Journal of Risk and Uncertainty, 17:49–78 (1998)

Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence

MICHAEL H. BIRNBAUM
California State University, Fullerton and Decision Research Center

JUAN B. NAVARRETE
California State University, Fullerton

Abstract

Choices between gambles show systematic violations of stochastic dominance. For example, most people choose (\$6, .05; \$91, .03; \$99, .92) over (\$6, .02; \$8, .03; \$99, .95), violating dominance. Choices also violate two cumulative independence conditions: (1) If $S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q)$ then $S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q)$. (2) If $S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r)$ then $S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r)$, where $0 < z < x' < x < y < y < y' < z'$.

Violations contradict any utility theory satisfying transivity, outcome monotonicity, coalescing, and comonotonic independence. Because rank-and sign-dependent utility theories, including cumulative prospect theory (CPT), satisfy these properties, they cannot explain these results.

However, the configural weight model of Birnbaum and McIntosh (1996) predicted the observed violations of stochastic dominance, cumulative independence, and branch independence. This model assumes the utility of a gamble is a weighted average of outcomes' utilities, where each configural weight is a function of the rank order of the outcome's value among distinct values and that outcome's probability. The configural weight, TAX model with the same number of parameters as CPT fit the data of most individuals better than the model of CPT.

Key words: choice, gambles, expected utility, prospect theory, rank dependent utility, rank- and sign-dependent utility, stochastic dominance violations

JEL Classification: C910, D810

Introduction

This study investigates a recipe that produces systematic violations of stochastic dominance in choices between gambles. The theoretical section shows that violations of stochastic dominance in this recipe refute any utility theory that implies outcome monotonicity, transitivity, and coalescing. These properties are assumed or implied by rank dependent or rank- and sign-dependent utility (RDU and RSDU) thoeries (Luce & Fishburn, 1991; 1995; Lopes, 1990; Quiggin, 1982; Wakker, Erev, & Weber, 1994), including cumulative prospect theory (CPT) (Tversky & Kahneman, 1992; Wakker & Tversky, 1993). The theoretical section also derives two cumulative independence conditions from the above three assumptions plus comonotonic independence, a property required by CPT and RSDU models (Wakker, 1996; Wakker & Tversky, 1993). Again, RDU and RSDU theories satisfy these conditions and therefore imply cumulative independence.

The recipes used to create these violations were developed from configural weight theory (CWT) by Birnbaum (1997). Configural weight models were introduced in psychology in the 1970s (e.g., Birnbaum, 1973; 1974; Birnbaum & Stegner, 1979) and have much in common with RDU models that were introduced to the economics literature in the 1980s (Quiggin, 1982). Birnbaum (1974) noted, "the configural weight averaging model assumes that the weight of a stimulus depends upon its rank within the set to be judged." However, despite similarities, the two classes of models are distinct because the nature of the rank-dependence is different. This study is designed to test the class of RDU/RSDU/CPT models against the configural weight models.

The empirical sections of the paper show that choices systematically violate stochastic dominance and cumulative independence in the manner predicted in advance of the experiment by the configural weight model and parameters of Birnbaum and McIntosh (1996). We also compare the fit of the configural weight, TAX model against that of CPT. Results show that the TAX model gives a better fit to the data of individuals, fit separately to the models, despite the fact that both models use the same number of parameters. The discussion shows how this study fits as a piece in the much broader puzzle of research testing descriptive models of decision making under risk and uncertainty.

1. Definitions and a Recipe for Violations

Let $\succ$ represent the empirical preference relation between gambles; $A \succ B$ means that gamble A is preferred to B, and let $\sim$ represent indifference between gambles. Assume preferences among gambles satisfy a weak order. Let $G = (x, p; y, q; z, r)$ denote the three-outcome gambling yielding monetary outcome x with probability p, y with probability q, and z with probability $r = 1 - p - q$.

1.1 Definitions

1. *Transitivity.* The preference relation, $\succ$, is defined to be *transitive* if and only if,

$$A \succ B \text{ and } B \succ C \Rightarrow A \succ C. \quad (1)$$

2. *Outcome monotonicity.* Increasing the value of one outcome, holding everything else constant in the gamble should improve the gamble. Therefore, for three-outcome gambles with outcomes selected such that, $x^+ > x$, $y^+ > y$, and $z^+ > z$, outcome monotonicity requires:

$$(x^+, p; y, q; z, r) \succ G = (x, p; y, q; z, r) \quad (2a)$$

$$(x, p; y^+, q; z, r) \succ G \quad (2b)$$

$$(x, p; y, q; z^+, r) \succ G \tag{2c}$$

Outcome monotonicity has been violated in judgment but not in direct choice, nor when all outcomes are positive (Birnbaum & Sutton, 1992; Birnbaum, 1992; 1997).

3. *Coalescing Equivalence.* Coalescing implies that if two or more outcomes have the same value within a gamble, they can be combined by adding their probabilities. For three-outcome gambles, coalescing requires,

$$(x, p; x, q; z, r) \sim (x, p + q; z, r) \tag{3a}$$

$$(x, p; y, q; y, r) \sim (x, p; y, q + r) \tag{3b}$$

Coalescing is implied by rank- and sign-dependent utility theories, as will be shown in Section 2.3. Coalescing was also proposed as an editing rule in original prospect theory (Kahneman & Tversky, 1979). "Event splitting effects" (Humphrey, 1995; Starmer & Sugden, 1993) are violations of coalescing combined with transitivity.

4. *Branch Independence.* Branch independence is the assumption that if two gambles have a common outcome for a given state of the world with known probability, then the value of the outcome on that common branch should have no effect on the preference order induced by other components of the gambles. Branch independence is weaker than Savage's (1954) axiom because it requires that the outcomes be distinct and their probabilities known; it does not presume coalescing. For three outcome gambles, restricted branch independence implies,

$$\begin{gathered}(x, p; y, q; z, r) \succ (x', p; y', q; z, r)\\ \text{if and only if}\\ (x, p; y, q; z', r) \succ (x', p; y', q; z', r)\end{gathered} \tag{4}$$

where the outcomes are distinct, and the probabilities are nonzero and sum to 1. Branch independence is termed *unrestricted* when the probability distributions or number of outcomes in the gambles are permitted to differ (cf. Birnbaum, Coffey, Mellers, & Weiss, 1992). The term *comonotonic branch independence* is used to refer to the special case in which x and x', y and y', and z and z' have the same ranks in all four gambles (Wakker, 1996; Wakker, et al., 1994). The complementary case of Expression 4 is termed *noncomonotonic branch independence*. Systematic violations of comonotonic branch independence with six distinct values (x, y, z, x', y', z') have not yet been reported, but systematic violations of noncomonotonic branch independence have been (Birnbaum & Chavez, 1997; Birnbaum & McIntosh, 1996; Weber & Kirsner, 1997).

1.2 Stochastic Dominance

Stochastic dominance is the relation between nonidentical gambles, $A \neq B$, such that

$$A \text{ stochastically dominates } B \text{ iff } P(x > t \mid A) \geq P(x > t \mid B) \text{ for all } t. \quad (5)$$

where $P(x > t \mid A)$ is the probability that an outcome of Gamble A exceeds t. Stochastic dominance can be viewed as a combination of outcome monotoxicity and event monotonicity (Luce, 1988; Luce & von Winterfeldt, 1994). Violations of outcome monotonicity imply violations of stochastic dominance, but stochastic dominance violations do not necessarily imply violations of outcome monotonicity when this distinction is maintained (Birnbaum, 1997; Luce & von Winterfeldt, 1994). In traditional representations of gambles as probability distributions, where coalescing is presupposed, these concepts are equivalent. The term, "transparent dominance" refers to dominance between gambles that are otherwise identical, except one gamble has at least one higher outcome for the same event and probability (outcome monotonicity), or for choices in which the outcomes are the same and one gamble has a higher probability of a better outcome (probability monotonicity).

The statement that preferences satisfy stochastic dominance means,

$$\text{If } A \text{ stochastically dominates } B, \text{ then } A \succ B. \quad (6)$$

A violation of stochastic dominance occurs when A dominates B but B is chosen over A. To test for systematic violations with fallible data, we will test the hypothesis that if A stochastically dominates B, then the probability of choosing A over B should exceed 1/2. This is a very conservative test.

Tversky and Kahneman (1986) reported a violation of stochastic dominance in which dominance was "masked" by the problem "frame." The framing was accomplished by making it seem that the "same" events (same color of marble selected from an urn) gave equal or higher outcomes in the dominated gamble (because the number of marbles of each color varied, the events were not really the same). Tversky and Kahneman (1986) found that 58% of 124 judges chose the dominated gamble over the dominant gamble. Although 58% is not significantly greater than 50% ($\alpha = .05$), results were quite different from another framing of the choice in which the same color of marble always gave the same or a higher outcome in the dominant gamble. Despite such possible violations, Tversky and Kahneman (1992) advanced CPT, which satisfies stochastic dominance.

1.3 Paradigm for Stochastic Dominance Violations

In this section we present a recipe for pairs of gambles that possess a dominance relation and yet most judges choose the dominated gamble. Birnbaum (1997) proposed this recipe as a test between the class of models that satisfy dominance against configural weight models. Define $G_0 = (x, 1 - p; y, p)$, where $y > x > 0$. Create a worse gamble, $G-$, by splitting the upper branch into two pieces, with probabilities $p - r$ and r, one of which has a slightly lower outcome of y^- instead of y ($x < y^- < y$). Let $G- = (x, 1 - p; y^-, r; y, p - r)$. G_0 stochastically dominates $G-$. Now create a better gamble, $G+$, by splitting the lower branch of G_0 into two parts, one with a slightly higher value than x ($x^+ > x$); let $G+ = (x, 1 - p - q; x^+, q; y, p)$. $G+$ dominates G_0, which dominates $G-$, and $G+$ stochastically dominates $G-$. For example, G_0 = ($12, .1; $96, .9), with $q = r = .05$, let $G+$ = ($12, .05; $14, .05; $96, .9), which dominates $G-$ = ($12, .1; $90, .05; $96, .85).

