subset case has those functions removed that have no entries in the Table 7's Subset column.

Table 9.
Models With Highest Average Within Participant Rankings

Model Versions			Average Rank	t-stat[a]
Value Fn	Risk Weight	Stochastic Process		
Pwr	Prl-I	Logit	48.9	
Log	Prl-I	Luce	53.3	0.808
Pwr	Prl-II	Logit	53.8	1.150
Pwr	Prl-I	Probit	54.0	*2.196
Bell	Prl-I	Logit	54.1	1.338
Pwr	Prl-I	Luce	55.1	1.097
Bell	Prl-I	Luce	55.4	1.037
Log	Prl-I	Logit	56.7	*2.197
Pwr	GE	Logit	57.6	*2.043
Hara	Prl-I	Luce	58.5	1.649

[a] The significance of the difference between a model's average rank and that of the top model were assessed using a pairwise t-test (* indicates models that are worse with .95 confidence). All models lying outside this top ten were worse than Pwr/Prl-I/Logit by this t-test with .95 confidence.

Figure Legends

Figure 1: Stochastic Choice. The figure is reproduced from Table 8 of "An experimental measurement of utility" by Mosteller and Nogee, 1951, *Journal of Political Economy, 59*, p. 384. The figure shows how often participant B-I accepted the offered gamble for various outcome levels. A Logit curve has also been fitted through the original data.

Figure 2: Goodness of Fit Distribution of Models. The distribution of maximized log likelihoods found during the model estimation process is shown. The average log likelihood is –35.8, equivalent to assigning a geometric mean probability of .67 to participant's actual choices.

Figure 3: Power Function Parameter Estimate Distribution. The plot shows the distribution across participants of the value function exponent (estimated in combination with a GE risky weighting function and Logit transformation). The chart plots participant frequency for categories that are 0.10 wide. Eighty seven of the 96 participants fall within the range plotted. Ticks mark the prior findings reviewed in Table 5.

Figures 4 and 5: Risky Weighting Function Parameter Estimate Distributions. The first figure shows the distribution across participants of the GE risky weighting function elevation parameter, *s*. The second plot shows the distribution of the curvature parameter, *r*. These were obtained using a Pwr value function and Logit transformation. In Figure 4, the chart plots participant frequency for categories that are 0.40 wide, and 91 of the 96 participants fall within the range plotted. In Figure 5, the categories are 0.20 wide, and 85 participants are plotted. Ticks mark the prior findings reviewed in Table 5.

Figure 6: The Explanatory Performance of Different Function Combinations. The figure shows the rankings achieved by each combination of value and risky weighting function, averaged across participants and four stochastic processes.

