MANAGEMENT SCIENCE
Vol. 51, No. 9, September 2005, pp. 1346–1358
ISSN 0025-1909 | EISSN 1526-5501 | 05 | 5109 | 1346

DOI 10.1287/mnsc.1050.0404

Three New Tests of Independence That Differentiate Models of Risky Decision Making

Michael H. Birnbaum
Department of Psychology, California State University, Fullerton, California 92834, mbirnbaum@fullerton.edu

This paper tests new "independence" properties to compare three models of risky decision making. According to the rank-affected multiplicative (RAM) weights model, all three properties should be satisfied; according to the transfer of attention exchange (TAX) model, two should be satisfied and one can be violated. However, according to cumulative prospect theory (CPT), all three properties will be violated if the probability weighting function is nonlinear. Although CPT is flexible enough to accommodate violations of these properties, its predicted violations based on previously estimated parameters failed to materialize. In 14 choices for which the CPT model disagreed with TAX, CPT correctly predicted the modal choice in only one case and TAX predicted the modal choice in the other 13. New versions of three other paradoxes were also tested and found to refute CPT.

Key words: choice; cumulative prospect theory; decision making; distribution independence; expected utility; paradoxes; prospect theory; rank-dependent utility
History: Accepted by Detlof von Winterfeldt, decision analysis; received April 1, 2004. This paper was with the author $2\frac{1}{2}$ months for 2 revisions.

1. Introduction

This study introduces three new "independence" properties and examines if they are descriptive of risky decision making. The three properties are all implied by expected utility (EU) theory, and all could be deduced from the "sure thing" principle of Savage (1954). Non-EU models may satisfy or violate these particular properties, however, allowing them to be used to compare rival descriptive models. These properties distinguish cumulative prospect theory (CPT) (Tversky and Kahneman 1992), rank-affected multiplicative (RAM) weights, and the transfer of attention exchange (TAX) models (Birnbaum and Chavez 1997, Birnbaum and Navarrete 1998).

It will be shown that RAM satisfies all three properties, that special TAX satisfies two properties, and that CPT can violate all three properties. Section 2 describes the models, and §3 derives their predictions. Section 4 describes two experiments designed to compare these models, and §5 presents results showing that data are most consistent with the TAX model. Section 6 discusses how these findings fit in with other evidence bearing on the accuracy of these three models.

2. Three Models of Risky Decision Making

2.1. Cumulative Prospect Theory

For gambles with positive consequences, $G = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$, where $x_1 > x_2 > \cdots > x_n > 0$, CPT can be written as follows:

$$CPU(G) = \sum_{i=1}^{n}\left[W\left(\sum_{j=1}^{i} p_j\right) - W\left(\sum_{j=1}^{i-1} p_j\right)\right]u(x_i), \qquad (1)$$

where $P_i = \sum_{j=1}^{i} p_j$ is the decumulative probability that a prize is equal to x_i or greater, and $W(P)$ is a monotonic function that assigns decumulative weight to decumulative probability; $W(0) = 0$ and $W(1) = 1$. With strictly positive consequences, this representation coincides with rank-dependent utility (RDU), as described in Quiggin (1993).

In practice, the decumulative weighting function (Tversky and Kahneman 1992) has been fit with the following equation:

$$W(P) = \frac{P^{\gamma}}{[P^{\gamma} + (1-P)^{\gamma}]^{1/\gamma}}, \qquad (2)$$

where P is decumulative probability and γ is a constant. With $\gamma < 1$, $W(P)$ is an inverse-S function of P, which predicts that participants are risk seeking for small values of P and risk averse for moderate to large values of P. The term "CPT model" will be used in this paper to refer to Equations (1) and (2) with parameters estimated by Tversky and Kahneman (1992): $\gamma = 0.61$ and $u(x) = x^{\beta}$, $\beta = 0.88$. Two parameter equations have also been suggested for $W(P)$, for example, by Tversky and Wakker (1995), but these make virtually identical predictions in this paper.

2.2. Rank Affected Multiplicative Model

The "configural weight" models were originally developed to describe how people combine evidence when making evaluative judgments (Birnbaum 1974). As Birnbaum (1974, p. 559) states, in these models, "the weight of an item depends in part on its rank within the set." These models thus have some similarities to the rank-dependent utility models that were later developed by Quiggin (1993) and extended by others (Luce 2000; Luce and Fishburn 1991, 1995; Tversky and Kahneman 1992). In particular, by selection of proper functions and parameters, the RDU models can be made to mimic predictions of the configural weight, RAM, and TAX models in certain experiments. However, these classes of models make different predictions for other experiments, some of which will be tested here.

In the RAM and TAX models, risky gambles are represented as trees in which each *branch* is a probability-consequence pair that is separately specified in the description given the decision maker. For example, the gamble $G = (\$100, 0.1;\ \$100, 0.1;\ \$0, 0.8)$ represents a three-branch gamble whose three branches are $(\$100, 0.1)$, $(\$100, 0.1)$, and $(\$0, 0.8)$. When gambles are treated as probability distributions rather than as trees, however, gamble G is equivalent to the following two-branch gamble, $F = (\$100, 0.2;\ \$0, 0.8)$, where the two branches leading to \$100 have been combined or *coalesced*. However, in RAM and TAX, gambles G and F are not the same. On the other hand, in any RDU, rank and sign-dependent utility (RSDU), or CPT model, G and F are equivalent.

In the RAM model, the importance of a branch depends not only on that branch's probability, but also on the relation of its consequence relative to consequences on other branches of the same gamble. In the RAM model, importance is represented by weight that is a product of a function of the branch's probability and a function of the rank of the consequence on that branch relative to the other branches. The key idea is that each branch carries a degree of importance that depends on its probability and rank of its consequence. In the case of risky decisions, the branch leading to the worst consequence will typically receive the most weight, and the branch leading to the best consequence will typically receive the least weight, if they had the same probability. These properties produce risk aversion in the RAM model.

For strictly positive consequences, the RAM model can be written as follows:

$$\mathrm{RAM}(G) = \frac{\sum_{i=1}^{n} a(i,n)s(p_i)u(x_i)}{\sum_{i=1}^{n} a(i,n)s(p_i)}, \tag{3}$$

where $\mathrm{RAM}(G)$ is the utility of the n-branch gamble $G = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$, where $x_1 > x_2 > \cdots > x_n > 0$, $a(i,n)$ is the weight of the branch whose consequence is ranked i among n branches, $s(p_i)$ is a function of branch probability, p_i, and $u(x_i)$ is the utility of the consequence on that branch. The RAM model takes its name from the multiplicative relationship between a function of branch probability and a function of branch's rank; i.e., the relative weight of branch i is $a(i,n)s(p_i)$, divided by the sum of these products.

The "RAM model" in this paper will refer to Equation (3) with $u(x) = x$ (for $0 < x < \$150$), with the rank weights equal to their ranks; i.e., $a(i,n) = i$ and $s(p) = p^{\gamma}$, where $\gamma = 0.6$. These parameters provide an approximate fit to data in Tversky and Kahneman (1992).

For two-branch gambles of the form $G = (x, p;\ 0, 1-p)$, this RAM model, like CPT, implies that certainty equivalents of gambles (the amounts of cash that are indifferent to the gambles) will be an inverse-S function of probability to win: $CE(G) = xp^{\gamma}/[p^{\gamma} + 2(1-p)^{\gamma}]$. If the effect of probability is negatively accelerated ($\gamma < 1$), this model, like that of CPT, implies risk seeking for small values of p, and risk aversion for moderate to large values of p.