Any theory in which preferences satisfy transitivity, outcome monotonicity and coalescing must satisfy stochastic dominance in this recipe. *Proof*: By outcome monotonicity, $G+ = (x, 1 - p - q; x^+, q; y, p) \succ (x, 1 - p - q; x, q; y, p) \sim G_0$, by coalescing. By coalescing, $G_0 \sim (x, 1 - p; y, r; y, p - r) \succ (x, 1 - p; y^-, r; y, p - r) = G-$, by outcome monotonicity. By transitivity, $G+ \succ G-$.

Birnbaum (1997) noted that the model of Birnbaum and McIntosh (1996), by violating coalescing, predicts violations of stochastic dominance in this recipe. As shown below, most judges chose $G-$ over $G+$, as predicted by this model and its previously published parameters.

1.4 Cumulative Independence

Birnbaum (1997) derived the following cumulative independence conditions for gambles selected such that $0 < z < x' < x < y < y' < z'$ and $p + q + r = 1$.

Lower Cumulative Independence:

$$\text{If } S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q)$$
$$\text{Then } S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q) \tag{7a}$$

Upper Cumulative Independence:

$$\text{If } S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r)$$
$$\text{Then } S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r) \tag{7b}$$

Any theory that staisfies comonotonic independence, monotonicity, transitivity, and coalescing must satisfy both lower and upper cumulative independence conditions.

Proof: $S \succ R$ implies $(x', r; x, p; y, q) \succ (x', r; x', p; y', q)$, by comonotonic independence (changing z to x' in both gambles). By monotonicity (increasing x to y in the gamble on the left), $(x', r; y, p; y, q) \succ (x', r; x, p; y, q) \succ (x', r; x', p; y', q)$; by transitivity, $(x', r; y, p; y, q) \succ (x', r; x', p; y', q)$; finally $(x', r; y, p + q) \succ (x', r + p; y', q)$, by coalescing, which is the same as $S'' \succ R''$, proving lower cumulative independence.

If $S' \prec R'$ then $(x, p; y, q; y', r) \prec (x', p; y', q; y', r)$, by comonotonic independence (reducing z' to y' in both gambles). By monotonicity (decreasing y to x in the gamble on the left), $(x, p; x, q; y', r) \prec (x, p; y, q; y', r) \prec (x', p; y', q; y', r)$; therefore, $(x, p; x, q;$

$y', r) < (x', p; y', q; y', r)$, by transitivity; finally, $(x, p + q; y', r) < (x', p; y', q + r)$, by coalescing, which is the same as $S''' < R'''$, thus proving upper cumulative independence.

1.5 Violations of Noncomonotonic Branch Independence

The same choices used in tests of upper and lower cumulative independence can also be used to test branch independence. In the notation of Section 1.4, there are two patterns of violation of noncomonotonic branch independence:

$$\text{SR}': \text{S} > \text{R and } S' < R' \quad (8a)$$

$$\text{RS}': \text{S} < \text{R and } S' > R'. \quad (8b)$$

As noted by Birnbaum and McIntosh (1996), EU and Subjectively Weighted Utility (SWU) theories imply branch independence. Both CPT and CWT allow violations of noncomonotonic branch independence. However, the theory of editing and elimination of common branches sometimes used in CPT (Wu, 1994) implies no systematic violations of branch independence.

2. Theories of Choice between Gambles

2.1 RSDU & Cumulative Prospect Theory

Cumulative prospect theory (Tversky & Kahneman, 1992) is a rank- and sign-dependent theory whose representation is the same as that of Luce and Fishburn (1991; 1995). Whereas cumulative prospect theory was derived from the assumption of comonotonic independence (Wakker & Tversky, 1993), Luce and Fishburn (1991; 1995) derived comonotonic independence from a theory of joint receipts (see also Luce, 1996). When the outcomes of the gamble are nonnegative, the RSDU representation simplifies to RDU. The RDU of a gamble, $G = (x_1, p_1; x_2, p_2; \ldots; x_i, p_i; \ldots x_n, p_n)$, where $x_1 > x_2 > x_3 > \ldots x_n > 0$; $\Sigma\, p_i = 1$, is given by the following:

$$RDU(G) = \sum_{i=1}^{n} \mathrm{u(x_i)[W(P_i) - W(P_{i-1})]} \quad (9)$$

where P_i is the (decumulative) probability of receiving x_i or more, and P_{i-1} is the probability of receiving more than x_i. The function, $u(x)$, is the utility function of the monetary outcome, x. $W(P)$ is a strictly increasing, monotonic, cumulative weighting function, which assigns $W(0) = 0$ and $W(1) = 1$. If $W(P) = P$, Equation 9 reduces to EU theory. In a choice between gambles, $A > B$ if and only if RDU(A) > RDU(B).

The model of Tversky and Kahneman (1992) further specified $W(P)$ to be an inverse-S function of P, with steeper slopes near zero and one than near 1/2, with the following equation:

$$W(P) = \frac{P^{\gamma}}{[P^{\gamma} + (1 - P^{\gamma})]^{\frac{1}{\gamma}}} \tag{10a}$$

where $\gamma = .61$; they also estimated $u(x) = x^{.88}$ for $x \geq 0$. Equation 10a with $\gamma < 1$ produces an inverse-S curve; when $\gamma > 1$, the function is S-shaped. [Tversky and Wakker (1995) used the term "S-shaped" for what we term "inverse-S"]. Certainty equivalents of gambles are calculated from the following equation: $\mathrm{CE}(G) = u^{-1}[\mathrm{RDU}(G)]$, where $\mathrm{CE}(G)$ is the certainty (cash) equivalent of gamble G, and u^{-1} is the inverse of $u(x)$. In this paper, when values of gambles are calculated from the cumulative prospect *model* (as opposed to the more general theory of Equation 9, which allows any $W(P)$ function), they will be calculated using this model and parameters.

Other forms for $W(P)$ have been discussed by Tversky and Wakker (1995) and Luce (1996). A two-parameter equation for $W(P)$ is as follows:

$$W(P) = \frac{cP^{\gamma}}{cP^{\gamma} + (1 - P)^{\gamma}} \tag{10b}$$

This formula can approximate Equation 10a, but it is more flexible (Tversky & Wakker, 1995). Equation 10b will be used in fitting the CPT model to the data of this study.

2.2 Configural Weight RAM and TAX Models

Birnbaum (1997) reviewed configural weight models investigated by Birnbaum and his colleagues (Birnbaum, 1973; 1974; Birnbaum & Chavez, 1997; Birnbaum et al., 1992; Birnbaum & McIntosh, 1996; Birnbaum & Beeghley, 1997; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992; Birnbaum & Veira, 1998; Birnbaum & Zimmermann, 1998), and described tests for distinguishing various configural models. Configural weight models are similar to rank-dependent models, in that both use rank-affected weights to account for violations of independence, but they make different predictions for this study.

The Rank Affected Multiplicative (RAM) model of Birnbaum, et al. (1992), as extended by Birnbaum and McIntosh (1996), assumes that the utility of a three-outcome gamble can be represented as a configurally-weighted average of the utilities of the outcomes:

$$\mathrm{CWU}(x, p; y, q; z, r) = \frac{Au(x) + Bu(y) + Cu(z)}{A + B + C} \tag{11}$$

where $u(x)$, $u(y)$, and $u(z)$ are the utilities of the monetary outcomes, x, y, and z ($0 < x < y < z$); A, B, and C are the weights of the outcomes, which depend on the ranks of the outcomes, the probabilities of the outcomes ($p + q + r = 1$), and the point of view of the judge, as follows:

$$A = w_L S(p) \tag{12a}$$

$$B = w_M S(q) \tag{12b}$$

$$C = w_H S(r) \tag{12c}$$

where w_L, w_M, and w_H are the configural weights of the lowest, medium, and highest outcomes in the gamble, and $S(p)$ is a function of probability. Configural weights are assumed to depend on the judge's point of view, which is different for buying prices, selling prices, and choices. However, $u(x)$ and $S(p)$ are assumed to be independent of viewpoint. Birnbaum et al. (1992) assumed that $w_L > w_M = w_H$, but Birnbaum and McIntosh (1996) concluded that $w_L/w_M < w_M/w_H$. When $w_L = w_M = w_H$, Expressions 11–12 reduce to subjectively weighted average utility, of which the models of Karmarkar (1978), Lattimore, Baker, and Witte (1992), and Viscusi (1989) are relatives with different $S(p)$ functions. When $w_L = w_M = w_H$ and $S(p) = p$, the model reduces to EU.

Birnbaum and McIntosh (1996), in a choice task, estimated w_L, w_M, and w_H to be .51, .33, and .16, respectively, and $u(x) = x$ for $0 < x < \$150$. Birnbaum and Beeghley (1997) found different weights for judgments of buying and selling prices. However, studies of choice, buyer's price, and seller's price all found configural weights satisfying $w_L/w_M < w_M/w_H$. This relation among weights predicts violations of branch independence in the opposite direction from those predicted by the inverse-S weighting function used in CPT.

Birnbaum (1997) showed that if $S(p) = p^{\gamma}$, where $\gamma = .6$, then Equations 11–12 can fit the certainty equivalents in Tversky and Kahneman (1992) with a linear utility function, for $0 < x < \$150$. When there are only two outcomes, the weight of the (absent) middle outcome in Equation 11 (B) is set to zero, and different configural weights are allowed for the lower and higher outcomes. To fit the Tversky and Kahneman (1992) data, the weights ($n = 2$) are $w_L = .63$ and $w_H = .37$. These weights agree with the results of Birnbaum, et al. (1992) in the "neutral" viewpoint and of Birnbaum, Thompson, and Bean (1997) who obtained judgments of strengths of preference to test interval independence. In the present paper, calculations based on the Birnbaum and McIntosh (1996) model will be illustrated using the parameters in that paper and the assumptions that $S(p) = p^{.6}$, and $u(x) = x$. These parameters fit certainty equivalents of binary gambles (Tversky & Kahneman, 1992), violations of common consequence independence (Wu & Gonzalez, 1996), and violations of branch independence (Birnbaum & McIntosh, 1996).

