

Evidence against Rank-Dependent Utility Theories: Tests of Cumulative Independence, Interval Independence, Stochastic Dominance, and Transitivity

Michael H. Birnbaum

California State University, Fullerton, and Decision Research Center

and

Jamie N. Patton and Melissa K. Lott

California State University, Fullerton

This study tests between two modern theories of decision making. Rank- and sign-dependent utility (RSDU) models, including cumulative prospect theory (CPT), imply stochastic dominance and two cumulative independence conditions. Configural weight models, with parameters estimated in previous research, predict systematic violations of these properties for certain choices. Experimental data systematically violate all three properties, contrary to RSDU but consistent with configural weight models. This study also tests whether violations of stochastic dominance can be explained by violations of transitivity. Violations of transitivity may be evidence of a dominance detecting mechanism. Although some transitivity violations were observed, most choice triads violated stochastic dominance without violating transitivity. Judged differences between gambles were not consistent with the CPT model. Data were not consistent with the editing principles of cancellation and combination. The main findings are interpreted in terms of coalescing, the principle that equal outcomes can be combined in a gamble by adding their probabilities. RSDU

Address correspondence and reprint requests to Michael H. Birnbaum, Department of Psychology, California State University, Fullerton, P.O. Box 6846, Fullerton, CA 92834-6846. E-mail: mbirnbaum@fullerton.edu. URL: <http://psych.fullerton.edu/mbirnbaum/home.htm>.

Support was received from National Science Foundation Grant SBR-9410572, to the first author, through California State University, Fullerton. We thank Teresa Martin and Juan Navarrete for assistance and advice, and we thank Chris Starmer for helpful comments on this paper.

models imply coalescing but configural weight models violate it, allowing configural weighting to explain violations of stochastic dominance and cumulative independence. © 1999 Academic Press

It has been known since the 1950s that expected utility (EU) and subjective expected utility (SEU) theories (Savage, 1954) fail to describe choices that people make between gambles. The paradoxes of Allais (1953; 1979) and other empirical violations were known to be inconsistent with these theories. Subjectively weighted utility (SWU) theories (Edwards, 1954; 1962; Karmarkar, 1978), including original prospect theory (Kahneman & Tversky, 1979), were proposed to explain empirical choices, including the “common ratio” and “common consequence” paradoxes of Allais.

SWU theories represent the value of a gamble as a weighted sum of the utilities of its outcomes (Edwards, 1954; Kahneman & Tversky, 1979). SWU explains the Allais paradoxes by allowing weights of outcomes to differ from their probabilities. However, the equation of SWU, like that of original prospect theory, makes unrealistic and inaccurate predictions. For example, it predicts that judges will violate *transparent dominance* (Fishburn, 1978), where transparent dominance is the premise that if one gamble has outcomes that are strictly better than another, the one with the better outcomes should be preferred. Similarly, if two gambles have the same outcomes, but one has higher probabilities of better outcomes, transparent dominance requires that the dominant gamble should be chosen. Consider the following choices from Birnbaum (1998a):

Choice 1: would you rather play *A* or *B*?

- | | |
|---|---|
| <i>A</i> : .5 probability to win \$100
.5 probability to win \$200 | <i>B</i> : .99 probability to win \$100
.01 probability to win \$200 |
|---|---|

Choice 2: would you rather play *C* or *D*?

- | | |
|---|---|
| <i>C</i> : .5 probability to win \$110
.5 probability to win \$120 | <i>D</i> : .01 probability to win \$101
.01 probability to win \$102
.98 probability to win \$103 |
|---|---|

According to SWU, people might prefer *B* to *A* and *D* to *C*. Such violations of transparent dominance seem quite implausible as descriptions of human behavior.¹

According to SWU, splitting an outcome's probability into smaller and smaller pieces might endlessly make a gamble better and better. To avoid this implication, Kahneman and Tversky (1979) restricted prospect theory to gambles with no more than two nonzero outcomes. They also added editing principles that would prevent strange predictions for cases such as Choices 1 and 2 above.

The *dominance* principle says that judges detect and conform to transparent dominance (as in *A* versus *B* and *C* versus *D*); in such cases, the judge will

¹ Let $SWU(G) = \sum_{i=1}^n w(p_i)u(x_i)$, where $SWU(G)$ is the subjectively weighted utility of gamble G . For this example, let $u(x) = x^\beta$ and $w(p) = cp^\gamma/[cp^\gamma + (1-p)^\gamma]$, where $\beta = c = \gamma = .4$. Then $SWU(A) = 4.2 < SWU(B) = 5.0$ and $SWU(C) = 3.8 < SWU(D) = 4.9$, both in violation of transparent dominance. See Birnbaum (1998a) for further discussion.

select the dominant gamble, even though the equation predicts the opposite. The principle of *combination* says that judges will combine equal outcomes by adding their probabilities. *Simplification* assumes that judges neglect nonessential differences, rounding off to simplify choices. The *cancellation* principle says that elements common to both gambles will be canceled and not affect choices. *Segregation* asserts that complex gambles are separated into risky and riskless components.

Although some considered these editing principles to be important insights into how people process gambles, others considered them an awkward way to “patch up” predictions of an otherwise errant model. Stevenson, Busemeyer, and Naylor (1991) noted that the editing principles not only contradict the equation, but also make different predictions when they are applied in different orders, making the theory too vague to test without further specification.

Tversky and Kahneman (1992) incorporated the representation of rank- and sign-dependent utility (RSDU) theory (Quiggin, 1982; Luce & Fishburn, 1991; 1995) into cumulative prospect theory (CPT). CPT seems an advance over prospect theory because it applies to any number of outcomes, and it automatically satisfies combination and dominance, without editing.

This study investigates new tests that are as damaging to CPT and RSDU as the Allais paradoxes are to EU. This paper will test the RDU representation basic to RSDU and CPT, and it will also investigate three of the editing principles: cancellation, combination, and dominance detection. The results will show systematic violations of stochastic dominance and two cumulative independence conditions. These violations refute not only CPT, but also all members of a general class of which it is a member. Data systematically violate the editing principles of cancellation and combination. The recipes that are used to find these violations of CPT were based on predictions from configural weight (CW) models. The predictions were made prior to the experiments, based on previously reported parameters (Birnbbaum, 1997).

Our story begins with recent experiments that found violations of branch independence and distribution independence (Birnbbaum & Chavez, 1997; Birnbbaum & McIntosh, 1996). Branch independence holds that if two gambles have a common branch (the same distinct, probability–outcome combination produced by the same event), that common branch should have no effect on the preference order induced by other branches. Branch independence can be illustrated by the following example (from Birnbbaum & Chavez, 1997). Would you prefer to play *S* or *R*?