It is important to keep in mind that in RAM, the rank weights represent ranks of consequences on discrete branches, not decumulative probabilities, as would be the case in RDU or CPT. For example, in a three-branch gamble, there are exactly three branch ranks, 1, 2, and 3, for the branches with the highest-, middle-, and lowest-valued consequences, respectively, regardless of their probabilities. In the parameterized RAM model, the branch with the lowest consequence has three times the weight of the highest consequence, and the branch with the second highest consequence has twice the weight of the highest consequence, assuming their probabilities were equal.

2.3. Transfer of Attention Exchange Model

The TAX model also assumes that the weight of a branch depends on the branch's probability. The key idea is that branches compete for attention, which redistributes their weights. The shifts of attention from branch to branch are represented by transfers of weight from one branch to another. For most people choosing between risky gambles, weight will be drawn from branches leading to higher-valued consequences and transferred to branches with lower-valued consequences. It is this shift of weight that accounts for risk aversion in this model, rather than nonlinear utility (as in EU theory). In the "special" TAX model, the amount of weight transferred is assumed to be a fixed proportion of the weight that the branch has to lose, which aside from the transfers, is a function of its probability, $t(p)$.

When lower-valued branches receive greater weight, this special TAX model can be written as follows for $G = (x_1, p_1; x_2, p_2; \dots; x_n, p_n)$, where $x_1 > x_2 > \cdots > x_n > 0$:

$$\mathrm{TAX}(G) = \frac{\sum_{i=1}^{n}\left[t(p_i) + (\delta/(n+1))\sum_{j=1}^{i-1} t(p_j) - (\delta/(n+1))\sum_{j=i+1}^{n} t(p_i)\right]u(x_i)}{\sum_{i=1}^{n} t(p_i)}, \quad (4)$$

where $\delta > 0$, which produces risk aversion in the case of fifty-fifty gambles, and $t(p) = p^{\gamma}$, where $\gamma < 1$. To understand this formula, it helps to break down the three terms affecting weight. First, if the configural weight parameter, δ, were 0, then the formula simplifies, and relative weights are a function of branch probabilities, $t(p)$, divided by the sum of these weights. If $t(p)$ is a negatively accelerated function, then certainty equivalents in two-branch gambles will be an inverse-S function of p. The second term represents the transfer of weight from each branch with a higher consequence ($j = 1, 2, \dots, i-1$) to branch i. The third term shows that branch i, in turn, gives up a proportion of its weight to each branch with a lower consequence ($j = i+1, i+2, \dots, n$).

Although this model is equivalent to previous versions, branches are ordered here from best to worst in conformity with the ranking used in CPT, and $\delta > 0$ in this equation corresponds to $\delta < 0$ in Birnbaum and Chavez (1997) and other papers using that older convention.

In the parameterized TAX model of this paper, predictions are calculated from Equation (4) with $u(x) = x$ for small positive sums ($0 < x < \$150$), $t(p) = p^{0.7}$, and $\delta = 1$. With $\delta = 1$, in a two-branch gamble, one-third of the weight (transformed probability) of the higher-valued branch is transferred to the lower-valued branch. In a three-branch gamble, one-fourth of the weight of any higher-valued branch is transferred to each lower-valued branch. These parameters were also chosen to approximate data of Tversky and Kahneman (1992).

2.4. Comparing the Models

The purpose of this paper is to investigate three new properties that distinguish RAM and TAX models from each other and from CPT. These three new properties differ from several properties previously investigated in that CPT can violate them all, whereas RAM and special TAX models must satisfy all or two of these three properties, respectively. So, in this case, CPT has the opportunity to go on the offense in contrast to previous contests (e.g., Birnbaum and Navarrete 1998), where CPT had to defend the null hypothesis (that properties it implied would be satisfied) against models that violated those properties. Here, TAX and RAM are put in the position of defending null hypotheses against violations predicted by the CPT model. These properties are also implied by EU, so they also can be considered as tests between CPT and EU.

Consider the two choices in Example 1:

S:	0.60 probability to win \$2 0.20 probability to win \$56 0.20 probability to win \$58	*R*:	0.60 probability to win \$2 0.20 probability to win \$4 0.20 probability to win \$96
*S*2:	0.10 probability to win \$2 0.45 probability to win \$56 0.45 probability to win \$58	*R*2:	0.10 probability to win \$2 0.45 probability to win \$4 0.45 probability to win \$96

Given its previous parameters, CPT implies certainty equivalents of 20.4 and 23.4 for S and R, and 39.7 and 36.2 for $S2$ and $R2$, respectively. Thus, CPT predicts that people will choose R over S and $S2$ over $R2$. However, TAX model calculations yield certainty equivalents of 21.7 and 13.8 for S and R, and 36.9 and 22.8 for $S2$ and $R2$, respectively, so people should choose the "safe" gamble in both choices. For RAM, predictions are 19.3 and 12.3, and 42.5 and 26.2, respectively. Therefore, both RAM and TAX models imply that people should choose both "safe" gambles, but CPT predicts a reversal from R to $S2$ as the probability of the lower branch is decreased from 0.6 to 0.1. It will be shown that the RAM and TAX models must defend the null hypothesis that there should be no change in preference against CPT's prediction of a specific violation.

Note that S and R share a common lower branch of 0.60 to win \$2, whereas the common lower branch in $S2$ and $R2$ has a probability of 0.10 to win \$2. The property tested in Example 1 is called *3-lower distribution independence* (3-LDI) because only the probability of the lowest common consequence in a three-branch gamble is changed, and the ratio of probabilities of the other branches is fixed. Section 3 defines this property (as well as others to be tested) more precisely, and derives predictions of the models.

3. Properties and Predictions of the Models

3.1. Lower Distribution Independence (3-LDI)

Let $G = (x, p; y; q; z, 1-p-q)$ represent a three-branch gamble to win x with probability p, y with probability q, and z otherwise. Let $x > y > z > 0$. Define 3-LDI as follows:

$$S = (x, p; y; p; z, 1-2p) \succ R = (x', p; y'; p; z, 1-2p) \quad \text{if and only if}$$
$$S2 = (x, p'; y; p'; z, 1-2p') \succ R2 = (x', p'; y'; p'; z, 1-2p'). \quad (5)$$

Note that the first comparison involves a distribution with probabilities of p and p to win x and y, with a common branch of $(z, 1-2p)$. The second choice is the same except for a different probability, p' on the two branches and a different probability on the common branch $(z, 1-2p')$.

It will be shown that the special cases of both RAM and TAX satisfy this property, whereas CPT can violate it. To understand the predictions, recall that all three models represent the utility of a gamble as a weighted average of the utilities of the consequences. When the number of branches, their ranks, and the probability distribution are all fixed, all three models reduce to what Birnbaum and McIntosh (1996, p. 92) call the "generic rank-dependent configural weight" theory, also known as the rank weighted utility model (Luce 2000, Marley and Luce 2001). This generic model can be written for the choice between S and R as follows:

$$S \succ R \;\Leftrightarrow\; w_1u(x) + w_2u(y) + w_3u(z) > w_1u(x') + w_2u(y') + w_3u(z), \qquad (6)$$

where w_1, w_2, and w_3 are the weights of the highest-, middle-, and lowest-ranked branches, respectively, which will depend on the value of p (differently in different models). The generic model allows us to subtract the term $w_3u(z)$ from both sides, which yields

$$S \succ R \;\Leftrightarrow\; \frac{w_2}{w_1} > \frac{u(x') - u(x)}{u(y) - u(y')}. \qquad (7)$$

Suppose that there is a violation of 3-LDI, in which $S2 \prec R2$. By a similar derivation, this means

$$S2 \prec R2 \;\Leftrightarrow\; \frac{u(x') - u(x)}{u(y) - u(y')} > \frac{w_2'}{w_1'}, \qquad (8)$$

where the (primed) weights now depend on the new level of probability, p'. Therefore, there can be a preference reversal from $S \succ R$ to $S2 \prec R2$ if and only if the ratio of weights changes as a function of probability and "straddles" the ratio of differences in utility, as follows:

$$\frac{w_2}{w_1} > \frac{u(x') - u(x)}{u(y) - u(y')} > \frac{w_2'}{w_1'}. \qquad (9)$$

A reversal from $S \prec R$ to $S2 \succ R2$ can occur with the opposite ordering:

$$\frac{w_2}{w_1} < \frac{u(x') - u(x)}{u(y) - u(y')} < \frac{w_2'}{w_1'}. \qquad (10)$$

The two patterns of violation, represented in Expressions (9) and (10), will be labeled $SR2$ and $RS2$, respectively.