Birnbaum (1997) showed that the RAM model predicts violations of stochastic dominance and cumulative independence, and the present study was designed based on the predictions of that model and its previously published parameters. We use choices between gambles that are predicted by that model to violate stochastic dominance and cumulative independence.

Subsequent to the completion of our study, Birnbaum and Chavez (1997) reported violations of distribution independence, which constitute evidence against the RAM model. These violations are consistent with the configural weight model of Birnbaum and Stegner (1979, Equation 10). It is called the TAX model, because weights are taken from outcomes as a function of their probabilities and redistributed to other outcomes, analogues to taxation and redistribution of wealth.

Consider gamble, $G = (x_1, p_1; x_2, p_2; \ldots; x_j, p_j; \ldots; x_i, p_i; \ldots; x_n, p_n)$, where the outcomes are ordered such that $0 < x_1 < x_2 < \ldots < x_j < x_i < \ldots x_n$ and $\Sigma\, p_i = 1$. The CW model can be written as follows:

$$U(G) = \frac{\sum_{i=1}^{n} S(p_i)u(x_i) + \sum_{i=2}^{n}\sum_{j=1}^{i-1} [u(x_i) - u(x_j)]\omega(i, j, G)}{\sum_{i=1}^{n} S(p_i)} \tag{13}$$

where U(G) is the utility of the gamble; $S(p)$ is a function of probability; $u(x)$ is the utility function of money, and $\omega(i, j, G)$ is the configural transfer of weight between outcomes x_i and x_j. Note that $x_i > x_j$, so if the configural term is negative, then the higher-valued outcome loses weight and the lower valued outcome gains this same weight. Following the simplifications of Birnbaum and Chavez (1997), $S(p) = p^{\gamma}$, $u(x) = x^{\beta}$, and the configural terms are restricted as follows:

$$\omega(i, j, G) = S(p_i)\delta/(n + 1), \text{ if } \delta \leq 0 \tag{14a}$$

$$\omega(i, j, G) = S(p_j)\delta/(n + 1), \text{ if } \delta > 0 \tag{14b}$$

where δ is the single configural parameter, n is the number of distinct outcomes in the gamble, $\delta/(n + 1)$ is the proportion of weight taken from one outcome and transferred to another. Birnbaum (in press) showed that this model can account for a variety of choice phenomena in the literature with $u(x) = x$, $S(p) = p^{.7}$, and $\delta = -1$. With these parameters, the TAX model predicts Allais paradoxes, event-splitting effects, violations of distribution independence, stochastic dominance, and cumulative independence. It makes essentially the same predictions for the present experiment as the RAM model, and it is equivalent to that model for experiments with a fixed number of outcomes of fixed probability, as in Birnbaum and McIntosh (1996).[1]

2.3 Coalescing

RDU (Equation 9) implies coalescing, for any $W(P)$ function. Suppose $x > y > z > 0$ and $p + q + r = 1$. *Proof*: If $x > y > z > 0$ and $p + q + r = 1$, then RDU$(x, p; y, q; z, r) = u(x)W(p) + u(y)[W(p + q) - W(p)] + u(z)[W(1) - W(p + q)]$, by Equation 9. If $x = y$, RDU$(x, p; x, q; z, r) = u(x)W(p) + u(x)[W(p + q) - W(p)] + u(z)[1 - W(p + q)] =$

$u(x)W(p + q) + u(z)[1 - W(p + q)]=$ RDU$(x, p + q; z, r)$, since $r = 1 - p - q$. If $y = z$, then RDU$(x, p; y, q; y, r) = u(x)W(p) + u(y)[W(p + q) - W(p)] + u(y)[1 - W(p + q)] = u(x)W(p) + u(y)[1 - W(1 - p)] =$ RDU$(x, p; y, 1 - p) =$ RDU$(x, p; y, q + r)$, since $1 - p = q + r$. In original prospect theory (Kahneman & Tversky, 1979), coalescing was proposed as a separate editing rule. In CPT, the representation (Equation 9) guarantees satisfaction of coalescing, even without the editing rule.

The RAM and TAX configural weight models (without an editing rule) do not in general satisfy coalescing. For example, if $S(p + q) < S(p) + S(q)$, an outcome with probability $p + q$ can gain weight by splitting its probability into two smaller probabilities, with the result that the sum of the weights can exceed the weight of the sum. This aspect of configural weight theory implies violations of stochastic dominance and cumulative independence (Birnbaum, 1997; in press). Luce (in press) has developed the implications of the property of coalescing to demonstrate its theoretical power.

2.4 Predictions Concerning Stochastic Dominance

Because RSDU, RDU, and CPT satisfy outcome monotonicity, transitivity, and coalescing, this class of theories must satisfy stochastic dominance in the paradigm of Section 1.3. Suppose that if $A > B$, then P$(A, B) \geq 1/2$; RDU models imply that RDU$(G+) >$ RDU$(G-)$; therefore, these models imply that stochastic dominance must be satisfied at least 50% of the time.

The configural weight model of Birnbaum and McIntosh (1996) satisfies outcome monotonicity and transitivity, but violates coalescing and therefore it can violate stochastic dominance. For example, $G+$ = ($12, .05; $14, .05; $96, .90) dominates $G-$ = ($12, .10; $90, .05; $96, .85), but certainty equivalents calculated using the parameters of Birnbaum and McIntosh (1996), are $56.99 and $63.23, for $G+$ and $G-$ respectively, a predicted violation of more than $6! If the same model and parameters hold in this study, the model predicts that violations of stochastic dominance will exceed 50%. The TAX model, with parameters of Birnbaum (in press), also violates stochastic dominance in this choice. (See Footnote 1).

In contrast, the CPT of Tversky and Kahneman (1992) implies certainty equivalents that satisfy dominance, $70.27 and $69.73, for $G+$ and $G-$, respectively. Cumulative prospect theory, or any RSDU or RDU model (with any $W(P)$), implies that scholastic dominance must be satisfied.

2.5 Predicted Violations of Cumulative Independence and Branch Independence

Whereas RDU and RSDU theories satisfy cumulative independence, the configural weight model of Birnbaum and McIntosh (1996) predicts violations of both lower and upper cumulative independence. This study was designed to investigate cases where the model and its previously published parameters would predict violations. For example, let S = (\$2, .8; \$40, .1; \$44, .1); R = (\$2, .8; \$10, .1; \$98, .1); S'' = (\$10, .8; \$44, .2); and R'' = (\$10, .9; \$98, .1). According to the Birnbaum and McIntosh (1996) model,

CWU(S) = \$10.50 $>$ CWU(R) = \$9.94 and CWU(S'') = \$17.01 $<$ CWU(R'') = \$22.12, so $S > R$ and $S'' < R''$, contrary to Equation 7a. This preference pattern will be denoted SR''.

Similarly, let S' = (\$40, .1; \$44, .1; \$110, .8); R' = (\$10, .1; \$98, .1; \$110, .8); S''' = (\$40, .2; \$98, .8); R''' = (\$10, .1; \$98, .9). The model predicts,

CWU(S') = \$68.86 $<$ CWU(R') = \$70.66 and CWU(S''') = \$73.54 $>$ CWU(R''') = \$70.76, so $S' < R'$ and $S''' > R'''$, in violation of Equation 7b, a pattern denoted, $R'S'''$.

Note also that $S > R$ and $S' < R'$, showing the SR' pattern of violation of branch independence. The TAX model makes similar predictions (Birnbaum, in press). (See footnote 1).

In contrast, CPT, with parameters of Tversky and Kahneman (1992), gives predicted certainty equivalents for the gambles as follows: CE(S) = \$11.35 $<$ CE(R) = \$17.57 and CE(S'') = \$18.27 $<$ CE(R'') = \$24.52, satisfying lower cumulative independence; CE(S') = \$82.01 $>$ CE(R') = \$77.72 and CE(S''') = \$74.52 $>$ \$70.72 = CE(R'''), satisfying upper cumulative independence. For branch independence, the CPT model predicts $S < R$ and $S' > R'$, giving the opposite pattern (RS') of violations from that predicted by the model of Birnbaum and McIntosh (1996). Equation 10a of CPT can predict the SR' pattern if $\gamma > 1$ and it can predict RS' when $\gamma < 1$. However, for any $W(P)$ function, CPT must satisfy cumulative independence and stochastic dominance.

3. Method

The judge's task was to choose between gambles. Judges circled the gamble they preferred, and judged how much they would pay to receive their preferred gamble rather that the other one.[2]

3.1 Designs

There were 4 choices that tested stochastic dominance. These choices were of the form,

$$G+ = (x, p - q; x^+, q; y, 1 - p) \text{ versus } G- = (x, p; y^-, r; y, 1 - p - r),$$

where $0 < x < x^+ < y^- < y$. The values of (x, x^+, y^-, y) in the four pairs were (\$12, \$14, \$90, \$96), (\$3, \$5, \$92, \$97), (\$6, \$8, \$91, \$99) and (\$4, \$7, \$89, \$95); the values of (p, q, r) were (.10, .05, .05), (.12, .06, .04), (.05, .03, .03), and (.02, .01, .02), respectively. Note that $G+$ stochastically dominates $G-$. These four trials testing stochastic dominance were embedded among many other choices. They were counterbalanced, so consistent choice of the gamble on the right (or left) would produce two violations and two satisfactions of dominance.