<i>S</i> : .50 probability to win \$2	<i>R</i> : .50 probability to win \$2
.25 probability to win \$40	.25 probability to win \$10
.25 probability to win \$44	.25 probability to win \$98

Note that *S* and *R* share a common branch, a .50 probability to win \$2, but they differ on other branches. According to branch independence, the outcome \$2 on the common branch can be changed to any other distinct outcome without changing the direction of choice, as follows:

S' : .25 probability to win \$40	R' : .25 probability to win \$10
.25 probability to win \$44	.25 probability to win \$98
.50 probability to win \$108	.50 probability to win \$108

Note that S' and R' are the same as S and R , except the outcome on the common branch has been changed from \$2 to \$108. According to branch independence, S is preferred to R if and only if S' is preferred to R' . However, Birnbaum and Chavez (1997) found that whereas 60% of 100 judges preferred S to R , 62% of the same subjects preferred R' to S' , in violation of branch independence. Similar results were observed for 11 other tests of branch independence.

Violations of branch independence refute SWU theories (including EU) and the cancellation principle (Birnbaum & Chavez, 1997). If judges canceled the common outcomes in S versus R and S' versus R' , they would not show systematic violations of branch independence.

Violations of branch independence can be explained if the weight of an outcome is affected, at least in part, by the rank of the outcome among the outcomes of a gamble. To illustrate how such a CW model works, suppose the weight of the lowest outcome is three times its probability, the weight of the middle outcome is twice its probability, and the weight of the highest outcome equals its probability. Suppose also that relative weights are computed by dividing each weight by the sum of weights, and a gamble is chosen according to its configurally weighted (CW) average outcome. In the choice between S and R , the sum of the weights would be $3(.5) + 2(.25) + 1(.25) = 2.25$, and the relative weights of lowest, middle, and highest outcomes in S and R would be .67, .22, and .11, respectively. The CW average outcome is \$15.11 for S , which is greater than \$14.44, the value of R . However, in S' and R' , the sum of weights is $3(.25) + 2(.25) + 1(.5) = 1.75$, and relative weights of the lowest, middle, and highest outcomes are .429, .286, and .286, respectively. The CW average for S' is \$60.57, which is less than \$63.14, the value of R' . These weights predict the reversal observed by Birnbaum and Chavez (1997). Thus, if weights are affected by the ranks of the outcomes, one can explain violations of branch independence (Birnbaum, Coffey, Mellers, & Weiss, 1992, pp. 338–339; Weber & Kirsner, 1997).

This study tests between two classes of theories that permit weights to be affected by ranks, yet the theories make different predictions for other tests. These two rival classes are (1) the configural weight models of Birnbaum and his associates (Birnbaum, 1974, 1997, 1998a; Birnbaum & Jou, 1990; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992) and (2) the class of RSDU theories (Luce & Fishburn, 1991; 1995), including CPT (Tversky & Kahneman, 1992) and rank-dependent utility (RDU) theory (Quiggin, 1982). Many articles have studied variants of these models (Camerer, 1992; Champagne & Stevenson, 1994; Lopes, 1990; Luce, 1992, 1996; Miyamoto, 1989; Starmer & Sugden, 1989; Wakker, 1996; Wakker, Erev, & Weber, 1994; Weber, 1994; Wu, 1994; Wu & Gonzalez, 1996), but few studies have tested between them. Both account for Allais paradoxes and violations of branch independence. However, as will be shown in the next sections, these models make different predictions for stochastic dominance and cumulative independence.

CONFIGURAL WEIGHT AND RANK-DEPENDENT UTILITY MODELS

RDU Model

The RDU of a gamble can be written as follows:

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

where $u(x_i)$ is the utility of outcome x_i ($x_1 > x_2 > x_3 > \dots$); P_i is the probability that the outcome is greater than or equal x_i ; and P_{i-1} is the probability that the outcome is greater than x_i . $W(P)$ is a strictly monotonic function of decumulative probability that assigns $W(0) = 0$ and $W(1) = 1$. It is assumed that G_1 is chosen over G_2 if and only if $\text{RDU}(G_1) > \text{RDU}(G_2)$.

The *certainty equivalent* (CE) of a gamble is given by $CE(G) = u^{-1}(\text{RDU}(G))$.

When gambles are restricted to strictly positive outcomes, RSDU theory (Luce & Fishburn, 1991, 1995) and CPT (Tversky & Kahneman, 1992) reduce to RDU (Eq. (1)). In their model of CPT, Tversky and Kahneman (1992) used the equation

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

where $\gamma = .61$. Tversky and Kahneman (1992) also estimated $u(x) = x^\beta$, where $\beta = .88$.

A more general, two-parameter function for $W(P)$ has been discussed by Tversky and Wakker (1995). This weighting function can be written as

$$W(P) = \frac{cP^\gamma}{cP^\gamma + (1 - P)^\gamma}, \quad (2b)$$

where c can be interpreted as an index of risk aversion, apart from the $u(x)$ function. When $P = 1/2$, $W(P) = c/(c + 1)$ for any γ . Equations (2a) and (2b) can both produce an S-shaped ($\gamma > 1$) or inverse-S ($\gamma < 1$) shaped curve relating decumulative weight, $W(P)$, to decumulative probability, P . Tversky and Wakker (1995) used the term "S-shaped" for what we term the inverse-S (Tversky & Fox, 1995).

Violations of branch independence and distribution independence show the opposite pattern from that predicted by the inverse-S function (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997). For example, the model of CPT, with parameters of $\gamma = .61$, $\beta = .88$, and $c = .724$, predicts that judges should prefer R to S and S' to R' in the choice problem illustrated above, exactly

opposite the modal preference order found by Birnbaum and Chavez (1997). ($CE(S) = \$17.58 < CE(R) = \25.83 and $CE(S') = \$68.29 > CE(R') = \63.12).²

Although observed violations are opposite those predicted by the model of Tversky and Kahneman (1992), violations of branch independence and distribution independence would be consistent with Eq. (1) with a different $W(P)$ function (also assuming one deletes the editing principle of cancelation). Birnbaum and Chavez (1997) fit Eqs. (1) and (2b) to their data and estimated $\gamma = 1.59$, quite different from the inverse-S function with $\gamma < 1$.

Configural Weight Models

Birnbaum and McIntosh (1996) argued that the apparent contradiction between violations of branch independence and the weighting functions of Tversky and Kahneman (1992) and Wu and Gonzalez (1996) might be evidence that CPT is wrong, because the contradiction in CPT is not a contradiction in configural weight models. This contradiction in weighting functions led Birnbaum (1997) to deduce the cumulative independence conditions, described below, to provide a direct test between RSDU/RDU/CPT and the configural weight models (see Appendix).