If the ratio of weights is independent of p or p' (for example, as in EU where both ratios are 1), then there can be no violations of this property.

According to the RAM model, this ratio of weights is given by the following:

$$\frac{w_2}{w_1} = \frac{a(2,3)s(p)}{a(1,3)s(p)} = \frac{a(2,3)s(p')}{a(1,3)s(p')} = \frac{w_2'}{w_1'}. \qquad (11)$$

Therefore, RAM satisfies 3-LDI. With the further assumption that branch rank weights equal their objective ranks, this ratio will be $2/1$, independent of the value of p (or p').

According to the special TAX model, this ratio of weights can be written as follows:

$$\frac{w_2}{w_1} = \frac{t(p) + (\delta t(p)/4) - (\delta t(p)/4)}{t(p) - (2/4)\delta t(p)} = \frac{t(p)}{t(p)[1 - 2\delta/4]} = \frac{w_2'}{w_1'}. \qquad (12)$$

In addition, if $\delta = 1$, as in the previously parameterized model, then this ratio will be $2/1$. Therefore, special TAX, like the RAM model, implies 3-LDI.

According to CPT, there should be violations of 3-LDI, if the decumulative weighting function, $W(P)$, is nonlinear. The relation among the weights will be as follows:

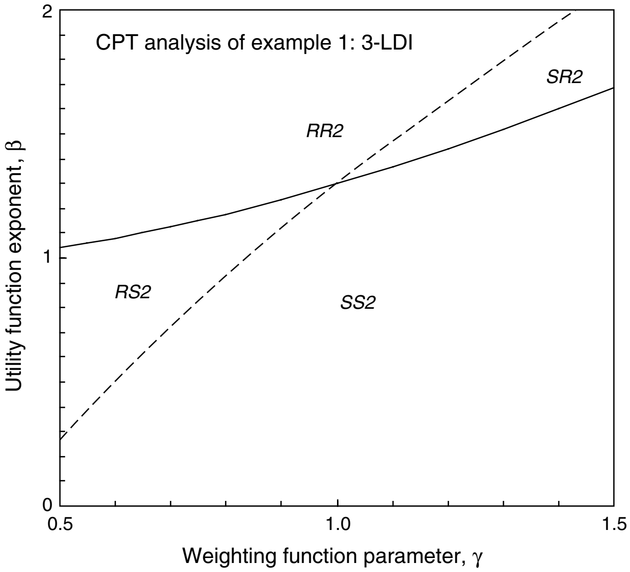
$$\frac{w_2}{w_1} = \frac{W(2p) - W(p)}{W(p) - W(0)} \neq \frac{W(2p') - W(p')}{W(p') - W(0)} = \frac{w_2'}{w_1'}.$$

For the model and parameters of Tversky and Kahneman (1992), the ratios of weights for $p = 0.2$ and $p' = 0.45$ are $0.1093/0.2608 = 0.4191$ and $0.3165/0.3952 = 0.80$, respectively. Note that the inverse-S weighting function of CPT implies that both of these ratios are less than 1, and they differ by a factor of almost 2.

Figure 1 illustrates how regions of the parameter space of the CPT model lead to different predicted patterns in Example 1. Note that when $\gamma = 1$, CPT reduces to EU, and no violations are allowed (i.e., only $SS2$ and $RR2$ are predicted). Based on $\beta = 0.88$ and $\gamma = 0.61$, the predicted pattern is $RS2$, as it is for many other combinations of parameters in this neighborhood. As noted by the reviews of Neilson and Stowe (2002) and Blavatskyy (2004), most studies supporting CPT have reported parameter estimates in this region (Abdellaoui 2000, Tversky and Kahneman 1992, Tversky and Fox 1995).

Two studies (Camerer and Ho 1994, Wu and Gonzalez 1996) reported smaller values of β, which imply the response pattern $SS2$ for Example 1. For that reason, it is important to provide a series of tests of a given property so that one can test a property without knowing the exact values of parameters in

Figure 1 **Predicted Pattern of Choices in Example 1 as a Function of Parameters of CPT**



Note. As shown, CPT is flexible enough to allow any pattern of behavior, but given its published estimates ($\gamma = 0.61$, $\beta = 0.88$), CPT predicts *RS*2 reversal of preferences. Each test can be analyzed in this way.

advance. Choices 9 and 12 of Table 1 provide a test in which the parameters reported in those studies imply a violation of the form *RS*2 of 3-LDI. Other tests of this paper widen still further the space of parameters explored by the study. Similar figures, drawn for each test of each property, show that the tests combine to provide a fairly wide and partially overlapped coverage of plausible CPT parameters.

3.2. Lower Distribution Independence (3-2 LDI)

Now consider the two choices in Example 2:

A: 0.50 probability to win \$40 0.50 probability to win \$44	*B*: 0.50 probability to win \$4 0.50 probability to win \$96
C: 0.04 probability to win \$2 0.48 probability to win \$40 0.48 probability to win \$44	*D*: 0.04 probability to win \$2 0.48 probability to win \$4 0.48 probability to win \$96

According to either TAX or RAM, with their previously estimated parameters, a person should choose *A* over *B* and *C* over *D*. For TAX, calculated certainty equivalents are 41.3, 34.7, 29.1, and 24.5, for *A*, *B*, *C*, and *D*, respectively. For RAM, the computed values are 41.3, 34.7, 35.5, and 29.8, respectively. However, according to the CPT model, the certainty equivalents are 41.7, 39.3, 33.9, and 37.9, respectively, so a person should prefer *A* over *B* and *D* over *C*. This property is named 3-2 *lower distribution independence* because the probability of the (common) lowest branch has vanished, reducing the choice from a comparison of three-branch gambles to a decision between two-branch gambles.