The design testing cumulative independence and branch independence was composed of 27 variations of each of the following four choices, making 108 trials:

$$S = (z, r; x, p; y, q) \text{ versus } R = (z, r; x', p; y', q);$$

$$S'' = (x', r; y, p + q) \text{ versus } R'' = (x', r + p; y', q);$$

$$S' = (x, p; y, q; z', r) \text{ versus } R' = (x', p; y', q; z', r); \text{ and}$$

$$S''' = (x, p + q; y', r) \text{ versus } R''' = (x', p; y', q + r).$$

There were 4 subdesigns with different (r, p, q): (.5, .25, .25), (.8, .1, .1), (.6, .3, .1), and (.6, .1, .3). Within each subdesign, there were 6, 7, or 8 levels of (z, x', x, y, y', z'), which were factorially combined with the four types of comparisons. All subdesigns used the following 6 levels of (z, x', x, y, y', z') = ($2, $11, $52, $56, $97, $108), ($3, $10, $48, $52, $98, $107), ($2, $11, $45, $49, $97, $107), ($2, $10, $40, $44, $98, $110), ($4, $11, $35, $39, $97, $111) and ($5, $12, $30, $34, $96, $110). For (r, p, q) = (.6, .3, .1), an additional level was added: ($3, $10, $25, $29, $98, $109); when (r, p, q) = (.6, .1, .3), two extra levels were added: ($4, $10, $61, $65, $98, $108) and ($3, $12, $56, $60, $96, $107).

The "check" design consisted of 12 choices with transparent dominance, six in which all probabilities and outcomes were the same except one outcome was better in one gamble, and six in which outcomes were the same but the probability of a better outcome was higher in one gamble. These were the same as in Birnbaum and Chavez (1997). Half of each type of check trial required selecting the gamble on the right and half on the left.

3.2 Procedure and Judges

Each booklet contained 3 pages of instructions, examples, and 10 practice trials, followed by 134 experimental choices. There were 108 choices testing cumulative independence, 12 "check" trials, 4 trials testing stochastic dominance, and 5 unlabeled warm-up trials at the beginning and end of the booklet. Choices were printed in random order, with restrictions that no two successive trials repeat the same design, subdesign, or outcomes. Half of the judges made choices in reverse order.

Judges were 100 undergraduates who completed the experiment in one hour, working at their own paces. Of these, 68 had no violations of transparent dominance in the 12 check trials (overall rate of violation was 4%). The data of 12 additional judges were not used, including 4 who did not finish and 4 with more than 2 violations in the check trials (2 had 3 violations and 2 had 4).

4. Results

4.1 Violations of Stochastic Dominance

Table 1 shows the number of judges who violated stochastic dominance in each of the four tests. For example, 73* of 100 chose $G-$ = ($12, .10; $90, .05; $96, .85) over $G+$ = ($12, .05; $14, .05; $96, .90), even though $G+$ dominates $G-$. (Asterisks designate statistical significance throughout the results. The critical value is 60 for a two-tailed, binomial sign test, with $p = .5$, $n = 100$ and $\alpha = .05$). We can separately reject the null hypothesis that stochastic dominance is satisfied 50% or more of the time in all four tests.

Predicted values in Table 1 are the difference in certainty equivalents, calculated from the previously published model and parameters of Birnbaum and McIntosh (1996). For the model and parameters of Tversky and Kahneman (1992), predicted values have the opposite sign from the data, varying from −$.45 to −$.62.

If RDU($G+$) ≥ RDU($G-$) implies that $P(G+, G-) \geq 1/2$, where $P(G+, G-)$ is the probability of choosing $G+$ over $G-$, then RDU models imply that stochastic dominance should be violated less than half the time. Instead, the percentage of violations of stochastic dominance is significantly greater than 50% in all four tests. Suppose judges "really" satisfy stochastic dominance, except they choose randomly on some trials because of inability to discriminate the gambles, boredom, lack of motivation, or mistakes. Such excuses imply that the probability of violating stochastic dominance, p, should still be less than or equal to 1/2. Because errors by different individuals are independent, these theories imply that each judge has a probability, $p \leq .5$, of violating stochastic dominance on any variation of the test. The theory that every judge has $p = .5$ is therefore a conservative limit. For 100 judges, this theory implies that the observed number of violations in one test will have a binomial distribution with expected value of $100p = 50$, and standard deviation of 5. This null hypothesis can be rejected in favor of the theory that more than half of the choices in these problems violate stochastic dominance, because the observed percentage of violations is significantly greater than 50% in all four tests. Averaged over 4 tests in this recipe, 70% of the 400 tests violated stochastic dominance. Assuming the binomial, the observed proportion of violations of .70 has a 95% confidence interval from .654 to .746.

Table 1. Violations of stochastic dominance.

G+	G−	% Viol	Mean	Pred
($12, .05; $14, .05; $96, .9)	($12, .1; $90, .05; $96, .85)	73*	$ 9.40	$6.21
($3, .06; $5, .06; $97, .88)[a]	($3, .12; $92, .04; $97, .84)	61*	$ 2.70	$6.87
($6, .02; $8, .03; $99, .95)	($6, .05; $91, .03; $99, .92)	73*	$11.44	$4.04
($4, .01; $7, .01; $95, .98)[a]	($4, .02; $89, .02; $95, .96)	73*	$ 8.39	$2.98

Notes: G+ dominates G− in all cases. In cases marked [a], the dominant gamble was on the right; otherwise, it was on the left. %Viol indicates percentage of choices violating stochastic dominance. Mean judgments indicate the average amount offered to purchase the dominated gamble, with negative numbers averaged in for choices of the dominant gamble. Positive means indicate that more money was offered on the average for the dominated (G−) rather than the dominant gamble. EV favors the dominant gambles by −$.15 to −$.40.

When data are examined for individuals, there are 65* judges with 3 or 4 violations of stochastic dominance (75% or 100% violations), compared to only 17 judges who had 25% or 0% violations, and 18 split equally. By a binomial test, significantly more judges show 75% or more violations than show 25% or fewer violations, $z = 5.30$. There were 41* judges who violated stochastic dominance on all four choices (100% violations), compared to only 9 who satisfied stochastic dominance on all four choices ($z = 4.53$). The binomial theory that all judges have $p = .5$ predicts that only 6 judges would have 4 out of 4 violations, far below the observed 41.

The binomial theory that *all* judges have a $p = .7$ rate of violations (the average in this study) would imply that only 1 judge would be expected to satisfy stochastic dominance on all 4 tests (observed = 9) and only 24 should have 4 violations out of 4 (observed = 41). The variance of the observed distribution indicates that there are individual differences in the rates of violation.

Mean judgments indicate that more money on the average was offered to get the dominated gamble rather than the dominant gamble. The largest violation occurred for ($6, .05; $91, .03; $99, .92) over ($6, .02, $8, .03, $99, .95), where the average offer was $11.44 to get the dominated gamble instead of the dominant gamble! The smallest violations occurred in the cases where $r \neq q$.

We summed the four responses by each judge and found that 80* judges offered more money for the four dominated gambles than for the dominant gambles, compared to only 18 who offered more for the dominant gambles, and two split evenly. The mean judgment of the amount offered to get the dominated gamble, averaged over judges and choices, with positive numbers reflecting violations of dominance and negative numbers representing satisfaction of dominance, was $7.98!* This empirical difference exceeds the average difference predicted by the prior model of Birnbaum and McIntosh (1996), which was $5.02.

In summary, the data show that the recipe proposed by Birnbaum (1997) to test the RAM model against the CPT/RSDU/RDU models produces 70% violations of stochastic dominance. Stochastic dominance is rarely violated in the "check trials" of transparent dominance, which are tests of outcome monotonicity or probability monotonicity (violations are only 4% in the sample of 100, and 4.7% overall). Thus, our results do not imply that people always violate stochastic dominance. Instead, they show that people systematically violate stochastic dominance in the special recipe, and rarely violate dominance in transparent tests of outcome or probability monotonicity.

4.2 Violations of Cumulative Independence

Table 2 shows the number of judges with each preference pattern in tests of lower cumulative independence: If $S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q)$ then $S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q)$. The pattern SR'' violates lower cumulative independence, yet SR'' is more frequent than RS'' (which is consistent with the property) in 19 of the 27 tests (shown in bold type), including all 8 choices with $r = .6$, $p = .1$, and $q = .3$, and 5 of 6 tests with $r = .8$, $p = q = .1$. Violations are least frequent when $r =$

Table 2. Tests of Lower Cumulative Independence: If $S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q)$ then $S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q)$.

r	p	q	(x, y)	(x′, y′)	SS″	SR″	RS″	RR″
.50	.25	.25						
			($52, $56)	($11, $97)	38	**29***	14	19
			($48, $52)	($10, $98)	42	**28***	11	19
			($45, $49)	($11, $97)	39	**25**	16	20
			($40, $44)	($10, $98)	28	19	24	29
			($35, $39)	($11, $97)	25	**22**	21	32
			($30, $34)	($12, $96)	16	11	22	51
.80	.10	.10						
			($52, $56)	($11, $97)	34	**35***	11	20
			($48, $52)	($10, $98)	30	**38***	9	23
			($45, $49)	($11, $97)	27	**29***	14	30
			($40, $44)	($10, $98)	18	**31***	12	39
			($35, $39)	($11, $97)	19	**35***	16	30
			($30, $34)	($12, $96)	17	18	19	46
.60	.30	.10						
			($52, $56)	($11, $97)	57	**16**	15	12
			($48, $52)	($10, $98)	55	17	18	10
			($45, $49)	($11, $97)	49	11	23*	17
			($40, $44)	($10, $98)	40	**26**	15	19
			($35, $39)	($11, $97)	41	21	24	14
			($30, $34)	($12, $96)	44	11	26*	19
			($25, $29)	($10, $98)	31	16	20	33
.60	.10	.30						
			($61, $65)	($10, $98)	26	**32***	14	28
			($56, $60)	($12, $96)	22	**32***	13	33
			($52, $56)	($11, $97)	16	**36***	10	38
			($48, $52)	($10, $98)	23	**29***	15	33
			($45, $49)	($11, $97)	11	**31***	14	44
			($40, $44)	($10, $98)	12	**21**	11	56
			($35, $39)	($11, $97)	10	**22***	7	61
			($30, $34)	($12, $96)	8	**21**	12	59
Totals					778	**662***	426	834

.6, $p = .3$, and $q = .1$. Asterisks in the tables indicate the 13 cases of significant deviation from the two-tailed binomial null hypothesis (with $\alpha = .05$). Overall, there are 662* choices in the *SR″* pattern compared to 426 showing the *RS″* pattern.