Configural weight averaging models were proposed to represent judgments in a variety of tasks, including impression formation (Birnbaum, 1974), moral evaluation (Birnbaum, 1973), and judgments of the value of goods or investments based on estimates from sources of varied expertise and bias (Birnbaum & Stegner, 1979). Birnbaum and Stegner (1979) used a configural weight averaging model to describe judgments of the values of used cars in the buyer's, seller's, and neutral's ("fair price") viewpoints. Configural weighting also successfully predicts judgments of buying and selling prices of investments based on estimates provided by sources of varied expertise in predicting the future values of investments (Birnbaum & Zimmermann, 1998).

Analogies among judgment tasks (Birnbaum & Mellers, 1983) show that a configural weight averaging model that accounts for these different judgment tasks is also a viable model of both judgments and choices between gambles (Birnbaum & Beeghley, 1997). In averaging models, there are two parameters for a stimulus, subjective scale value (analogous to utility) and weight, which is analogous to probability weighting. Averaging models, unlike the equation used in original prospect theory, do not violate transparent dominance (Birnbaum, 1998a), but as we will show in a later section, they can violate stochastic dominance in other situations. Averaging models have two important properties: the effect of a manipulation that affects scale value (value of the outcomes)

² Computer programs, DMALC and DMALC2, have been written in BASIC by the first author to compute the predictions of the models compared in this study. These programs are useful for designing experiments to distinguish these models, and for exploring the properties of the models. These programs have also been converted (in collaboration with Rob Bailey) to on-line calculators that will run through a JavaScript compatible (e.g., Netscape Navigator 3.0 or above) web browser. Requests may be sent to the first author, and these programs are also available at URL <http://psych.fullerton.edu/mbirnbaum/programs.htm>.

will be amplified by a factor that affects weight (probability), and the effect of the same manipulation will be inversely related to the total weight of other factors.

RAM Model

The rank-affected multiplicative (RAM) model of Birnbaum and McIntosh (1996) can be written for gambles of strictly positive outcomes as

$$U(G) = \frac{\sum_{i=1}^n u(x_i) S(p_i) a_V(r_i, n)}{\sum_{i=1}^n S(p_i) a_V(r_i, n)}, \quad (3)$$

where $u(x_i)$ is the utility of outcome x_i ($x_i > 0$); $S(p_i)$ is the psychological weight of the probability of outcome x_i ; r_i is the rank of x_i among the n distinct outcomes (r_i ranges from 1 = highest rank to n = lowest rank); and $a_V(r_i, n)$ is the configural weight of this rank in point of view, V . Point of view is affected by instructions, such as those to identify with a buyer or seller, to identify with a “neutral” judge, or to take the viewpoint of choice. (For the treatment of negative and zero outcomes in this model, see Birnbaum, 1997.)

The RAM model can be deduced from the assumption that the value of a gamble is the value that minimizes a loss function defined on discrepancies between the judged value and the possible outcomes of a gamble. If this loss function is the expected value of a (symmetric) squared loss function defined on differences in utility, minimizing the expected loss leads to the EU representation (Birnbaum et al., 1992). If the loss function is asymmetric, however, then this approach implies a configural weight, averaging model in which the configural weights represent the relative costs of overestimating as opposed to underestimating the value of a gamble.

Derivations of the configural models from the loss function rationale applied to gambles are given in Birnbaum et al. (1992) and Birnbaum and McIntosh (1996, Appendix A). The derivations show (Birnbaum et al., 1992, p. 336, Eq. 6) that judgments of binary gambles can be an inverse-S function of probability if the psychophysical function of probability is negatively accelerated (Varey, Mellers, & Birnbaum, 1990). Thus the inverse-S observed by Tversky and Kahneman (1992) for certainty equivalents of binary gambles as a function of probability can be viewed as the consequence of the relative weight feature of an averaging model combined with a negatively accelerated psychophysical function for probability.

According to the loss function rationale for configural weighting, factors that affect the relative costs of over- or underestimating the value of a good (taking the viewpoint of the buyer or seller) should affect the configural weights. Birnbaum and Zimmermann (1998) noted that changing configural weights in different viewpoints can explain data for buyer’s and seller’s prices, but the reference level (“loss aversion”) notion coupled with the specific model of CPT

cannot. In configural weight theory, choice is a viewpoint intermediate between buying and selling, similar to that of the “neutral” or “fair price” judgment viewpoint.

The RAM model is an averaging model with configural weights that are the product of a function of the outcome’s probability, $S(p_i)$, and a function of the outcome’s rank in a given point of view. Birnbaum and McIntosh (1996) estimated (for choice) that the weights for three-outcome gambles are approximately proportional to their ranks (3: 2: 1): .51, .33, and .16 for the lowest, middle, and highest, respectively.

Birnbaum and McIntosh (1996) noted that the RAM model (with $S(p_i) = p_i^{\beta}$, $u(x) = x$, and configural weights of .63 and .37 for the lower and higher of two outcomes) makes virtually the same predictions as the model of Tversky and Kahneman (1992) for two-outcome gambles. Birnbaum and McIntosh (1996) noted that the RAM model also accounts for the results of Wu and Gonzalez (1996) without changing parameters; however, the CPT model cannot explain the data of both Birnbaum and McIntosh and Wu and Gonzalez without changing parameters.

Although the RAM model accounts for a variety of data (Birnbaum, 1998a), the RAM model implies distribution independence, which is illustrated by the following choices:

<i>E</i> : .59 probability to win \$4	<i>F</i> : .59 probability to win \$4
.20 probability to win \$45	.20 probability to win \$11
.20 probability to win \$49	.20 probability to win \$97
.01 probability to win \$110	.01 probability to win \$110

Note that there are two common branches, a .59 probability to win \$4, and a .01 probability to win \$110. According to distribution independence, the probabilities of the common branches can be changed, and the preference order should not be altered. In choices E' and F' , the probabilities of the common outcomes have been changed as follows:

E' : .01 probability to win \$4	F' : .01 probability to win \$4
.20 probability to win \$45	.20 probability to win \$11
.20 probability to win \$49	.20 probability to win \$97
.59 probability to win \$110	.59 probability to win \$110

The RAM model implies $E > F$ if and only if $E' > F'$, because the configural weight of an outcome is a function of rank and probability (Birnbaum & Chavez, 1997, pp. 176–177). However, most judges preferred E to F and most judges preferred F' to E' (Birnbaum & Chavez, 1997). Violations of distribution independence refute original prospect theory and the RAM model, but they are consistent with either RDU, or with another configural weighting model, known as the TAX model.