The property of 3-2 LDI requires

$$\begin{aligned} A &= (x, 1/2; y; 1/2) \\ &\succ B = (x', 1/2; y'; 1/2) \quad \text{if and only if} \\ C &= (x, p; y; p; z, 1-2p) \\ &\succ D = (x', p; y'; p; z, 1-2p). \end{aligned} \tag{13}$$

By a similar derivation to that for 3-LDI, the RAM model predicts the same choices between *A* and *B* as between *C* and *D*. This conclusion follows when

Table 1 **Tests of 3 Distribution Independence and Branch Independence (Main Design of Experiment 1)**

No.	Choice: *S*	Choice: *R*	Percentage choosing *R*: All (1,075)	2i (524)	2i* (551)	Prior TAX (RAM): *S*	Prior TAX (RAM): *R*	Prior CPT: *S*	Prior CPT: *R*
9	0.80 to win \$2 0.10 to win \$40 0.10 to win \$44	0.80 to win \$2 0.10 to win \$4 0.10 to win \$96	42.4	42.3	42.5	**11.4 (9.4)**	9.79 (8.2)	10.4	14.5
12	0.10 to win \$2 0.45 to win \$40 0.45 to win \$44	0.10 to win \$2 0.45 to win \$4 0.45 to win \$96	30.2	34.3	26.3	**27.1 (31.2)**	22.9 (26.2)	30.5	35.9
15	0.10 to win \$40 0.10 to win \$44 0.80 to win \$100	0.10 to win \$4 0.10 to win \$96 0.80 to win \$100	56.0	57.5	54.7	61.6 (68.6)	**63.4** (68.1)	77.4	71.7
18	0.45 to win \$40 0.45 to win \$44 0.10 to win \$100	0.45 to win \$4 0.45 to win \$96 0.10 to win \$100	33.2	32.4	33.9	45.9 (45.4)	43.9 (44.7)	50.3	42.6
5	0.04 to win \$2 0.48 to win \$40 0.48 to win \$44	0.04 to win \$2 0.48 to win \$4 0.48 to win \$96	34.0	35.0	33.0	**29.1 (35.5)**	24.5 (29.8)	34.7	37.7
6	0.50 to win \$40 0.50 to win \$44	0.50 to win \$4 0.50 to win \$96	31.4	31.6	31.1	41.3 (41.3)	34.7 (34.7)	41.7	39.3

Note. Choices 9 and 12 test 3-LDI, choices 15 and 18 test 3-UDI, and choices 5 and 6 test 3-2 LDI.

$a(2,2)/a(1,2) = a(2,3)/a(1,3)$, as it does in the parameterized RAM model, where this ratio is 2/1 in both cases.

The special TAX model also implies that people should make the same choices in A and B as between C and D because the weight ratio in 50-50, two-branch gambles is

$$\frac{w_2}{w_1} = \frac{t(1/2) + (\delta/3)t(1/2)}{t(1/2) - (\delta/3)t(1/2)},$$

which also equals 2/1, so this model also satisfies 3-2 LDI.

The CPT model predicts that people should choose A over B and choose D over C. The ratio of weights (second to first in the two-branch case) is $0.579/0.421 = 1.38$, but in the three-branch case, the weights of third, second, and highest branches are 0.185, 0.405, and 0.410, so the ratio of the second (middle) branch to the highest has been reduced to 0.99. Thus, CPT violates 3-2 LDI, leaving RAM and TAX to again defend the null hypothesis.

3.3. Upper Distribution Independence (3-UDI)

3-upper distribution independence, 3-UDI, can be written as follows:

$$\begin{aligned} S' &= (z', 1-2p; x, p; y; p) \\ &\succ R' = (z', 1-2p; x', p; y'; p) \quad \text{if and only if} \\ S2' &= (z', 1-2p'; x, p'; y; p') \\ &\succ R2' = (z', 1-2p'; x', p'; y'; p'). \end{aligned} \tag{14}$$

Example 3 illustrates 3-UDI:

S':	0.10 to win \$40 0.10 to win \$44 0.80 to win \$100	R':	0.10 to win \$4 0.10 to win \$96 0.80 to win \$100
$S2'$:	0.45 to win \$40 0.45 to win \$44 0.10 to win \$100	$R2'$:	0.45 to win \$4 0.45 to win \$96 0.10 to win \$100

Here, the probability of the common upper branch is changed. This property will be violated, with $S' \succ R'$ and $S2' \prec R2'$, if and only if

$$\frac{w_3}{w_2} > \frac{u(x') - u(x)}{u(y) - u(y')} > \frac{w_3'}{w_2'}. \tag{15}$$

The opposite pattern of preference ($R'S2'$) is predicted when the order above is reversed:

$$\frac{w_3}{w_2} < \frac{u(x') - u(x)}{u(y) - u(y')} < \frac{w_3'}{w_2'}. \tag{16}$$

The RAM model implies that

$$\frac{w_3}{w_2} = \frac{a(3,3)s(p)}{a(2,3)s(p)} = \frac{a(3,3)s(p')}{a(2,3)s(p')} = \frac{w_3'}{w_2'}.$$

Therefore, the RAM model implies no violations of 3-UDI.

In the special TAX model, however, this ratio of weights is as follows:

$$\frac{w_3}{w_2} = \frac{t(p) + (\delta/4)t(p) + (\delta/4)t(1-2p)}{t(p) - (\delta/4)t(p) + (\delta/4)t(1-2p)} \neq \frac{w_3'}{w_2'},$$

which shows that this weight ratio is not independent of p. For TAX with its previous parameters, this weight ratio increases from 1.27 to 1.60, which straddles the ratio of differences $(96-44)/(40-4) = 1.44$; therefore, this model predicts that $S' \prec R'$ and $S2' \succ R2'$. TAX implies this pattern ($R'S2'$) of violation of 3-UDI.

CPT implies violations of 3-UDI because the ratios of weights are as follows:

$$\begin{aligned} \frac{w_3}{w_2} &= \frac{1 - W(1-p)}{W(1-p) - W(1-2p)} \\ &\neq \frac{1 - W(1-p')}{W(1-p') - W(1-2p')} = \frac{w_3'}{w_2'}. \end{aligned}$$

In particular, for the inverse-S weighting function of Tversky and Kahneman (1992), this ratio of weights increases and then decreases as p increases, with a net decrease from 2.76 to 2.11 as p goes from 0.1 to 0.45. Therefore, CPT implies violations of the opposite pattern ($S'R2'$) from that implied by TAX. CPT predicts no violation in Example 3, given its previous parameters, but it does predict violations of the form $S'R2'$ in Choices 8 and 10 in Table 3. To account for the same pattern as predicted by TAX, CPT would need an S-shaped probability weighting function (i.e., with $\gamma > 1$), rather than inverse-S.

3.4. Restricted Branch Independence

The experimental design also allows one to test a special form of restricted branch independence (RBI), which in this experiment can be written as follows:

$$\begin{aligned} S &= (x, p; y; p; z, 1-2p) \\ &\succ R = (x', p; y'; p; z, 1-2p) \quad \text{if and only if} \\ S' &= (z', 1-2p; x, p; y; p) \\ &\succ R' = (z', 1-2p; x', p; y'; p). \end{aligned} \tag{17}$$

This version of RBI is a special case in which both (x, p) and (y, p) branches have equal probabilities. There will be an SR' violation of RBI (i.e., $S \succ R$ and $R' \succ S'$) if and only if the following holds:

$$\frac{w_2}{w_1} > \frac{u(x') - u(x)}{u(y) - u(y')} > \frac{w_3'}{w_2'}, \tag{18}$$

where w_1 and w_2 are the weights of the highest and middle branches, respectively (when both have probability p in gambles S and R), and w_2' and w_3' are the weights of the middle and lowest branches with probability p in gambles S' and R', respectively. Both RAM and TAX models imply this type of violation, SR'. With their parameters, the ratios are $2/1 > 3/2$ in both of these models. Although RAM and TAX satisfy all three or two of the three distribution independence properties, respectively, they both systematically violate RBI.

CPT also systematically violates RBI; however, it violates it in the opposite way from that of RAM and TAX, given its inverse-S weighting function. Define a *weakly inverse*-S function as a strictly increasing function satisfying the following: for all $p < p^*$: $W(2p) - W(p) < W(p) - W(0)$ and $W(1-p) - W(1-2p) < W(1) - W(1-p)$. Therefore, $w_2 < w_1$ and $w_2' < w_3'$. It follows that $w_2/w_1 < 1 < w_3'/w_2'$. Therefore, CPT with a weakly inverse-S weighting function implies that violations of RBI should have the relations $S \prec R$ and $R' \prec S'$; i.e., the RS' pattern. So, if the opposite pattern is observed, as predicted by RAM and TAX, it contradicts any weakly inverse-S function. (A strongly inverse-S function is one that is weakly inverse-S and also satisfies the following: $W(p) > p\ \forall p < p'$ and $W(p) < p\ \forall p > p'$. If one can reject the weakly inverse-S, one rejects the stronger form.)