Each judge's choices were also summed over the 27 choices, and it was found that 64* judges had more than *SR″* than *RS″* choices, violating lower cumulative idependence, 7 showed an equal split, and only 29 showed more *RS″* choices ($z = 3.63$).

Table 3 shows the number of judges with each preference pattern in the tests of upper cumulative independence: If $S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r)$ then $S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r)$. In this case, the $R'S'''$ pattern is a violation and the $S'R'''$ pattern is consistent with the principle. In 25* of the 27 tests, the violation

Table 3. Tests of Upper Cumulative Independence: If $S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r)$ then $S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r)$.

r	p	q	(x, y)	(x', y')	$S'S'''$	$S'R'''$	$R'S'''$	$R'R'''$
.50	.25	.25						
			($52, $56)	($11, $97)	48	13	**22**	17
			($48, $52)	($10, $98)	40	9	**23***	28
			($45, $49)	($11, $97)	42	10	**29***	19
			($40, $44)	($10, $98)	36	7	**25***	32
			($35, $39)	($11, $97)	20	12	**30***	38
			($30, $34)	($12, $96)	27	7	**31***	35
.80	.10	.10						
			($52, $56)	($11, $97)	41	9	**30***	20
			($48, $52)	($10, $98)	36	10	**29***	25
			($45, $49)	($11, $97)	28	13	**30***	29
			($40, $44)	($10, $98)	25	8	**34***	33
			($35, $39)	($11, $97)	28	8	**32***	32
			($30, $34)	($12, $96)	22	9	**33***	36
.60	.30	.10						
			($52, $56)	($11, $97)	55	13	**20**	12
			($48, $52)	($10, $98)	49	12	**23**	16
			($45, $49)	($11, $97)	47	14	**22**	17
			($40, $44)	($10, $98)	46	14	**24**	16
			($35, $39)	($11, $97)	38	8	**33***	21
			($30, $34)	($12, $96)	43	8	**28***	21
			($25, $29)	($10, $98)	22	8	**34***	36
.60	.10	.30						
			($61, $65)	($10, $98)	33	18	14	35
			($56, $60)	($12, $96)	27	14	**21**	38
			($52, $56)	($11, $97)	28	18	16	38
			($48, $52)	($10, $98)	24	9	**17**	50
			($45, $49)	($11, $97)	24	13	**18**	45
			($40, $44)	($10, $98)	19	10	**22***	49
			($35, $39)	($11, $97)	20	11	**23***	46
			($30, $34)	($12, $96)	15	8	**15**	62
Totals					883	293	**678***	846

pattern is more frequent than the consistent pattern (bold type), including 16 that are individually significant (asterisks). Overall, 678* choices were of the $R'S'''$ pattern against only 293 of the $S'R'''$ type.

Of the 100 individual judges, 67* had more $R'S'''$ choices than $S'R'''$, in violation of upper cumulative independence, 11 showed an equal split, and only 22 had more $S'R'''$ choices.

In separate analyses, there were 17, 29, and 22 judges whose data showed more choices compatible with than in violation of stochastic dominance, lower cumulative independence, and upper cumulative independence, respectively. Based on these individual counts alone, there might be a minority of as many as 17 judges whose data are more consistent (than in violation) with all three properties. However, there were only 7 judges whose preference shifts were in better agreement with both cumulative independence properties than in violation. Of these 7, only 2 had fewer than 2 violations of stochastic dominance. Thus, only 2 judges have data more consistent with all three of these implications of RDU than in violation.

4.3 Violations of Branch Independence

Table 4 summarizes tests of noncomonotonic branch independence. Whereas EU and SWU theories imply no violations of branch independence, both RDU and CWT allow violations. If judges were to cancel common components prior to choice, then common

Table 4. Tests of Noncomonotonic Branch Independence: $S = (z, r; x, p; y, q) > R = (z, r; x', p; y', q)$ if and only if $S' = (x, p; y, q; z', r) > R' = (x', p; y', q; z', r)$.

r	p	q	(x, y)	(x', y')	SS'	SR'	RS'	RR'
.50	.25	.25						
			($52, $56)	($11, $97)	48	**19**	13	20
			($48, $52)	($10, $98)	39	**31***	10	20
			($45, $49)	($11, $97)	44	**20***	8	28
			($40, $44)	($10, $98)	31	**16**	12	41
			($35, $39)	($11, $97)	22	**25***	10	43
			($30, $34)	($12, $96)	15	12	19	54
.80	.10	.10						
			($52, $56)	($11, $97)	41	**28***	9	22
			($48, $52)	($10, $98)	38	**30***	8	24
			($45, $49)	($11, $97)	27	**29***	14	30
			($40, $44)	($10, $98)	23	**26***	10	41
			($35, $39)	($11, $97)	28	**26***	8	38
			($30, $34)	($12, $96)	15	**20**	16	49
.60	.30	.10						
			($52, $56)	($11, $97)	55	**18**	13	14
			($48, $52)	($10, $98)	55	**17**	6	22
			($45, $49)	($11, $97)	47	13	14	26
			($40, $44)	($10, $98)	47	**19**	13	21
			($35, $39)	($11, $97)	40	**22***	6	32
			($30, $34)	($12, $96)	37	**18**	14	31
			($25, $29)	($10, $98)	21	**26***	9	44
.60	.10	.30						
			($61, $65)	($10, $98)	39	**19**	12	30
			($56, $60)	($12, $96)	31	**23***	10	36
			($52, $56)	($11, $97)	32	**20**	14	34
			($48, $52)	($10, $98)	21	**31***	12	36
			($45, $49)	($11, $97)	20	**22**	17	41
			($40, $44)	($10, $98)	16	**17**	13	54
			($35, $39)	($11, $97)	18	**14**	13	55
			($30, $34)	($12, $96)	11	**18**	12	59
Totals					861	**579***	315	945

branches should have no effect, and there should be no systematic violations of branch independence. Apart from such editing, the inverse-S weighting function fit by Tversky and Kahneman (1992) implies that there should be more RS' than SR' violations. The model of Birnbaum and McIntosh (1996) predicts the opposite pattern; namely, more SR' violations. In 25* of the 27 tests, SR' is more frequent than RS' (bold type in Table 4), including 12 that are individually significant (asterisks); overall, there were 579* choices of the SR' violation and only 315 of the RS' type. Of 100 judges, 65* had more SR' violations than RS' violations, 11 showed an equal split, and only 24 showed the opposite pattern. These data replicate and extend previous findings.

4.4 Fit of CPT and TAX Models

In preceding analyses, we compared a specific model and its parameters (the RAM model fit by Birnbaum and McIntosh, 1996) against a general class of models, the RSDU and RDU models with any parameters. Analyses showed that significantly more individuals showed the violations predicted by the CW RAM model than had choices consistent with any RDU model. Those analyses required no estimation of parameters from the present data.

We now change tacks and fit data to specific models, to address the following questions: How well do previously published models (and parameters) predict (1) the average data and (2) individual data in the present study? When CPT and CWT models are allowed to fit parameters to the same data, how well do they fit (3) average data and (4) individual data? (5) Can CPT be saved by allowing its parameters to depend on the number of outcomes in each gamble?

The configural weight, TAX model has the same number of parameters as CPT; it makes predictions for this study that are nearly identical to the RAM model, and it can account for violations of distribution independence (Birnbaum & Chavez, 1997), which violate the RAM model. Therefore, our contest of fit is between the TAX and CPT models.

Both CPT and TAX models were fit with the following equations:

$$D(R, L) = a[U(R) - U(L)] \tag{15}$$

$$P(R, L) = F[b(U(R) - U(L))] \tag{16}$$

$$u(x) = x^{\beta} \tag{17}$$

where $D(R, L)$ is the predicted judgment of the strength of preference (amount offered to receive) gamble R instead of gamble L (including the sign that indicates the direction of choice); $P(R, L)$ is the predicted probability of choosing R over L; a and b are scaling constants; F is the logistic function, $F[x] = 1/(1 + \exp[-x])$; $u(x)$ is the utility of outcome, x; β is the exponent.

For CPT, $U(R)$ and $U(L)$ are the utilities of the gambles as given in Equations 9 and 10b. In Equation 10b, $W(P)$ is characterized by two constants, c and γ. This CPT model has five parameters: a, b, c, γ, and β.

For the configural weight TAX model, $U(R)$ and $U(L)$ were computed by Equations 13 and 14a–b instead of 9 and 10b. This model also have five parameters to estimate, a, b, δ, γ, and β. The parameter δ in the TAX model is analogous to c in CPT model.

Models were fit to the 112 choices and judgments (excluding the check trials) using computer programs of Birnbaum and Chavez (1997) to minimize the following compromise:

$$\mathcal{B} = h \sum_{j=1}^{m} (D_j - \hat{D}_j)^2 - (1 - h)\log[\prod_{j=1}^{m} P(C_j)] \tag{18}$$

where D_j and $\hat{D}_j$ the observed and predicted judgment for each choice, which can be either positive or negative, depending on the direction of choice; C_j is the observed choice (gamble on the right or left), $P(C_j)$ is the probability of that choice given the model [i.e., $P(R, L)$ or $1 - P(R, L)$ from Equation 16]. The term, $\text{SUM} = \sum_{j=1}^{m} (D_j - \hat{D}_j)^2$, is the familiar sum of squared deviations between observed and predicted judgments for the $m = 112$ judgments, and the term, $\mathcal{L} = -\log [\prod_{j=1}^{m} P(C_j)]$, is the (negative log) likelihood of the 112 observed choices given the model; h and $1 - h$ are relative weights of the two sub-indices of fit. This compromise loss function requires the model to account for both strength of preference judgments and choices using the same parameters. FORTRAN programs, CPTFIT and TAXFIT, used Chandler's (1969) subroutine, STEPIT, to minimize $\mathcal{B}$ (See footnote 1 for further details).