TAX Model

Judgments of a person’s morality as a function of the person’s deeds show a violation of asymptotic independence that is also observed in judgments of the

buying prices of used cars (Birnbbaum, 1973; Birnbbaum & Stegner, 1979). If one deed (or one estimate of a used car) is low, the judgment is low and other deeds (or estimates) do not appear to be capable of compensating for the low estimate. This violation of asymptotic independence (Birnbbaum, 1997) has been described by a model in which there are transfers of weight among the stimuli (Birnbbaum & Stegner, 1979).

If one thinks of weights as measures of importance, or attention that a stimulus receives, then this model can be termed a *transfer of attention exchange* (TAX) model. In this model, there is a fixed amount of attention (weight), which is transferred among stimuli according to their relative positions by a policy that depends on viewpoint. The amount of weight transferred from a stimulus is proportional to the amount of weight that the stimulus has to lose. For example, if the worst deed that a person has done takes weight from good deeds, then its relative weight may not be driven to zero by doing an infinite number of good deeds. Thus, in the TAX model, a person who has done one very bad deed might no longer be judged "moral" no matter how many good deeds that person might do. Similarly, a used car that has received one low estimate might not be judged high in buying price no matter how many other mechanics give it a good report.

In the TAX model, weight is transferred among stimuli as a proportion of the weight of the stimulus losing weight. A general weight transfer model can be written as

$$U(G) = \frac{\sum_{i=1}^n S(p_i)u(x_i) + \sum_{i=2}^n \sum_{j=1}^{i-1} [u(x_i) - u(x_j)]\omega(i, j, G)}{\sum_{i=1}^n S(p_i)}, \quad (4)$$

where $U(G)$ is the utility of the gamble; the outcomes are ranked from lowest to highest, $x_1 < x_n$; $S(p)$ is a function of probability; and $\omega(i, j, G)$ are the configural transfers of weight among the outcomes. If the configural transfers, $\omega(i, j, G)$, are all zero, this model reduces to a subjectively weighted average utility model; if transfers are all zero and $S(p) = p$, the model reduces to EU. Birnbbaum and Chavez (1997) proposed the following simplification for choice,

$$\omega(i, j, G) = S(p_i)\delta/(n + 1) \quad \text{if } \delta < 0; \quad (5a)$$

$$\omega(i, j, G) = S(p_j)\delta/(n + 1) \quad \text{if } \delta \geq 0; \quad (5b)$$

where the single configural parameter is δ , and n is the number of distinct outcomes in the gamble. The ratio $\delta/(n + 1)$ is the proportion of weight transferred from one outcome to another, analogous to a tax rate. As in the other models, $CE(G) = u^{-1}[U(G)]$.

Birnbbaum and Chavez (1997) noted that the TAX model, with $u(x) = x$, $S(p) = p_i^7$, and $\delta = -1$, makes the same predictions as the RAM model for the experiment of Birnbbaum and McIntosh (1996). Birnbbaum (1998a) noted

that with these same parameters, the TAX model can account for the Allais paradoxes, violations of branch independence, risk aversion for high probabilities of positive outcomes, risk seeking for small probabilities of positive outcomes, and violations of distribution independence. For example, the TAX model predicts that $CE(E) = \$21.70 > CE(F) = \20.85 and $CE(E') = \$49.85 < CE(F') = \50.03 . The CPT model of Tversky and Kahneman (1992), deleting the editing rule of cancelation, predicts the opposite preference order.

Both RAM and TAX models are averaging models with the following property: if the function of probability, $S(p)$, is negatively accelerated (i.e., if there is a nonlinear psychophysical function), then a given probability–outcome branch can gain weight by splitting into two branches. This means that configural weight models violate the property of coalescing that will be introduced below; we will argue that violation of coalescing is the key to understanding violations of RSDU/RDU/CPT reported in this paper.

Summary of the Story So Far

Table 1 summarizes implications of EU, SWU, RAM, RDU/RSDU, and TAX models, and it compares predictions to previous results. EU theory implies that choices should satisfy common consequence independence and common ratio independence; therefore, EU is refuted by the Allais paradoxes. SWU and original prospect theory can explain the Allais paradoxes, but they imply branch independence and distribution independence, contrary to data of Birnbaum and McIntosh (1996) and Birnbaum and Chavez (1997). The configural weight, RAM model explains violations of branch independence, but it implies distribution independence. Although the inverse-S weighting function is not consistent with the particular violations of branch independence reported by Birnbaum and McIntosh (1996) and Birnbaum and Chavez (1997), the general form of RDU or RSDU can account for violations of branch independence and distribution independence. With a different weighting function, RDU explains the Allais paradoxes. The configural weight TAX model can also account for the findings reviewed thus far. Thus, two classes of theories are left standing,

TABLE 1

Summary of Previous Research and Implications for the Decision Models

Model	Common consequence independence	Common ratio independence	Branch independence	Distribution independence
EU/SEU	Satisfied	Satisfied	Satisfied	Satisfied
SWU	Violated	Violated	Satisfied	Satisfied
CW RAM	Violated	Violated	Violated	Satisfied
RDU/RSDU	Violated	Violated	Violated	Violated
CW TAX	Violated	Violated	Violated	Violated
Empirical Results	Violated	Violated	Violated	Violated

the RDU/RSDU/CPT models (with inconsistencies in the weighting function between studies) and the configural weight, TAX model.

In the next section, we show that configural models violate stochastic dominance and cumulative independence, which the class of RDU/RSDU/CPT must satisfy. We assemble these implications from simpler components, to clarify possible theoretical interpretations.

TESTABLE PROPERTIES

Simple Properties

Let $G = (x, p; y, q; z, r)$ represent a three-outcome gamble that yields a consequence of x with probability p , y with probability q , and z with probability $r = 1 - p - q$; all probabilities are nonzero and all outcomes distinct. Let $>$ represent the preference relation and \sim represent indifference.

1. *Transitivity* of preference means

$$A > B \text{ and } B > C \quad \text{implies } A > C.$$

Any theory that assumes that gamble A is preferred to gamble B if and only if $U(A) > U(B)$ will inherit transitivity of preference from the transitivity of numbers (utilities) that represent the gambles. All of the models considered here imply transitivity because they represent the utilities of gambles on a single dimension.