4. Methods

In each of two experiments, participants made 20 or 22 choices between pairs of gambles. They viewed materials via the Internet. They were informed that three lucky participants in each experiment would be selected at random to play one of their chosen gambles for money, with prizes as high as $108, so they should choose carefully. Prizes were awarded as promised. Each choice was displayed as in the following example:

1. Which do you choose?
   O A: 0.50 probability to win $0
   0.50 probability to win $100
   or
   O B: 0.50 probability to win $25
   0.50 probability to win $35

Instructions read (in part) as follows:

> Think of probability as the number of tickets in a bag containing 100 tickets, divided by 100. Gamble A has 50 tickets that say $100 and 50 that say $0, so the probability to win $100 is 0.50 and the probability to get $0 is 0.50. If someone reaches in bag A, half the time they might win $0 and half the time $100. But in this study, you only get to play a gamble once, so the prize will be either $0 or $100. Gamble B's bag has 100 tickets also, but 50 of them say $25 and 50 of them say $35. Bag B thus guarantees at least $25, but the most you can win is $35. Some will prefer A and others will prefer B. To mark your choice, click the button next to A or B....

4.1. Experimental Design

Table 1 shows the main experimental design of Experiment 1. The "safe" gambles (denoted S in Table 1) contain branches with $(x, y) = (\$44, \$40)$. The "risky" gambles (denoted R in Table 1) had two branches with a greater range of consequences, $(x', y') = (\$96, \$4)$. Choices 9 and 12 provide a test of 3-LDI. Choices 15 and 18 provide a test of 3-UDI. Choices 5 and 6 form a test of 3-2 LDI.

Choices 9 and 15 allow a test of RBI, because the consequence on the common branch of 0.80 to win $2 has been changed from $2 to $100. Choices 12 and 18 form another test of RBI where the common branch has a probability of 0.10.

In addition to the choice sets displayed in Table 1, there were 10 other variations of the first five choices in Table 1, where only the values for the prizes in the "safe" gambles were altered. These values were $(\$38, \$34)$ and $(\$32, \$28)$ instead of $(\$44, \$40)$, respectively.

The first four choices served as a "warm-up," and were the same as the first four in Birnbaum (1999b). These choices replicated previous findings, showing risk aversion for 50-50 gambles and satisfaction of consequence monotonicity in 90% or more of the choices. There were two conditions, labeled 2i and 2i*, which had opposite trial orders for choices 5–20.

Experiment 2 was designed to create violations of 3-LDI and 3-UDI, according to the parameterized CPT model. It used a different mechanism of probability (drawing colored marbles randomly from an urn). Order of branches within gambles and positions of S and R in each choice (first or second) were also counterbalanced in Experiment 2 relative to Experiment 1. Experiment 2 also tested new variations of other properties previously published, which will be reported in the discussion by way of reviewing the case against CPT. Complete materials can be viewed at http://psych.fullerton.edu/mbirnbaum/archive.htm.

4.2. Participants

Participants in Experiment 1 consisted of 1,075 people recruited by links on the Internet and from the usual psychology "subject pool" at California State University, Fullerton. The two conditions (with opposite trial orders for choices 5–20) had 524 and 551 participants, respectively. Of these, 688 (65%) indicated they were female and 374 were males (some did not report gender). Age ranged from 18 to 83, with a mean of 29.5 years; 17.5% indicated that they were 40 or older; 44% reported that they held college degrees; and 3% had doctoral degrees. Experiment 2 had 503 participants, with a greater proportion of college students. Their mean age was 25.5, with 12.4% older than 40; 65% females; 28.3% were college graduates and 3% held doctorates.

5. Results

Tables 1 and 2 present choice percentages for the two groups and overall percentages (based on $n = 1{,}075$). All choice percentages in Table 1 are significantly different from 50%, by individual binomial tests. For $n = 524$, the binomial ($p = 1/2$) has a mean of 262 and a standard deviation of 11.45, so the 95% confidence interval on the mean proportion ranges from 0.46 to 0.54. For $n = 1{,}075$, the 95% confidence interval is 0.47 to 0.53. All values in Table 1 fall outside these intervals, and are therefore significantly different from 50%. Throughout this paper, the term "significant" and asterisks refer to this standard.

Predictions of TAX, RAM, and CPT models using parameters estimated from previous data are presented in the six columns in the right of Table 1. Calculated certainty equivalents are shown for each gamble according to each model. In Table 1, we see that the TAX and CPT models disagree in four rows: Choices 9, 12, 15, and 5. Bold type indicates cases where a model correctly predicts the modal choice when models disagree.

In all four cases in Table 1 where TAX and CPT disagree, TAX correctly predicted the modal choice (bold type). The TAX model implies no violations of 3-LDI (row 9 versus 12), but it correctly predicts a violation of 3-UDI (row 15 versus 18). Note also the predicted violation of restricted branch independence between choices 9 and 15. Here, CPT with its inverse-S weighting function erroneously predicts that $R \succ S$ in choice 9 and $R' \prec S'$ in choice 15; however, the TAX model correctly predicts the modal choices. Although the changes in choice proportion are not large by absolute standards ($56\% - 42\% = 14\%$), they are each significantly different from 50%, in opposite directions. By the more sensitive (within-subjects) test of correlated proportions, there were 273 who reversed choices between 9 and 15 in the direction predicted by TAX compared to only 128 who switched in the opposite direction ($z = 7.29$).

According to the CPT model, choice 6 should yield the opposite decision from that in choices 5 and 12. However, this predicted violation of 3-2 LDI did not materialize; percentages in choices 5 and 6 are 34% and 31%, respectively. Failure to reject does not "prove the null hypothesis," of course, so it is possible that with some other choices of gambles this prediction might still hold up. Nevertheless, CPT predicted violations of RAM and TAX, but the results did not confirm its predictions.

TAX and RAM differ in only one predicted choice in Table 1 (choice 15). RAM allows no violations of either 3-LDI or 3-UDI, whereas the TAX model predicts a particular violation of 3-UDI in choices 15 and 18. Although CPT can violate 3-UDI, it failed to predict this violation correctly predicted by TAX. In sum, the data in Table 1 are most consistent with the TAX model, which correctly predicted the modal choice in all six rows of Table 1. Next best was the RAM model, which predicted five of six choices correctly. The data are least consistent with the CPT model, which predicted the correct choice only in the two cases where it agreed with TAX and RAM.