A series of analyses addressed the five questions stated above. The first assumed previously published parameters for CPT and TAX models, and estimated only the scaling parameters, a and b from these data (a only affects the sum of squared deviations, SUM, and b only affects the negative log likelihood of choices, $\mathcal{L}$). For the prior model of CPT, the values of $\gamma = .61$, $\beta = .88$, and $c = .72$ were taken from the fit to Tversky and Kahneman (1992). For the prior TAX model, the values of $\gamma = .70$, $b = 1$, and $\delta = -1$ were taken from Birnbaum (in press). Fitting means, the TAX model has SUM = 1715.1 compared to 3362.2 for CPT, and the TAX model has a negative log likelihood, $\mathcal{L} = 49.94$ compared to 66.81 for CPT. Fitting the data of individuals, the mean value of $\mathcal{L}$ was significantly lower for the TAX model than for CPT, $t(99) = 4.28$; the prior TAX model fit the choices of 63* judges better than the prior CPT model, and 37 were better fit by CPT.

When all 5 parameters were estimated from the data, the TAX model again fit the mean judgments better. With $h = .01$, the TAX model had SUM = 1189 and $\mathcal{L} = 41.1$ compared to SUM = 2300 and $\mathcal{L} = 53.7$ for CPT. See the upper portion of Table 5.

Similar results were obtained when the models were fit to individuals. Median parameter estimates for individuals are given in the lower portion of Table 5. The TAX model fit

Table 5. Parameter Estimates and Indices of Fit of CPT and TAX models.

Fit to Mean Judgments

	Parameters					Index of Fit
Model	γ	β	δ or *c*	*a*	*b*	−log(Π*P*)
TAX (5)	0.891	0.728	−0.536	3.617	1.744	41.10
TAX (4)	0.973	(1.0)	−0.837	0.867	0.405	44.07
CPT (5)	0.922	0.599	0.543	5.852	2.346	53.74

Fit to Individuals (Medians of Individual Parameters and Index)

	Parameters					Index of Fit
Model	γ	β	δ or *c*	*a*	*b*	−log(Π*P*)
TAX (5)	0.790	.414	−.954	9.549	1.010	58.96
TAX (4)	0.739	(1.0)	−1.09	0.102	0.714	61.416
CPT (5)	0.956	0.956	0.273	1.287	0.249	64.76

Note: Each entry is the median parameter estimate or median subindex of fit ($\mathscr{L} = -\log(\Pi P)$ refers to negative log likelihood. For the null model will all probabilities = 0.5, this index would be 77.63. In TAX(4), β is fixed to 1.

better on both subindices of fit for 52* judges, compared to only 16 whose data were better fit by CPT on both subindices ($z = 4.37$). The average SUM was significantly lower for the TAX model, $t(99) = 2.06$. The TAX model predicted choices better than CPT ($\mathscr{L}$ was smaller) for 81* of judges against only 19 whose choices were better predicted by CPT. The mean $\mathscr{L}$ is also significantly lower for TAX than CPT, $t(99) = 5.21$. For the 5-parameter TAX model, 89* judges had $\delta < 0$; 60 had $\gamma < 1$, and 71* had $\beta < 1$. When β is fixed to 1, the TAX(4) model fits only slightly worse than TAX(5), and this simpler TAX model still fits better than the CPT(5) model.

Both TAX models gave good fits to violations of cumulative independence and stochastic dominance, which the CPT model can not do with any parameters. For example, the TAX model with β fixed to 1 gives mean predictions (averaged over judges) for the 4 violations of stochastic dominance of $10.79, $11.83, $12.00, and $11.53 for the gambles in Table 1, respectively. Predictions for any CPT model for these four choices are all negative.

One interpretation of violations of cumulative independence within RDU is to theorize that the weighting function depends on the number of outcomes in the gamble. Tversky and Kahneman (1992, p. 317) gave a “pessimistic assessment” that their model (estimated from two-outcome gambles) might not generalize to other values of n. Derivations show that the SR' pattern of violations of branch independence (with $n = 3$) implies opposite relations among ratios of weights (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997) from those estimated by Tversky and Kahneman (for $n = 2$). It can be shown that the preference order we observe for $n = 2$, $R'' \succ S''$ and $S''' \succ R'''$, would be consistent with the Tversky and Kahneman weighting function, with $\gamma < 1$. CPT might therefore improve its fit by allowing its parameters to depend on n.

When data were fit to this more general CPT(10) model, with 10 parameters for each judge, $W(P)$ functions were systematically different for $n = 2$ and $n = 3$. Median estimated parameters, were as follows: $\gamma = .831$ and 1.179, $c = .372$ and .119; $\beta = .819$ and 1.525, $a = 1.54$ and .058, $b = .539$ and .078; SUM = 9420 and 7302; $\mathscr{L} = 27.6$ and 28.93, for $n = 2$ and $n = 3$ (excluding the 4 choices testing stochastic dominance), respectively. (These are significant. For example, 69* judges had higher values of γ when $n = 3$ than when $n = 2$; mean γ was also significantly higher, $t(99) = 2.10$. Note that the median $\gamma < 1$ for $n = 2$, consistent with previous estimates (Tversky & Kahneman, 1992; Wu & Gonzalez, 1996), and median $\gamma > 1$ for $n = 3$, consistent with Birnbaum and McIntosh (1996) and Birnbaum and Chavez (1997).

Even when CPT is allowed 10 parameters, CPT still must satisfy stochastic dominance. Median parameters predict the following proportions for violations of stochastic dominance (CPT(10) for $n = 3$): 0.23, 0.32, 0.22, and 0.33. These predictions are significantly (and substantially) less than corresponding observed proportions (Table 1): .73, .61, .73, and .73. In contrast, median parameters from TAX(5) yield predictions of .68, .72, .72, and .73, which are much more accurate predictions of the observed data. Thus, even when CPT is allowed twice as many parameters, it gives a markedly worse fit to the data than the TAX model.

5. Discussion

5.1 Violations of Stochastic Dominance and Cumulative Independence

Violations of stochastic dominance and cumulative independence are to the class of RDU, RSDU, and CPT theories as the Allais paradoxes are to EU: There is no $W(P)$ function and $u(x)$ function that can explain these phenomena with Equation 9.

As predicted by the equations and parameters of Birnbaum and McIntosh (1996), Birnbaum's (1997) recipe creates systematic violations of stochastic dominance. The recipe produces an amazingly high rate of 70% violations. Apparently, a gamble with two high outcomes and one low one can seem bettr than a gamble with two low outcomes and one high one, even when the gamble with two low outcomes stochastically dominates the gamble with two high outcomes.

Violations of stochastic dominance and cumulative independence violate basic tenants of RDU and RSDU. These violations go beyond testing implications of a particular weighting function from one study to the next. Cumulative independence creates a contradiction in the RDU weighting function within the same experiment. This contraindication is illuminated in our fits of the general CPT model with different parameters for the cases of $n = 2$ and $n = 3$, which yielded median estimates of $\gamma < 1$ and $\gamma > 1$, respectively.

Thus, this study finds empirical confirmation of three striking predictions of the model of Birnbaum and McIntosh (1996) that refute the class of RSDU/CPT models—violations of stochastic dominance, upper cumulative independence, and lower cumulative independence. Such violations are inconsistent with any RDU or RSDU theory. The configural weight models also successfully predict a fourth aspect of the data, the SR′ pattern of violation of branch independence, which contradicts the inverse-S weighting function estimated by Tversky and Kahneman (1992).

5.2 Possible Sources of Violations

If violations of stochastic dominance were observed only in choice, then one might attempt to explain them using some theory of the comparison process, such as editing and cancellation of similar branches between gambles (cf. Kahneman & Tversky, 1979; Wu, 1994; Leland, 1994). However, Birnbaum and Yeary (1997) presented the 8 gambles in Table 1 for judgment and found that mean judgments of the dominated gamble in each case were higher than the mean judgments of the corresponding dominant gamble. These violations were observed on all four tests in both the buyer's and seller's points of view. Thus, the present findings for choice are corroborated using two judgment procedures. Therefore, the origin of violations of stochastic dominance appears to be due to processes of combination, rather than of choice.

Because stochastic dominance in this paradigm (Section 1.3) can be viewed as a combination of outcome monotonicity, transitivity, and coalescing, violations might be due to violations of one or more of these simpler principles. Similarly, cumulative independence can be deduced from the above three assumptions plus comonotonic independence. According to the configural weight theory, these violations are due to violations of coalescing.

Monotonicity Violations? Although violations of monotonicity have been observed in judgments comparing gambles with and without a zero ($0) outcome (Birnbaum, 1992; 1997; Mellers, Weiss, & Birnbaum, 1992), they have not been reported in direct choices (Birnbaum & Sutton, 1992) nor in cases with all positive outcomes. Furthermore, it seems unlikely that monotonicity is the problem because those 4 who violated outcome monotonicity or probability monotonicity more than twice in 12 check trials were excluded. Because violations of these properties are so infrequent (68 judges had no violations), it seems unlikely that violations of stochastic dominance or cumulative independence can be attributed to violations of outcome or probability monotonicity. The rarity of these violations also suggests that our judges were not lacking in attention.

"Check" trials have been used in previous research to ensure that judges have at least superficial understanding of the task, and to identify those who might be confused, careless, or inattentive. If a judge were choosing randomly, then the probability of making 2 or fewer violations of transparent dominance in 12 check trials is .02. A person who always chose the gamble on the right would have 6 violations. To study if the check procedure itself affects our results, we tested 38 additional undergraduates with a shorter booklet without check trials (and no one was excluded). Results replicated the main study. For example, 13 of these 38 judges had 4 violations of stochastic dominance out of 4 tests, 9 had 3, 10 had 2, 3 had 1 violation, and 3 had no violations. The overall rate of violation is therefore 67.1% in this group, significantly greater than 50% and not significantly different from 70%, the rate found in our main experiment.