Figures

Figure 1: Stochastic Choice

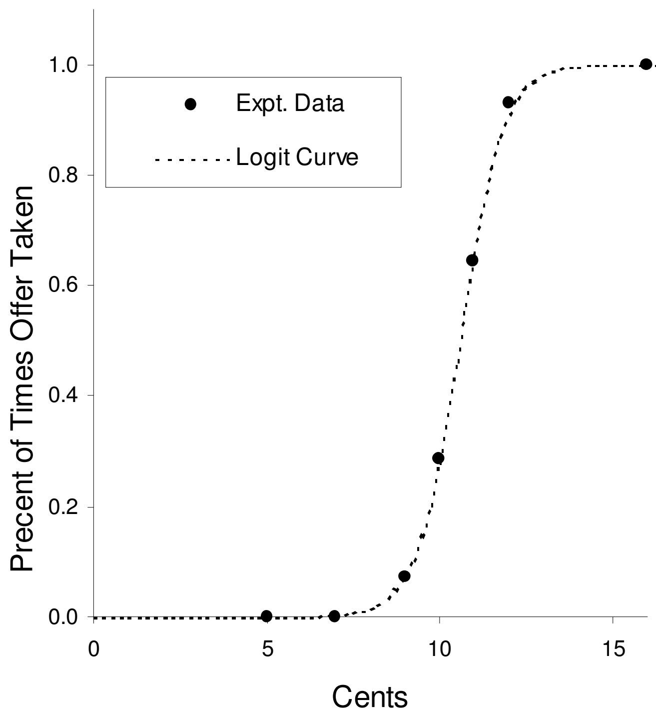


Figure 2: Goodness of Fit Distribution of Models

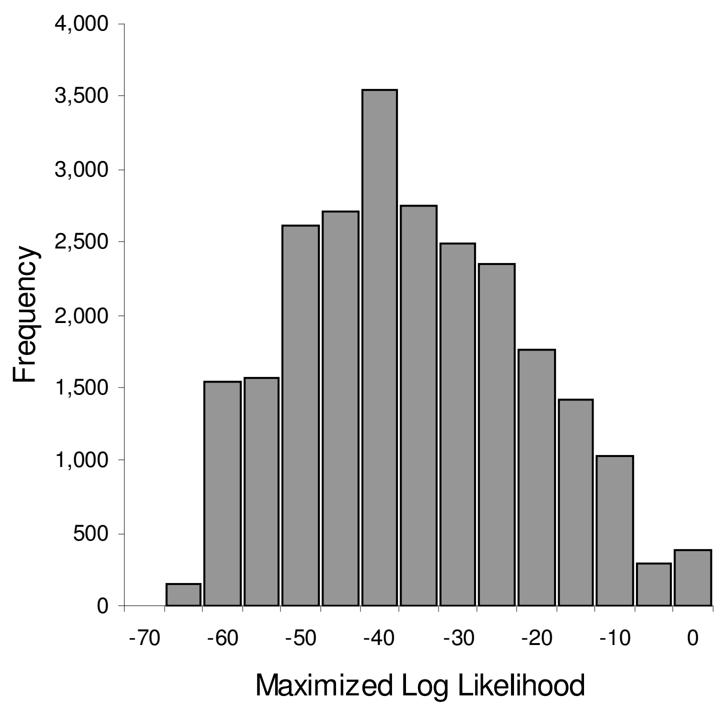
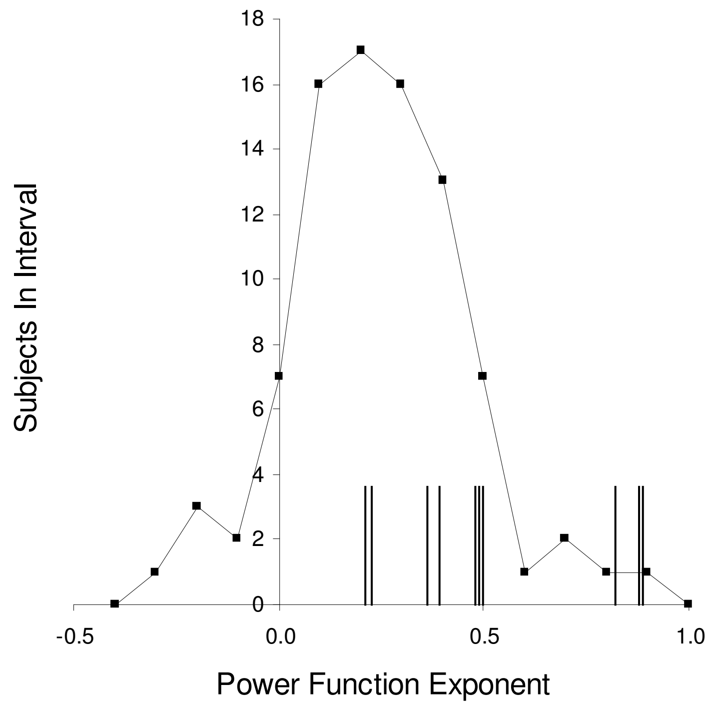


Figure 3: Power Function Parameter Estimate Distribution



Figures 4 and 5: Risky Weighting Function Parameter Estimate Distributions

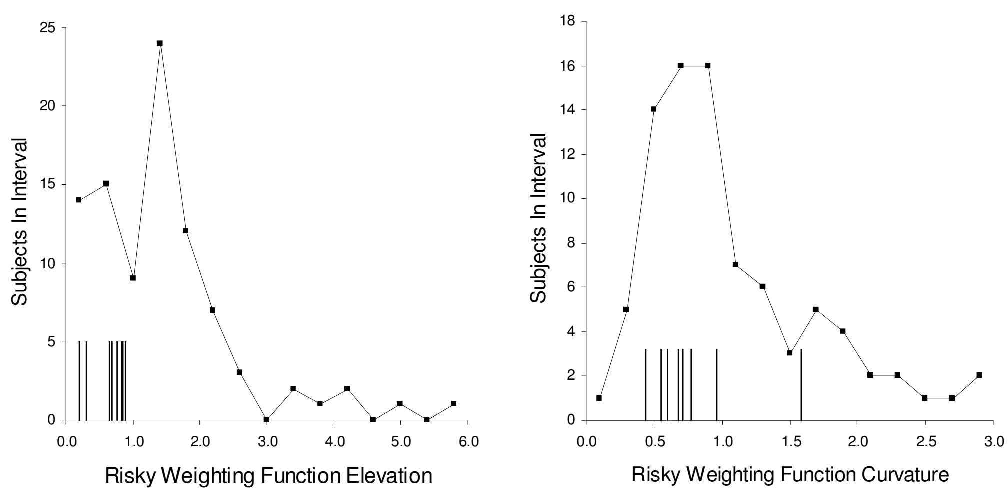


Figure 6: The Explanatory Performance of Different Function Combinations