Table 2 shows the same tests as in Table 1, except that consequences of the "safe" (S) gambles are reduced. In Table 2, RAM and TAX agree in all predictions of the direction of choice (hence, RAM values are not shown). In Table 2, the CPT model disagrees with TAX/RAM on all five choices. In four of these cases, TAX/RAM makes the correct prediction, and CPT is correct in one case. Choices 16 and 19 show

Table 2 Tests of Distribution Independence and Branch Independence

	Choice		Condition			Prior TAX		Prior CPT	
No.	*S*	*R*	All (1,075)	2i (524)	2i* (551)	*S*	*R*	*S*	*R*
10	0.80 to win \$2 0.10 to win \$34 0.10 to win \$38	0.80 to win \$2 0.10 to win \$4 0.10 to win \$96	47.1	44.3	49.8	**10.0**	9.8	9.2	14.5
13	0.10 to win \$2 0.45 to win \$34 0.45 to win \$38	0.10 to win \$2 0.45 to win \$4 0.45 to win \$96	37.2	40.5	34.1	**23.3**	22.9	26.3	35.8
16	0.10 to win \$34 0.10 to win \$38 0.80 to win \$100	0.10 to win \$4 0.10 to win \$96 0.80 to win \$100	59.9	58.9	60.7	57.7	**63.4**	75	71.7
19	0.45 to win \$34 0.45 to win \$38 0.10 to win \$100	0.45 to win \$4 0.45 to win \$96 0.10 to win \$100	44.2	42.3	46.0	40.3	43.9	**45.1**	42.6
7	0.50 to win \$34 0.50 to win \$38	0.50 to win \$4 0.50 to win \$96	38.8	38.5	39.1	**35.3**	34.7	35.7	39.3

Note. Entries show percentages choosing *R*.

another violation of 3-UDI, similar to that in Table 1. However, none of the previously fit models correctly predicted this violation. However, it would be consistent with TAX if some participants had slightly different parameters from those used in the calculations. It would be consistent with CPT only if the parameters are very different from those estimated in previous studies ($\gamma > 1$).

Among other results of Experiment 1 (not shown), CPT and TAX/RAM disagreed only in choice 17, where TAX and RAM correctly predicted that significantly more than half (63%) would prefer $R' =$ (\$100, 0.8; \$96, 0.1; \$4, 0.1), whereas CPT predicted that the majority should have chosen $S' =$ (\$100, 0.8; \$32, 0.1; \$28, 0.1) instead.

Table 3 shows results from Experiment 2 of two tests of 3-UDI and 3-LDI. These tests were designed so that CPT predicts violations, whereas RAM and TAX predict no changes. Here, the CPT model predicts that people should choose *R* in choice 6 and *S* in choice 12, *S* in choice 8, and *R* in choice 10. Instead, most people (69%) chose *S* in both choices 6 and 12, and the majority chose *R* in both choices 8 and 10. In fact, the percentage who chose *R* in choice 10 is significantly smaller than that in choice 8 ($z = 5.08$), contrary to the CPT model. Both TAX and RAM models correctly predicted all four modal choices, and CPT was correct only in the two cases for which it agreed with those models.

Experiment 2 also included a test of restricted branch independence in choices 20 and 16 (not shown). It was found that 109 people chose $S =$ (\$40, 0.2; \$38, 0.2; \$2, 0.6) over $R =$ (\$96, 0.2; \$4, 0.2; \$2, 0.6) and chose $R' =$ (\$100, 0.6; \$96, 0.2; \$4, 0.2) over $S' =$ (\$100, 0.6; \$40, 0.2; \$38, 0.2), whereas only 51 had the opposite pattern, RS', $z = 4.58$. This pattern is opposite the predictions of any weakly inverse-S weighting function in CPT but agrees with RAM and TAX.

In sum, TAX and CPT models with parameters estimated from previous data made different predictions in 10 cases in Experiment 1, and in nine of these, the majority choice was correctly predicted by TAX. In one case, RAM and TAX disagreed, with the TAX model correctly predicting that choice. In one case, prior CPT correctly predicted the majority. In none of the tests of 3-LDI or 3-2 LDI were there significant reversals of the majority preference; so CPT's predictions failed to materialize. However, in two of four tests of 3-UDI there were significant reversals of the majority choice, of the type consistent with TAX. Therefore, these data indicate that 3-UDI can be rejected, but that 3-LDI and 3-2 LDI can be retained, pending further tests. In Experiment 2, CPT predicted violations of both 3-LDI and 3-UDI, but the majority did not switch preferences, consistent with predictions of RAM and TAX. In both experiments, violations of RBI were systematic and opposite the predictions of the inverse-S weighting function of CPT.

Table 3 **Tests of Distribution Independence and Branch Independence (Experiment 2)**

No.	Choice *S*	Choice *R*	% *R* $n = 503$	Prior TAX *S*	Prior TAX *R*	Prior CPT *S*	Prior CPT *R*
6	60 blue to win \$2 20 red to win \$56 20 white to win \$58	60 green to win \$2 20 black to win \$4 20 purple to win \$96	23.6*	**21.7**	13.8	19.9	21.3
12	10 black to win \$2 45 green to win \$56 45 purple to win \$58	10 white to win \$2 45 red to win \$4 45 blue to win \$96	18.7*	36.9	22.9	41.0	35.8
8	20 white to win \$28 20 blue to win \$30 60 red to win \$100	20 yellow to win \$4 20 green to win \$96 60 black to win \$100	70.8*	47.3	**57.4**	61.7	60.7
10	45 white to win \$28 45 purple to win \$30 10 blue to win \$100	45 black to win \$4 45 green to win \$96 10 red to win \$100	59.1*	34.1	43.9	39.3	42.6

Note. In these tests, prior CPT predicts violations of both 3-LDI and 3-UDI, and RAM makes the same predictions as TAX in these tests. Entries show percentages who chose *R*.

*Indicates percentages significantly different from 50%.

6. Discussion

These studies add three new tests to the growing case against the CPT model. Whereas RAM must satisfy all three properties and TAX satisfies two, CPT can violate them all. There were significant violations in tests of 3-UDI, which is evidence against the RAM model. This violation was consistent with (and predicted by) TAX with its previous parameters. The CPT model can violate this property, but it failed to predict this violation with its previous parameters.

When CPT was used to design tests of 3-2 LDI, 3-UDI, and 3-LDI, its predictions were not confirmed empirically in either experiment. Failure to find predicted violations does not prove that there are no violations, but when a model fails to make accurate predictions to new tests and rival models are successful, it seems reasonable to prefer those models that continue to be accurate.

Defenders of CPT sometimes assert that RAM and TAX are more flexible (able to account for more data patterns than CPT), and that is why they have outperformed CPT in previous tests comparing these models. This claim of greater flexibility for RAM and TAX is overstated, however, and it does not apply to the present tests. In this paper, CPT is the most flexible model (in principle) because it could (post hoc) account for violations of all of the properties tested in Tables 1–3. Furthermore, this paper compared models without estimating parameters from the data, so the models compared are equally inflexible.

One can change parameters (giving up the inverse-S weighting function) and fit CPT to the data in Tables 1–3. Indeed, with the following weighting function,

$$W(P) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}, \tag{19}$$

where $\gamma = 1.292$ and $\delta = 0.724$, and with $u(x) = x^{\beta}$, where $\beta = 0.608$, CPT correctly reproduces the modal choices in Tables 1–3. However, this weighting function with $\gamma > 1$ is S-shaped, rather than inverse-S shaped, so this version of CPT would not account for the data of Tversky and Kahneman (1992), nor those of Wu and Gonzalez (1996), nor other sets of data, including the Allais paradoxes (Birnbaum 2004a). Two other empirical properties contradict CPT with an inverse-S weighting function: violations of restricted branch independence in four-branch gambles, and violations of four-distribution independence (Birnbaum 2004a, Birnbaum and Chavez 1997, Birnbaum and Navarrete 1998).

These experiments had many trials in which a person could have cancelled the common branch. If people cancel consistently, as proposed by Kahneman and Tversky (1979), there would have been no systematic violations of RBI, or 3-UDI. If we treat the editing principle of cancellation as a freestanding scientific theory rather than as a partial process, we must reject it when we observe systematic violations of restricted branch independence or upper distribution independence, as in Table 1. On the other hand, if we treat cancellation as something that occurs only part of the time, we conclude that without it, violations of the inverse-S weighting function would have been even stronger (Birnbaum 2004a).