Transitivity. Birnbaum, Patton, and Lott (in press) tested the possibility that the present violation of stochastic dominance might induce systematic violations of transitivity, if judges detect and conform to stochastic dominance in simpler choices. For example, suppose judges recognize that $G+ = (\$12, .05; \$14, .05; \$96, .9)$ dominates $G_0 = (\$12, .1; \$96, .9)$, that G_0 dominates $G- = (\$12, .10; \$90, .05; \$96, .85)$, but they still choose $G-$ over $G+$. If so, there would be a violation of transitivity. With a new sample of 110 judges and five new variations of the choice recipe, Birnbaum, et al. (in press) found that stochastic dominance was violated in 73.6% of choices between $G-$ and $G+$. In about one third of the choice triads where stochastic dominance was violated in the choice of $G-$ over $G+$, stochastic dominance was satisfied in both choices of $G+$ and $G-$ against G_0, producing a violation of transitivity. However, in the other two thirds of the cases, there was at least one other violation of stochastic dominance, preserving transitivity. The conditional probability of violating stochastic dominance, given satisfaction of transitivity, was .67, suggesting that transitivity is not the culprit producing violations of stochastic dominance. Nevertheless, the data suggested that violations of weak stochastic transitivity could be attributed to an editing mechanism that detects dominance in choices against G_0 and does not detect dominance in the comparison of $G-$ and $G+$.

Coalescing. The RAM and TAX models obey outcome monotonicity (for positive outcomes), restricted comonotonic independence, and transitivity; however, both of these models violate coalescing. When the TAX model is fit to the present data, it fits better than the CPT models, although both models use the same number of parameters. The success of these configural weight models in predicting the violations of stochastic dominance, lower cumulative independence, and upper cumulative independence is therefore consistent with the hypothesis that violation of coalescing is the key to explaining violations of these properties.

The term "event-splitting" effects (Humphrey, 1995; Starmer & Sugden, 1993) refers to violation of a combination of coalescing and transitivity. Assuming transitivity, event-splitting effects are violations of coalescing. Starmer and Sugden (1993) and Humphrey (1995) explain event-splitting effects by means of a SWU model, sometimes termed "stripped" prospect theory (in which editing rules that imply coalescing have been removed). These SWU models can be tested against CWT models by the following property of event-splitting independence.

Event-splitting Independence: Event-splitting independence asserts that if splitting an event has an effect, it should have the same effect whenever the same event (and probability) is split, independent of the rank and value of the outcome split, as long as the sign of the split outcome is the same. For example, suppose all outcomes are positive and $p + q + r = 1$; *event-splitting independence* implies,

$$(x, p; x, q; z, r) \succ (x, p + q; z, r) \text{ iff } (x', r; y, p; y, q) \succ (x', r; y, p + q). \tag{19}$$

SWU models that are represented by $\text{SWU}(G) = \Sigma\, w(p)u(x)$ satisfy event-splitting independence. *Proof:* According to SWU, $(x, p; x, q; z, r) \succ (x, p + q; z, r)$ iff $w(p)u(x) + w(q)u(x) + w(r)u(z) > w(p + q)u(x) + w(r)u(z)$, iff $w(p) + w(q) > w(p + q)$. Since we can multiply both sides by $u(y)$ and add $w(r)u(x')$ to both sides, it follows that $w(r)u(x')$

$+ w(p)u(y) + w(q)u(y) > w(r)u(x') + w(p + q)u(y)$, which holds iff $(x', r; y, p; y, q) > (x', r; y, p + q)$.

The configural weight models [and models of the form, $\text{SWAU}(G) = \Sigma\, w(p)u(x)/\Sigma\, w(p)$] violate event-splitting independence. Because the configural weight RAM and TAX models are averaging models, splitting the probability of a positive outcome can either increase or decrease the value of the gamble, depending on whether the branch split was associated with the highest or lowest outcome in the gamble. For example, the model and parameters of Birnbaum and McIntosh (1996) yield the following predicted certainty equivalents:

CWU($12, .9; $96, .05; $96, .05) = $24.18 > CWU($12, .9; $96, .1) = $23.60; however,
CWU($12, .05; $12, .05; $96, .9) = $55.6 < CWU($12, .1; $96, .9) = $70.0,

in violation of event-splitting independence.

Changes in procedure. Some have suggested that other procedures might alter our results: (1) Perhaps with financial incentives, people might be more motivated to conform to stochastic dominance. (2) Perhaps with fewer choices, judges wouldn't get bored or careless. (3) Perhaps with a different display format, results would be different. (4) If choices were all on the same page, perhaps judges would make their choices consistent. (5) Perhaps the procedure of asking judges to evaluate the differences between gambles affects their choices.

Birnbaum (1998) applied these suggestions and replicated our results. There were 14 choices, all printed on a single sheet, with gambles displayed in the format used by Kahneman and Tversky (1979) and others. Thirty-one undergraduates were given a chance to play their chosen gamble on one randomly selected trial for face value, half of face value, or double face value. Prizes could be a high as $220, and people seemed excited. Results were quite compatible with those reported here. For example, the new study found 76%* violations of stochastic dominance with two variations of the recipe used here, significantly greater than 50% but not significantly different from the rate of 70% found here.

Birnbaum (1998) also tested for reversals of preference due to event-splitting, predicted by configural weight models. For example, *G*+ = ($12, .05; $14, .05; $96, .90) versus *G*− = ($12, .10; $90, .05; $96, .85), and *GS*+ = ($12, .05; $14, .05; $96, .05; $96, .85) versus *GS*− = ($12, .05; $12, .05; $90, .05; $96, .85). Note that *GS*+ versus *GS*− is the same choice as *G*+ versus *G*−, except for coalescing. However, the split versions reduce stochastic dominance to outcome monotonicity. Birnbaum (1998) found that 29* of 31 judges (90%) satisfied monotonicity by preferring *GS*+ over *GS*−, and that 23* of these violated stochastic dominance by preferring *G*− over *G*+; 21* of 31 (67%) reversed preferences from *G*− over *G*+ to *GS*+ over *GS*−, and none switched in the other direction. Violations of cumulative independence were also replicated. Apparently, our findings are not sensitive to financial incentives or the other changes in procedure.

Comonotonic Independence. Violations of cumulative independence may be due to violation of transitivity, monotonicity, coalescing, or comonotonic branch independence, which has been sustained in tests of judgment and choice, but which deserves more strenuous tests than it has yet received (Birnbaum & McIntosh, 1996; Birnbaum & Beeghley, 1997; Wakker, et al., 1994).

Wu (1994) reported violations of tail independence, with a test that is similar to but distinct from our test of upper cumulative independence. Tail independence can be derived from comonotonic branch independence, coalescing, and transitivity. Wu (1994) noted that violations are inconsistent with CPT. Wu's explanation also assumes that the key to the violations is coalescing, though his theory is different. Wu theorized that subjects edit and cancel common components between gambles, when the common components are transparent (not coalesced). Wu's (1994) editing theory implies that choices should satisfy branch independence.

5.3 Violations of Branch Independence

The present results replicate the pattern of violations of noncomonotonic branch independence observed in previous research; namely, the SR' pattern is more frequent than the RS' pattern. The present data also show that this pattern is found in choices between three-outcome gambles in which all three outcomes have different probabilities. The violations of branch independence observed here [and by Birnbaum and McIntosh (1996), Birnbaum and Beeghley (1997), Birnbaum and Veira (1998), Birnbaum and Chavez (1997), and Weber and Kirsner (1997)] are opposite those predicted by the inverse-S weighting function used in the cumulative prospect model.

5.4 Allais Common Consequence Paradox

The common consequence paradox of Allais (1953/1979), which violates EU theory, can be interpreted as a violation of Allais independence, which can be deduced from branch independence, coalescing, and transitivity (Birnbaum, in press). Because the paradox combines these properties, it can be explained by theories that satisfy branch independence but violate coalescing (and stochastic dominance), such as SWU and original prospect theory (Edwards, 1954; Karmarker, 1978; Kahneman & Tversky, 1979). However, the Allais paradox can also be explained by theories that satisfy coalescing (and stochastic dominance) but violate branch independence, such as RDU, RSDU, and CPT (Luce, 1992; Quiggin, 1985; Tversky & Kahneman, 1992; Wu & Gonzalez, 1996). Violations of branch and distribution independence (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997; the present data) show that SWU models and original prospect theory can be rejected. The present results and event-splitting effects (Starmer & Sugden, 1993; Humphrey, 1995) suggest that coalescing is violated, allowing rejection of the RDU, RSDU and CPT models. We are left with configural weight theories, which violate branch independence and coalescing.

5.5 Testing Among Rival Configural Weight Models

As noted by Birnbaum and McIntosh (1996), the configural weight RAM model can fit the results of Tversky and Kahneman (1992), Wu and Gonzalez (1996), and Birnbaum and McIntosh (1996) without changing parameters. It can also explain the Allais paradoxes (Birnbaum, in press). Although the RAM model with its previously published parameters has spectacular success in predicting the present results, the RAM model cannot account for violations of *distribution independence* (Birnbaum & Chavez, 1997). The preference order SR' is more frequent than RS', when the distribution of common outcomes was changed as follows:

$$(z, .55; x, .2; y, .2; \text{v}, .05) = \text{S} \succ R = (z, .55; x', .2; y', .2; \text{v}, .05) \text{ and}$$

$$(z, .05; x, .2; y, .2; \text{v}, .55) = \text{S}' \prec R' = (z, .05; x', .2; y', .2; \text{v}, .55).$$

This pattern of violations was observed in twelve variations of the above test (Birnbaum & Chavez, 1997). It is also opposite that predicted by the inverse-S weighting function.

The TAX model can explain this pattern of violations of distribution independence with the same parameters as needed to explain violations of Allais independence, common ratio independence, branch independence, stochastic dominance, and cumulative independence (Birnbaum, in press). The simple version of the TAX model (the restrictions of Equations 14a-b) imply that configural weights of equally likely outcomes will be a monotonic function of the rank, unlike RDU and unlike the RAM model. Studies of judgment have concluded that judges in the seller's point of view apply the most configural weight to the middle outcome (Birnbaum & Beeghley, 1997; Birnbaum & Veira, 1998).