Because transitivity might be violated in choices by random error, transitivity has also been defined in terms of choice probabilities (e.g., Tversky, 1969). Weak stochastic transitivity (WST) is defined as follows: if $P(A, B) > 1/2$ and $P(B, C) > 1/2$ then $P(A, C) > 1/2$, where $P(A, B)$ is the probability of choosing A over B . According to strong stochastic transitivity, $P(A, C)$ should exceed the larger of $P(A, B)$ and $P(B, C)$.

2. *Outcome monotonicity* means that increasing the value of one outcome, holding everything else in the gamble constant, will improve the gamble. For example, with a three-outcome gamble, $G = (x, p; y, q; z, r)$, monotonicity implies

$$(x^+, p; y, q; z, r) > G \quad \text{iff } x^+ > x;$$

$$(x, p; y^+, q; z, r) > G \quad \text{iff } y^+ > y;$$

$$(x, p; y, q; z^+, r) > G \quad \text{iff } z^+ > z.$$

Although monotonicity has been violated in judgment when the lowest outcome is increased from zero to a small positive value, that recipe produces few violations in direct choices between gambles (Birnbbaum & Sutton, 1992; Birnbbaum, 1997). Systematic violations have also not been reported when the number of outcomes is fixed and all outcomes are greater than zero. All three models above imply satisfaction of monotonicity when there is a fixed number of positive outcomes.

3. *Coalescing* assumes that equal outcomes can be combined by adding their probabilities; for example, for three-outcome gambles,

$$GS = (x, p; x, q; z, r) \sim G = (x, p + q; z, r),$$

and

$$FS = (x, p; y, q; y, r) \sim F = (x, p; y, q + r),$$

where *GS* and *FS* are *split* versions and *G* and *F* are the coalesced versions of the same actual gambles. Note that mathematically equivalent gambles can be presented in (psychologically) different formats, which should not make a difference if coalescing holds.

“Event-splitting effect” (Humphrey, 1995; Starmer & Sugden, 1993) refers to a violation of the following: $A > B$ if and only if $AS > BS$, where *AS* and *BS* are split versions of *A* and *B*. Assuming coalescing and transitivity, there should be no event-splitting effects. *Proof*: Coalescing implies that $AS \sim A$ and $BS \sim B$; therefore $A > B$ iff $AS \sim A > B \sim BS$. By transitivity, $AS > BS$.

Coalescing is implied by RDU models with any $W(P)$ function (Birnbbaum & Navarrete, 1998; Luce, 1998). In original prospect theory (Kahneman & Tversky, 1979), coalescing was imposed by the editing rule of combination, but in RDU, RSDU, and CPT, coalescing is implied.

Coalescing is violated by both RAM and TAX models. For example, in the TAX model with $S(p) = p^\gamma$, if $\gamma < 1$, then splitting the lowest outcome of a gamble can make the gamble worse, and splitting the highest outcome of a gamble can make it better. Suppose $u(x) = x$, $\delta = 0$, and $\gamma = .7$; then $CE(\$0, .25; \$0, .25; \$100, .5) = \$44.82 < CE(\$0, .5; \$100, .5) = \$50 < CE(\$0, .5; \$100, .25; \$100, .25) = \$55.18$. (See footnote 2.) These violations of coalescing occur because splitting an outcome can increase that outcome’s total weight. Intuitively, two outcomes with the same total probability can have more weight than one combined outcome, as if each discrete outcome gets some attention or consideration beyond its objective probability.

When δ is not zero, event-splitting has additional effects in the TAX model (beyond the effect of γ), because the relative weight of an outcome of a given rank will depend on the other outcomes in the gamble. For example, with $\gamma = 1$ and $\delta = -1$, $CE(\$0, .25; \$0, .25; \$100, .5) = \$25, < CE(\$0, .5; \$100, .5) = \$33.33 < CE(\$0, .5; \$100, .25; \$100, .25) = \$37.5$. Violations of coalescing occur even when $\gamma = 1$ because splitting the lowest of two outcomes converts that outcome into the lowest *and* middle outcome of a three-outcome gamble. Splitting the highest outcome gives the outcome the weight due to the middle *and* highest outcomes of the gamble.

4. *Branch independence* asserts that if two gambles have a common branch, then the choice between them will be independent of the outcome on that common branch. The term “branch” designates that the probability–

outcome combination is distinct in the problem presentation. For three-outcome gambles, restricted branch independence requires

$$(x, p; y, q; z, r) > (x', p; y', q; z, r)$$

if and only if

$$(x, p; y, q; z', r) > (x', p; y', q; z', r),$$

where (z, r) is the common branch, the outcomes (x, y, z, x', y', z') are all distinct, and the probabilities are not 0 and they sum to 1. This principle is weaker than Savage's (1954) independence axiom because it is restricted to distinct branches of known probability, and it does not presume coalescing. Branch independence is implied by cancelation (Kahneman & Tversky, 1979), but it is violated by RDU (Eq. (1)), RAM (Eq. (3)), and TAX (Eqs. (4) and (5)), models.

The special case of branch independence in which the outcomes retain the same ranks in all gambles is termed *comonotonic branch independence* (Wakker et al., 1994). All of the models specified above imply restricted comonotonic branch independence for positive outcomes.

In summary, the models agree on transitivity, monotonicity, and restricted comonotonic branch independence, but they disagree on coalescing. We next show that violations of coalescing can create violations of stochastic dominance and cumulative independence.

PREDICTIONS OF THE MODELS

Stochastic Dominance

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

A stochastically dominates B if and only if

$$P(x > t|A) \geq P(x > t|B) \text{ for all } t \quad (6)$$

where $P(x > t|A)$ is the probability that an outcome of Gamble A exceeds t .

The statement *preferences satisfy stochastic dominance* means

$$\text{if } A \text{ stochastically dominates } B, \text{ then } A > B. \quad (7)$$

It would then be a violation of stochastic dominance if A stochastically dominates B but a judge chooses B over A . With fallible data, one can test the (very conservative) hypothesis that if A stochastically dominates B , then the probability of choosing A over B should exceed 1/2.

Birnbaum (1997, 1998a) noted that whereas RDU, RSDU, and CPT models imply stochastic dominance, RAM and TAX models imply systematic violations in choices constructed from a special recipe. Birnbaum's (1997) recipe is illustrated by example: Start with $G_0 = (\$12, .1; \$96, .9)$. Split the lower outcome

of G_0 , making G^+ dominant over G_0 as follows: $G^+ = (\$12, .05; \$14, .05; \$96, .9)$. Next, split the higher outcome of G_0 , making G^- dominated by G_0 : $G^- = (\$12, .1; \$90, .05; \$96, .85)$. G^+ stochastically dominates G^- .