There are three ways to interpret the results: First, it is possible that CPT might hold, but its parameters change from choice to choice and experiment to experiment in such a way that one cannot use it to predict. Second, it is possible that these studies might have yielded more compatible results if some missing ingredient had been included in the experiments (different people, different procedures, different incentives, etc.). Third, it is possible that CPT is simply not an accurate description of how people make decisions.

In my opinion, the best interpretation is that CPT is not an accurate descriptive model. It may approximate some empirical results, but it appears to be accurate only when its predictions agree with those of TAX. This conclusion runs against what has been a growing consensus in economics in favor of the CPT model (Starmer 2000). However, it is important to keep two considerations in mind: first, CPT and TAX models give very similar predictions for two-branch gambles and agree in 94% of randomly devised choices between three-branch gambles. Second, those studies supporting CPT did not compare it against RAM or TAX.

In addition to results from my lab, others have also reported results that do not fit with CPT. For example, Neilson and Stowe (2002) found that CPT parameters estimated in lab studies that test one phenomenon do not necessarily agree with results of other tests. Gonzalez and Wu (2003) tried to predict the values of three-branch gambles from parameters estimated from either original prospect theory (PT) or CPT applied to two branch gambles. They concluded that neither PT nor CPT provided predictions as accurate as the RAM model.

6.1. Strong Paradoxes That Refute Cumulative Prospect Theory

Because the findings in Tables 1–3 can be described by CPT with different parameters, these results might be described as "weak" paradoxes. However, there are six "new, strong paradoxes" that contradict any form of RDU, RSDU, or CPT.

First, event-splitting effects (violations of coalescing combined with transitivity) were reported by Starmer and Sugden (1993), Humphrey (1995), and replicated by others. Birnbaum (1999a) noted that these results are consistent with the TAX model and violate any form of CPT. Although Luce (2000) expressed doubt about the between-subjects tests in those studies, Birnbaum (1999b, 2004a, b) reported strong violations within subjects.

A new test of event splitting is shown in choices 9 and 15 in Table 4, from Experiment 2. Here, R is the same and S is presented with its higher branch either split (choice 9) or coalesced (choice 15). Results show that significantly less than half (41.6%) chose R when it is paired with the split version of S. However, significantly more than half (69.5%) chose the same R paired with the coalesced version of S; 177 participants switched from S to R compared with only 36 who had the opposite reversal, $z = 9.66$. The TAX model correctly predicted this reversal. CPT with any functions and parameters requires the same decisions in choices 9 and 15 of Table 4.

Choice 7 in Table 4 might be described as a "transparent" new test of coalescing. These two gambles should be identical under any version of CPT; how-

Table 4 New Tests of Event-Splitting (Choices 9, 15, and 7), Upper Tail Independence (Choices 22 and 19), and Stochastic Dominance (Choice 5) in Experiment 2

	Choice		% *R*	Prior TAX		Prior CPT	
No.	*S*	*R*	$n = 503$	*S*	*R*	*S*	*R*
9	10 red to win \$40 10 blue to win \$40 80 white to win \$2	10 green to win \$90 10 yellow to win \$4 80 orange to win \$2	41.6*	**11.1**	9.3	9.9	13.7
15	20 red to win \$40 80 blue to win \$2	10 green to win \$90 10 white to win \$4 80 yellow to win \$2	69.5*	9.0	9.3	9.9	13.7
7	90 red to win \$98 05 white to win \$12 05 blue to win \$12	85 green to win \$98 05 yellow to win \$98 10 orange to win \$12	67.4*	46.0	**65.9**	73.5	73.5
22	80 red to win \$110 10 yellow to win \$44 10 blue to win \$40	80 black to win \$110 10 purple to win \$96 10 green to win \$10	67.3*	65.0	**69.0**	83.5	79.9
19	80 red to win \$96 10 white to win \$44 10 blue to win \$40	90 green to win \$96 10 yellow to win \$10	33.1*	60.3	57.2	75.0	71.4
5	*G*: 90 red to win \$99 05 white to win \$14 05 blue to win \$12	*H*: 85 red to win \$98 05 white to win \$96 10 blue to win \$11	57.4*	47.0	**65.4**	74.4	73.4

*Indicates percentages significantly different from 50%.

ever, significantly more than half (67.4%) chose the gamble in column *R*, which has two high-valued branches, as opposed to the gamble in column *S*, which has two low-valued branches, $z = 5.53$, consistent with TAX.

Second, Wu (1994) reported violations of upper tail independence, which violate CPT. He theorized that the violations might be due to editing, rather than to a basic flaw in the CPT model. Wu's findings were replicated (with variations) by Birnbaum (2001), who noted that the pattern of violation is consistent with the TAX model. A new test of upper tail independence from Experiment 2 is shown in choices 22 and 19 of Table 4. Unlike previous tests of this property, this new test does not use a smallest consequence of \$0 in all four gambles.

In choice 22, there is a common branch (80 marbles to win \$110) in both *S* and *R*. This consequence can be reduced from \$110 to \$96 in both gambles without changing the preference, according to CPT. In addition, CPT implies that the two branches that now yield \$96 can be coalesced in *R* without changing its utility. However, the percentage choosing *R* drops from significantly more than half in choice 22 (67.3%) to significantly less than half in choice 19 (33.1%). 209 participants showed this switch, compared with only 37 who showed the opposite reversal, $z = 11.0$. The TAX model with its previously estimated parameters predicts this reversal, and no version of CPT can reproduce it.

Third, Birnbaum (1997) deduced that his models should show violations of first-order stochastic dominance in a special choice recipe. Birnbaum and Navarrete (1998) then tested this recipe and found that undergraduates show about 70% violations in four variations of the recipe, significantly more than 50% in each test. Choice 5 in Table 4 is a new version of this test.

Note that in choice 5 of Table 4, *G* dominates *H*, because for any prize, the probability of winning that amount or more is either the same or greater in *G* than it is in *H*. The probability of winning \$99 or more, \$98 or more, \$14 or more, and \$12 or more is higher in *G* than in *H*, and for other values (\$96, \$11, etc.), it is the same in *G* and *H*. In this test, unlike previous ones, two of the three branches even have higher consequences in *G* than in *H*. So, if a person were using a heuristic of comparing consequences on corresponding branches, and choosing the gamble with the greater number of higher consequences, that person should also conform to stochastic dominance in this test. Any version of CPT requires that people should choose *G* over *H*, except for random error. However, TAX predicts that the majority will choose *H* over *G*. In fact, significantly more than half the participants (57.4%, $z = 2.34$) chose *H*, in violation of stochastic dominance, the counting heuristic, and any version of CPT.

It is important to distinguish tests of first-order stochastic dominance as above from other types of "stochastic dominance" such as those debated by Levy and Levy (2002) and Wakker (2003). Although CPT with its nonlinear weighting function may be able to account for the results of Levy and Levy (2002), it cannot account for violations of first-order stochastic dominance, as reported by Birnbaum and Navarrete (1998) or in the new example of choice 5 in Table 4.

Forth, Wu and Markle (2004) developed an example from Levy and Levy (2002) to construct a test of gain-loss separability, which they concluded is violated, in contradiction to CPT. They fit a type of configural weight model to their results. If their results hold up, violations of this property represent another important refutation of CPT. The pattern of violation is consistent with either RAM or TAX, even with a linear utility function, with the assumption that the configural weight parameter depends on whether the gambles include all gains, all losses, or mixed consequences.