Another configural model that can account for these phenomena is minimum asymmetric loss theory (see Birnbaum, et al., 1992; Birnbaum & McIntosh, 1996, Appendix A). This theory attributes configural weighting to asymmetric costs of over- or underestimating the value of an uncertain or risky good (Birnbaum et al., 1992). This idea was proposed by Birnbaum and Stegner (1979) to explain the different configural weighting patterns between buyer's and seller's prices. A rival to the configural weighting theory of viewpoint effects was the notion of loss aversion suggested to explain the "endowment effect," in the economics literature of the 1980s. Unfortunately, studies in the economics literature, reviewed in Kahneman, Knetsch, and Thaler (1991), did not compare loss aversion to configural weighting.

Birnbaum and Zimmermann (1998) compare implications of configural weighting theory, loss aversion, and anchoring and adjustment theories of buying and selling prices (WTP and WTA). The theory that configural weights depend on viewpoint (buyer vs. seller) explains the phenomena reported by Birnbaum and Stegner (1979), Birnbaum and Beeghley (1997), Birnbaum and Sutton (1992), Birnbaum and Veira (1998) and Birnbaum and Zimmermann (1998); however, two interpretations of the notion of loss aversion coupled with CPT fail to explain the major features of judgment data. Anchoring and adjustment theory (that sellers anchor on the highest outcome and buyers anchor on the lowest outcome and adjust to other outcomes) also fails to account for the data.

The TAX and RAM models imply comonotonic independence, whereas minimum loss theory can violate comonotonic independence (see Appendix A of Birnbaum and McIntosh, 1996). Minimum loss theory does satisfy regional comonotonic independence. As noted by Birnbaum (1997), the TAX model also implies a systematic pattern of violation of asymptotic independence, unlike the RAM model. These three properties, comonotonic independence, monotonicity of configural weights as a function of rank, and violation of asymptotic independence are testable implications of the TAX model that deserve careful testing in future research.

Because CWT models were successful in predicting where to find violations of stochastic dominance, it is reasonable to ask if there are cases in which these models predict violations and yet judges satisfy dominance. SWU models, such as original prospect theory without the editing principles, imply violations of transparent dominance (Birnbaum, in press). Although SWU predicts violations in such cases, neither RAM nor TAX models violate transparent dominance. There may be other cases in which the model predicts violations and judges do not violate dominance. One place to begin a search is in "translucent" comparisons of $G+$ and $G-$ aginast G_0, where Birnbaum et al. (in press) found fewer violations than predicted by any transitive model.

5.6 Editing Principles Need Revision

Both versions of prospect theory (Kahneman & Tversky, 1979; Leland, 1994; Tversky & Kahneman, 1986; 1992; Wu, 1994) proposed editing principles, including two (cancellation and combination) that are addressed by the present study. *Combination* assumes that judges simplify gambles by combining equal outcomes within gambles. This principle implies coalescing and rules out "event-splitting" effects (Humphrey, 1995; Starmer & Sugden, 1993).

The *cancellation* principle assumes that judges eliminate components common to both gambles before choice. If judges obeyed this editing principle, they should not violate branch independence or distribution independence in any systematic fashion. Systematic violations of branch independence, distribution independence, and coalescing (including its consequences of stochastic dominance and cumulative independence) contradict the principles of cancellation and combination.

Stevenson, Busemeyer, and Naylor (1991) noted that the same editing principles can make drastically different predictions depending on the order in which they are applied. For example, if judges coalesce nearly equal outcomes before comparing gambles, they should obey stochastic dominance in choice. However, if judges cancel equal outcomes with nearly equal probabilities (before coalescing within gambles), then they might violate stochastic dominance.

Although editing principles have some flexibility, it is difficult to see how they would explain violations of stochastic dominance in judgment, where gambles are presented on separate trials, and presumably cancellation of branches between gambles is not possible.

The fact that similar violations of branch independence and stochastic dominance are observed in both judgment and choice suggests that editing will have a difficult time explaining these phenomena.

6. Conclusions

In summary, the present study shows systematic violations of stochastic dominance, lower cumulative independence, and upper cumulative independence. These violations are inconsistent with RSDU theories (Luce & Fishburn, 1991; 1995) including CPT (Tversky & Kahneman, 1992) and rank-dependent utility (Quiggin, 1982). The present data also replicate the pattern of violations of branch independence that is inconsistent with the inverse-S weighting function. The majority pattern of violations of all four properties agree with predictions made in advance of the experiment by the configural weight, RAM model of Birnbaum and McIntosh (1996). The configural weight TAX model fits the data of most individuals better than CPT even though it has the same number of parameters, and it even outperforms the CPT model when CPT is allowed more parameters.

Acknowledgements

Support was received from National Science Foundation Grant, SBR-9410572, to the first author through California State University, Fullerton. Thanks are due R. Duncan Luce and Peter Wakker for their helpful suggestions on an earlier draft, and we thank Teresa Martin, Nicole Patton, Melissa Lott, and Melissa A. Birnbaum for assistance and advice.

Notes

1. Programs to compute the predictions of CPT, RAM, and TAX models can be accessed from URL http://psych.fullerton.edu/mbirnbaum/programs.htm.
2. A detailed description of the procedure (including a copy of the materials with instructions) can be obtained from URL http://psych.fullerton.edu/mbirnbaum/BN.htm.

References

Birnbaum, M.H. (1973). Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology* 99, 395–399.

Birnbaum, M.H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology,* 102, 543–561.

Birnbaum, M.H. (1992). Violations of monotonicity and contextual effects in choice-based certainty equivalents. *Psychological Science,* 3, 310–314.

Birnbaum, M.H. (1997). Violations of monotonicity in judgment and decision making. In A.A.J. Marley (Eds.), *Choice, decision and measurement: Essays in honor of R. Duncan Luce* (pp.73–100). Mahwah, NJ: Erlbaum.

Birnbaum, M.H. (in press). Paradoxes of Allais, stochastic dominance, and decision weights. In J.C. Shanteau, Mellers, B.A., & Schum, D. (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards*. Norwell, MA: Kluwer Academic Publishers.

Birnbaum, M.H., & Beeghley, D. (1997). Violations of branch independence in judgments of the value of gambles. *Psychological Science*, 8, 87–94.

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

Birnbaum, M.H., Coffey, G., Mellers, B.A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 331–346.

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

Birnbaum, M.H., Patton, J.N., & Lott, M.K. (in press). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes*, 74, 000–000.

Birnaum, M.H., & Stegner, S.E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37, 48–74.

Birnbaum, M.H., & Sutton, S.E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes*, 52, 183–215.

Birnbaum, M.H., Thompson, L.A., & Bean, D.J. (1997). Testing interval independence versus configural weighting using judgments of strength of preference. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 939–947.

Birnbaum, M.H., & Veira, R. (1998). Configural weighting in judgments of two- and four-outcome gambles. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 216–226.

Birnbaum, M.H., & Yeary, S. (1997). Tests of cumulative independence, coalescing, and stochastic dominance in judgments of buying and selling prices of gambles. Working paper.

Birnbaum, M.H. & Zimmermann, J.M. (1998). Buying and selling prices of investments: Configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes*, 74, 145–187.

Chandler, J.P. (1969). STEPIT: Finds local minima of a smooth function of several parameters (CPA 312). *Behavioral Science*, 14, 81–82.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, 380–417.

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

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

Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1991). Experimental tests of the endowment effect and the coarse theorem. In R.H. Thaler (Eds.), *Quasi rational economics* (pp. 167–188.). New York: Russel Sage Foundation.

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

Lattimore, P.K., Baker, J.R., & Witte, A.D. (1992). The influence of probability on risky choice. *Journal of Economic Behavior and Organization*, 17, 377–400.

Leland, J.W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, 9, 151–172.

Lopes, L. (1990). Re-modeling risk aversion: A comparison of Bernoullian and rank dependent value approaches. In G.M. v. Furstenberg (Eds.), *Acting under uncertainty* (pp. 267–299). Boston: Kluwer.

Luce, R.D. (1988). Rank-dependent, subjective expected-utility representations. *Journal of Risk and Uncertainty*, 1, 305–332.

Luce, R.D. (1992). Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty*, 5, 5–27.

Luce, R.D. (1996). When four distinct ways to measure utility are the same. *Journal of Mathematical Psychology*, 40, 297–317.

Luce, R.D. (submitted). Coalescing, event commutatively, and theories of utility. *Submitted for publication*, 00, 000–000.

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

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

Luce, R.D., & von Winterfeldt, D. (1994). What common ground exists for descriptive, prescriptive, and normative utility theories? *Management Science*, 40(2), 263–279.

Mellers, B.A., Weiss, R., & Birnbaum, M.H. (1992). Violations of dominance in pricing judgments. *Journal of Risk and Uncertainty*, 5, 73–90.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3, 324–345.

Quiggin, J. (1985). Subjective utility, anticipated utility, and the Allais paradox. *Organizational Behavior and Human Decision Processes*, 35, 94–101.

Savage, L.J. (1954). Historical and critical comments on utility. In L.J. Savage (Eds.), *The foundations of statistics* (pp. 91–104). New York: Wiley.

Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty*, 6, 235–254.

Stevenson, M.K., Busemeyer, J.R., & Naylor, J.C. (1991). Judgment and decision-making theory. In M. Dunnette & L.M. Hough (Eds.), *New handbook of industrial-organizational psychology* (pp. 283–374). Palo Alto, CA: Consulting Psychologist Press.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, S251–S278.

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

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

Viscusi, K.W. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, 2, 235–264.

Wakker, P. (1996). The sure-thing principle and the comonotonic sure-thing principle: An axiomatic analysis. *Journal of Mathematical Economics*, 25, 213–227.

Wakker, P., Erev, I., & Weber, E.U. (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty*, 9, 195–230.

Wakker, P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, 7, 147–176.

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

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty*, 9, 39–60.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42, 1676–1690.