Assuming transitivity, monotonicity, and coalescing, choices must obey stochastic dominance in this recipe; therefore, RDU/RSDU/CPT theories satisfy stochastic dominance. However, RAM and TAX models (by violating coalescing) violate stochastic dominance. For example, the TAX model (with $\gamma = .7$ and $\delta = -1$) implies $CE(G^+) = \$45.77 < CE(G_0) = \$58.10 < CE(G^-) = \$63.10!$ (See footnote 2.)

Birnbaum and Navarrete (1998) found that 70% of 100 judges chose G^- over G^+ in four variations of this recipe, contrary to stochastic dominance. Because stochastic dominance follows from transitivity, monotonicity, and coalescing, it is possible that violations might be due to one of these principles. This study will investigate if violations of transitivity account for violations of stochastic dominance.

On the other hand, violations of stochastic dominance might cause violations of transitivity if judges detect and conform to stochastic dominance in comparisons of the three-outcome gambles against G_0 . In other words, if judges prefer G^+ to G_0 and G_0 to G^- , and if they persist in choosing G^- over G^+ , then their choices would be intransitive. A major purpose of the present study is to explore these possible connections between stochastic dominance and transitivity.

Cumulative Independence and Branch Independence

Birnbaum (1997) derived (from RDU) 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.

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

$$\text{then } S'' = (x', r; y, p + q) > R'' = (x', r + p; y', q) \quad (8)$$

Upper cumulative independence.

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

$$\text{then } S''' = (x, p + q; y', r) < R''' = (x', p; y', q + r) \quad (9)$$

Any theory that satisfies comonotonic independence, monotonicity, transitivity, and coalescing must satisfy both lower and upper cumulative independence (Birnbaum, 1997; Birnbaum & Navarrete, 1998); therefore, RDU/RSDU/CPT models imply cumulative independence (see also Appendix).

Figure 1 illustrates a test of lower cumulative independence. The branch with the lowest outcome (z, r) has been improved in both gambles to (x', r) , which should not reverse preferences, according to comonotonic branch independence. It has also been coalesced in R' . Furthermore, S has been improved by increasing x to y (and coalescing) to create S' . Therefore, it is consistent

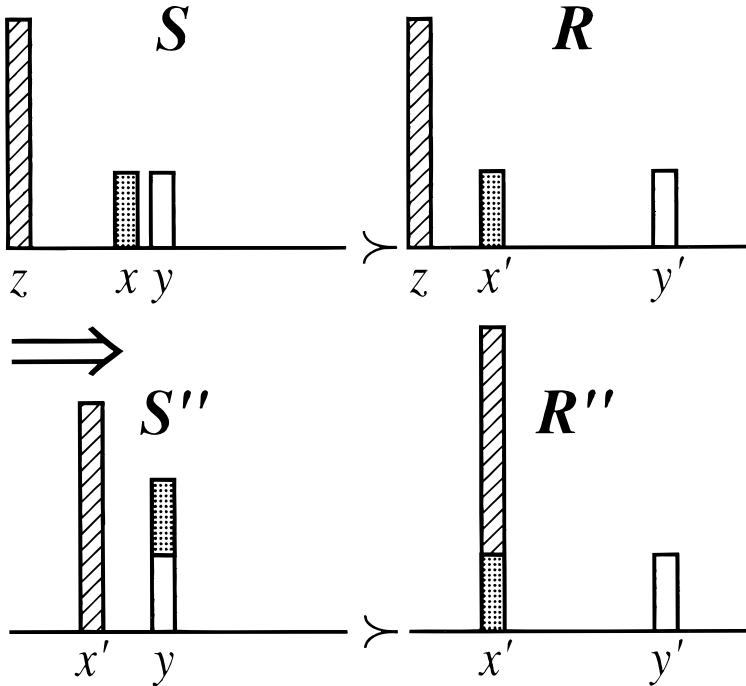


FIG. 1. A test of lower cumulative independence: If $S > R$, then $S' > R'$. Note that the second comparison is the same as the first, except z has been increased to x' on both sides, and x has been increased to y (and coalesced) in S' , which should make S' relatively better.

with lower cumulative independence to switch from $R > S$ to $S' > R'$; however, it is a violation to change preferences from $S > R$ to $S' < R'$.

In upper cumulative independence (Fig. 2), the branch with the highest outcome (z' , r) has been reduced in both gambles from z to y' , which should not change the preference order, by comonotonic branch independence. Furthermore, S' has been made worse by reducing y to x to create S'' . Therefore, it is consistent with upper cumulative independence to switch from $S' > R'$ to $R'' > S''$; however, it is a violation to switch in the opposite direction.

In contrast, RAM and TAX models predict violations of both cumulative independence conditions. For example, predicted certainty equivalents for the TAX model (with $u(x) = x$, $S(p) = p^7$, and $\delta = -1$) are

$$\begin{aligned} CE(S) &= CE(\$2, .6; \$40, .2; \$44, .2) = \$16.19 \\ &> CE(R) &= CE(\$2, .6; \$10, .2; \$98, .2) = \$15.47, \end{aligned}$$

but $CE(S') = CE(\$10, .6; \$44, .4) = \$19.74 < CE(R') = CE(\$10, .8; \$98, .2) = \26.12 , contrary to lower cumulative independence. In addition,

$$\begin{aligned} CE(S'') &= CE(\$40, .2; \$44, .2; \$108, .6) = \$58.89 \\ &< CE(R'') &= CE(\$10, .2; \$98, .2; \$108, .6) = \$62.72, \end{aligned}$$

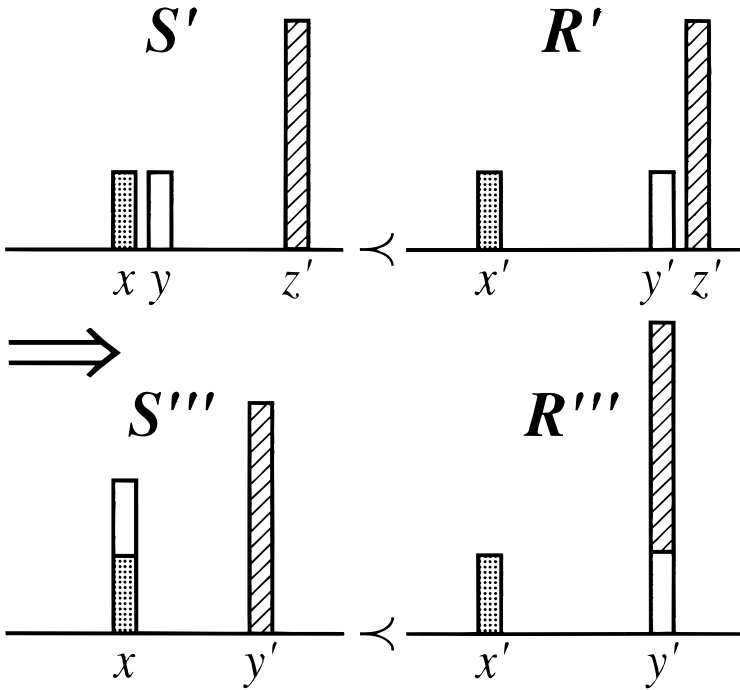


FIG. 2. A test of upper cumulative independence: If $S' < R'$, then $S'' < R''$. Note that the second comparison is the same as the first, except z' has been decreased to y' on both sides, and y has been decreased to x (and coalesced) on the left side, which should make S'' relatively worse.

but $CE(S''') = CE(\$40, .4; \$98, .6) = \$62.06 > CE(R''') = CE(\$10, .2; \$98, .8) = \52.55 , contradicting upper cumulative independence. (See footnote 2.)