Birnbaum (1997) deduced two theorems from RDU/CPT, called lower cumulative independence (the fifth paradox) and upper cumulative independence (the sixth paradox). These can be deduced from coalescing, consequence monotonicity, co-monotonic restricted branch independence, and transitivity. Violations can also be interpreted as contradictions in the probability weighting function of CPT (Birnbaum et al. 1999, Appendix). Birnbaum and Navarrete (1998) found systematic violations of both properties, as predicted by the TAX model but contradicting CPT.

These six "strong" paradoxes contradict any form of CPT—no choice of utility function and weighting function could resolve them. Although CPT might be modified to account for the three new independence properties in Tables 1–3 with new parameters and functions, there are no functions and parameters that allow CPT to account for "strong" paradoxes, as in Table 4.

Birnbaum (2004b) replicated four of these strong paradoxes with a dozen different variations of procedure, including different ways to present probability and format choices. Research with thousands of participants has been accumulated, showing that findings refuting CPT are quite robust with respect to these procedural variations. Tutorials on the "new paradoxes" are available from http://psych.fullerton.edu/mbirnbaum/talks/.

6.2. Concluding Comments

The present data add three new phenomena to the list that must be explained by descriptive theory: (1) Systematic violations of 3-UDI were observed, which were correctly predicted by TAX but not by CPT and which violate RAM. (2) Results failed to refute 3-LDI, a property that should be violated by CPT, but not by either RAM or TAX. (3) Data also failed to violate 3-2 LDI, another failure to confirm implications of CPT where RAM and TAX predicted correctly.

Subsequent to the completion of this research, Marley and Luce (in press) developed a general theoretical analysis of a class of ranked additive utility representations that includes the class of ranked weighted utility representations. They conclude that the latter class is equivalent to general TAX, and thus includes special cases such as the parameterized TAX and RAM models. The weighted utility class also includes RDU, CPT, and Marley and Luce's (2001) gains decomposition utility (GDU). Their (lower) GDU model, like the special TAX model, implies satisfaction of 3-LDI and violates 3-UDI. So, this GDU model is also consistent with the findings in Tables 1–3.

However, the lower GDU model (Marley and Luce 2001) implies upper coalescing (that one can coalesce upper branches of a gamble when their consequences are equal). Therefore, lower GDU is not consistent with the new test of upper coalescing (choices 9 and 15 of Table 4), nor can it describe violations of upper tail independence (choices 22 and 19 of Table 4). Marley and Luce developed more general forms of GDU and worked out the conditions under which each of the models will satisfy or violate these new independence properties. They identified tests that would in principle distinguish TAX from the new GDU models, which should lead to new empirical research to compare TAX and GDU models that remain consistent with empirical choice behavior.

Acknowledgments

Support was received from National Science Foundation Grants SBR-9410572, SES 99-86436, and BCS-0129453. The author thanks Detlof von Winterfeldt, Tony Marley, and two reviewers for helpful suggestions on the manuscript.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Sci.* **46** 1497–1512.

Birnbaum, M. H. 1974. The nonadditivity of personality impressions. *J. Experimental Psych.* **102** 543–561.

Birnbaum, M. H. 1997. Violations of monotonicity in judgment and decision making. A. A. J. Marley, ed. *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce.* Erlbaum, Mahwah, NJ, 73–100.

Birnbaum, M. H. 1999a. Paradoxes of Allais, stochastic dominance, and decision weights. J. Shanteau, B. A. Mellers, D. A. Schum, eds. *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Kluwer Academic Publishers, Norwell, MA, 27–52.

Birnbaum, M. H. 1999b. Testing critical properties of decision making on the Internet. *Psychological Sci.* **10** 399–407.

Birnbaum, M. H. 2001. A Web-based program of research on decision making. U.-D. Reips, M. Bosnjak, eds. *Dimensions of Internet Science*. Pabst Science Publishers, Lengerich, Germany, 23–55.

Birnbaum, M. H. 2004a. Causes of Allais common consequence paradoxes: An experimental dissection. *J. Math. Psych.* **48**(2) 87–106.

Birnbaum, M. H. 2004b. Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organ. Behavior Human Decision Processes* **95** 40–65.

Birnbaum, M. H., A. Chavez. 1997. Tests of theories of decision making: Violations of branch independence and distribution independence. *Organ. Behavior Human Decision Processes* **71**(2) 161–194.

Birnbaum, M. H., W. R. McIntosh. 1996. Violations of branch independence in choices between gambles. *Organ. Behavior Human Decision Processes* **67** 91–110.

Birnbaum, M. H., J. B. Navarrete. 1998. Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *J. Risk Uncertainty* **17** 49–78.

Birnbaum, M. H., J. N. Patton, M. K. Lott. 1999. Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic, dominance, and transitivity. *Organ. Behavior Human Decision Processes* **77** 44–83.

Blavatskyy, P. R. 2004. Back to the St. Petersburg paradox? *Management Sci.* **51**(4) 677–678.

Camerer, C. F., T.-H. Ho. 1994. Violations of the betweenness axiom and nonlinearity in probability. *J. Risk Uncertainty* **8** 167–196.

Gonzalez, R., G. Wu. 2003. Composition rules in original and cumulative prospect theory. Working paper, Department of Psychology, University of Michigan, Ann Arbor, MI.

Humphrey, S. J. 1995. Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *J. Risk Uncertainty* **11** 263–274.

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

Levy, M., H. Levy. 2002. Prospect theory: Much ado about nothing. *Management Sci.* **48** 1334–1349.

Luce, R. D. 2000. *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Lawrence Erlbaum Associates, Mahwah, NJ.

Luce, R. D., P. C. Fishburn. 1991. Rank- and sign-dependent linear utility models for finite first order gambles. *J. Risk Uncertainty* **4** 29–59.

Luce, R. D., P. C. Fishburn. 1995. A note on deriving rank-dependent utility using additive joint receipts. *J. Risk Uncertainty* **11** 5–16.

Marley, A. A. J., R. D. Luce. 2001. Rank-weighted utilities and qualitative convolution. *J. Risk Uncertainty* **23**(2) 135–163.

Marley, A. A. J., R. D. Luce. In press. Independence properties vis-à-vis several utility representations. *Theory Decision* Forthcoming.

Neilson, W., J. Stowe. 2002. A further examination of cumulative prospect theory parameterizations. *J. Risk Uncertainty* **24**(1) 31–46.

Quiggin, J. 1993. *Generalized Expected Utility Theory: The Rank-Dependent Model*. Kluwer, Boston, MA.

Savage, L. J. 1954. *The Foundations of Statistics*. Wiley, New York.

Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *J. Econom. Literature* **38** 332–382.

Starmer, C., R. Sugden. 1993. Testing for juxtaposition and event-splitting effects. *J. Risk Uncertainty* **6** 235–254.

Tversky, A., C. R. Fox. 1995. Weighing risk and uncertainty. *Psychological Rev.* **102**(2) 269–283.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* **5** 297–323.

Tversky, A., P. Wakker. 1995. Risk attitudes and decision weights. *Econometrica* **63** 1255–1280.

Wakker, P. 2003. The data of Levy and Levy (2002), "Prospect theory: Much ado about nothing?" Support prospect theory. *Management Sci.* **49** 979–981.

Wu, G. 1994. An empirical test of ordinal independence. *J. Risk Uncertainty* **9** 39–60.

Wu, G., R. Gonzalez. 1996. Curvature of the probability weighting function. *Management Sci.* **42** 1676–1690.

Wu, G., A. B. Markle. 2004. An empirical test of gain-loss separability in prospect theory. Working paper, Graduate School of Business, University of Chicago, Chicago, IL.