Birnbaum and Navarrete (1998) found violations of both cumulative independence properties. This study tests for violations predicted by RAM and TAX models, based on the previously published parameters, for six new sets of gambles that have not been previously tested.

Choices used in tests of cumulative independence (expressions (8) and (9)) can also be used to test branch independence, which asserts that $S > R$ if and only if $S' > R'$. The TAX model predicts that $CE(S) = \$16.19 > CE(R) = \15.47 but $CE(S') = \$58.89 < CE(R') = \62.72 . The CPT model with parameters of Tversky and Kahneman (1992) makes the opposite predictions.

Interval Independence

When subjects are asked how much they would pay to receive gamble A instead of B , it is reasonable to theorize that the greater the difference in utility between the gambles, the more they would be willing to pay (Birnbaum, Thompson, & Bean, 1997). Suppose that such judgments are a function of utility intervals,

$$D(A, B) = J[U(A) - U(B)], \tag{10}$$

where $D(A, B)$ is a judgment of strength of preference between two gambles, J is a strictly increasing monotonic judgment function, and $U(A)$ and $U(B)$ are the utilities of the two gambles.

Suppose two gambles are otherwise identical but differ on one branch. Interval independence asserts that the difference in utility depends entirely on the differing branch and is independent of the number, values, and probability distribution of outcomes that are common to both gambles.

Let $D(A, B)$ represent the judged strength of preference for A over B . Suppose A and B differ in their outcomes on one branch. Interval independence can be written as follows:

$$\begin{aligned} \text{If } A &= (x_1, p_1; x_2, p_2; \dots; x_i, p_i; \dots x_n, p_n) \text{ and} \\ B &= (x_1, p_1; x_2, p_2; \dots; y_i, p_i; \dots x_n, p_n), \text{ and if} \\ A' &= (y_1, q_1; y_2, q_2; \dots; x_i, p_i; \dots y_m, q_m) \text{ and} \\ B' &= (y_1, q_1; y_2, q_2; \dots; y_i, p_i; \dots y_m, q_m), \text{ then} \\ D(A, B) &= D(A', B'). \end{aligned} \tag{11}$$

Note that A and B differ only in the outcome for branch i (x_i instead of y_i), and A' and B' differ in the same branch. If judges edit and eliminate components common to both gambles in a comparison (Kahneman & Tversky, 1979; Wu, 1994), and if Eq. (10) holds, then interval independence should hold, because the common components will have no effect. SWU and EU also imply interval independence (Birnbbaum et al., 1997), assuming Eq. (10). Birnbbaum's (1974, Experiment 4) "scale free" test is a test of interval independence.

When gambles are also comonotonic, so that the rank position of the contrasting branch is the same in all four gambles, this special case is termed *comonotonic interval independence*. Assuming Eq. (10), RDU implies comonotonic interval independence, and it also implies that strength of preference should be independent of the number and values of common branches as long as the cumulative probabilities are the same. RAM and TAX models violate both properties.

For example, let $A = (\$12, .8; \$96, .2)$, $B = (\$12, 1.0)$; $A' = (\$12, .4; \$96, .6)$, $B' = (\$12, .6; \$96, .4)$. The difference between A and B is a .2 branch to win \$96 rather than \$12, as is the difference between A' and B' . Both configural weight models and RDU allow that $D(A, B)$ and $D(A', B')$ will not be equal in general.

According to RDU, however, the interval between gambles with five equally likely outcomes, $C = (\$2, \$4, \$6, \$7, \$96)$ and $D = (\$2, \$4, \$6, \$7, \$12)$, should be the same as the interval between C' and D' , where $C' = (\$12, .8; \$96, .2)$ and $D' = \$12$ for sure. These intervals should be the same because the cumulative probabilities and the outcomes of the differing branch are the same, so comonotonic common branches will drop out. Similarly, the interval between $E = (\$96, \$108, \$111, \$113, \$115)$ and $F = (\$12, \$108, \$111, \$113, \$115)$ should be the same as the interval between $E' = \$96$ for sure and $F' = (\$12, .2; \$96, .8)$, a

prediction that will be tested in this study. RAM and TAX models do not require that these intervals will be equal.

Comparison of the Models

All three models (RDU, RAM, and TAX) allow violations of branch independence, unlike SWU and EU models. For strictly positive outcomes, all three models satisfy transparent dominance. All three make similar predictions for two-outcome gambles. Despite these similarities, however, the models make different predictions for the tests in this paper. The class of RDU, RSDU, and CPT models imply transitivity, monotonicity, coalescing, and comonotonic branch independence. These models therefore imply stochastic dominance, lower cumulative independence, and upper cumulative independence, three properties that are violated by the configural weight TAX and RAM models. This study investigates cases where the configural weight models predict violations, based on their previously published parameters.

METHOD

Judges were instructed to choose between gambles and to judge how much they would pay to play their preferred gamble rather than the other gamble in each choice.

Designs

Stochastic dominance and transitivity design. There were 15 trials designed to test stochastic dominance and transitivity, composed of five variations of the following three choices: G^- versus G_0 , G_0 versus G^+ , and G^- versus G^+ . Variations 1 through 5 of $\{G_0, G^-, G^+\}$ are $\{(\$2, .04; \$98, .96), (\$2, .05; \$93, .02; \$98, .93), (\$2, .02; \$6, .02; \$98, .96)\}$, $\{(\$3, .12; \$92, .88), (\$3, .12; \$91, .02; \$92, .86), (\$3, .10; \$5, .02; \$92, .88)\}$, $\{(\$8, .04; \$97, .96), (\$8, .04; \$95, .02; \$97, .94), (\$8, .02; \$9, .02; \$97, .96)\}$, $\{(\$4, .05; \$88, .95), (\$4, .05; \$86, .02; \$88, .93), (\$4, .03; \$6, .02; \$88, .95)\}$, and $\{(\$2, .20; \$108, .80), (\$2, .20; \$96, .10; \$108, .70), (\$2, .10; \$12, .10; \$108, .80)\}$, respectively. The right–left positions of the gambles were as listed above, except in Variations 2 and 4, which used the opposite left–right positions.

Cumulative independence and branch independence design. This design was composed of six variations of each of the following four choices, making 24 trials:

$$\begin{aligned}
 S &= (z, .6; x, .2; y, .2) && \text{versus } R = (z, .6; x', .2; y', .2); \\
 S' &= (x', .6; y, .4) && \text{versus } R' = (x', .8; y', .2); \\
 S'' &= (x, .2; y, .2; z', .6) && \text{versus } R'' = (x', .2; y', .2; z', .6); \\
 S''' &= (x, .4; y', .6) && \text{versus } R''' = (x', .2; y', .8).
 \end{aligned}$$

The six levels of (z, x', x, y, y', z') were factorially combined with the four types of comparisons; these six levels were (\$2, \$11, \$52, \$56, \$97, \$108), (\$3, \$10, \$48, \$52, \$98, \$107), (\$2, \$11, \$45, \$49, \$97, \$109), (\$2, \$10, \$40, \$44, \$98, \$110), (\$4, \$11, \$35, \$39, \$97, \$111), and (\$5, \$12, \$30, \$34, \$96, \$108).

Interval independence design (transparent dominance). This design was composed of eight subdesigns. There were 32 choices with transparent dominance, in which both gambles were the same, except one outcome was higher in one gamble, or the probability of a better outcome was higher in one gamble.

In Subdesigns 1–6, the gamble on the right was the same as that on the left, except it had a .2 probability to receive \$96 rather than \$12. Subdesign 1 had five choices of the form

$$(\$12, p; \$96, 1 - p) \quad \text{versus} \quad (\$12, p - .2; \$96, 1.2 - p),$$

where $p = 1, .8, .6, .4, \text{ and } .2$. In Subdesign 2, there were four choices of the form

$$(\$2, .2; \$12, p; \$96, .8 - p) \quad \text{versus} \quad (\$2, .2; \$12, p - .2; \$96, 1 - p),$$

with $p = .8, .6, .4, .2$. Note that these are the same as in Subdesign 1, except the lowest outcome has been split to include a lower outcome (\$2). Subdesign 3 used four choices of the form

$$(\$12, p; \$96, .8 - p; \$108, .2) \quad \text{versus} \quad (\$12, p - .2; \$96, 1 - p; \$108, .2),$$

with $p = .8, .6, .4, \text{ and } .2$; these are the same as Subdesign 1, except the highest outcome has been split to include a higher branch (\$108). Subdesign 4 added a middle branch of \$48, using four choices,

$$(\$12, p; \$48, .2; \$96, .8 - p) \quad \text{versus} \quad (\$12, p - .2; \$48, .2; \$96, 1 - p),$$

where $p = .8, .6, .4, \text{ or } .2$. Subdesign 5 used choices between five-outcome gambles in which each outcome had probability of .2; these five choices were (\$2, \$4, \$6, \$7, \$12) versus (\$2, \$4, \$6, \$7, \$96), (\$3, \$6, \$8, \$12, \$108) versus (\$3, \$6, \$8, \$96, \$108), (\$3, \$5, \$12, \$108, \$112) versus (\$3, \$5, \$96, \$108, \$112), (\$5, \$12, \$107, \$109, \$113) versus (\$5, \$96, \$107, \$109, \$113), and (\$12, \$108, \$111, \$113, \$115) versus (\$96, \$108, \$111, \$113, \$115). Subdesign 6 used four choices,

$$(z, p; \$12, .2; z', .8 - p) \quad \text{versus} \quad (z, p; \$96, .2; z', .8 - p),$$

where $p = .8, .6, .4, \text{ or } .2$, and $(z, z') = (\$5, \$96), (\$4, \$108), (\$3, \$114), \text{ or } (\$2, \$110)$, respectively.

Subdesigns 7 and 8 used six choices composed of three variations of the following two forms,

$$(\$10, r; \$99, 1 - r) \quad \text{versus} \quad (\$10, r + .1; \$99, .9 - r),$$

and

$$(\$3, r; \$99, .1; \$107, .9 - r) \quad \text{versus} \quad (\$3, r; \$10, .1; \$107, .9 - r),$$

where $r = .85, .45, \text{ and } .05$. Note that these six gambles all have an additional .1 probability to win \$99 on the left instead of \$10 on the right.

Procedure and Judges

Each booklet contained three pages of instructions with examples, 10 practice trials, followed by 71 experimental choices ($15 + 24 + 32$), randomly ordered and embedded among an additional 22 unlabeled warmups and fillers, for a total of 93 trials. Choices were printed in random orders, restricted so that successive trials did not repeat the same design. Two booklets used different orderings (and different experimenters); half of the judges in each booklet worked in reverse order. Other details of stimulus presentation and procedure were as in Birnbaum and Chavez (1997).

The judges were 110 undergraduates who violated transparent dominance no more than twice out of 32 tests. Of the 110 judges, 58 had no violations of transparent dominance, 26 had one, and 26 had two (average rate of violation of transparent dominance is 2.2%).

RESULTS

Violations of Stochastic Dominance and Transitivity

Table 2 shows choice patterns for tests of stochastic dominance and transitivity. Rows represent variations, in the order listed under Method. The column labeled " $G^- > G^+$ " shows the number of judges (out of 110) who chose G^- over G^+ , violating stochastic dominance. For example, in the first variation, 90 out of 110 (81.8%) violated stochastic dominance by choosing $G^- = (\$2, .05; \$93, .02; \$98, .93)$ over $G^+ = (\$2, .02; \$6, .02; \$98, .96)$. In every variation of the recipe, significantly more judges violated stochastic dominance when comparing G^- to G^+ than allowed by the null hypothesis that stochastic dominance is satisfied half the time (critical value is 66 for a two-tailed, binomial sign test with $\alpha = .05$). Averaged over variations, 73.6% violated stochastic dominance. Therefore, we can reject the hypothesis that people conform to stochastic dominance (at least half the time) in favor of the hypothesis that they systematically choose the dominated gambles in this recipe.

Examining each person's data, we found 88 judges (80%*) who chose G^- over G^+ 3, 4, or 5 times (including 46* who violated stochastic dominance on

TABLE 2

Number of Judges Who Showed Each Combination of Preferences in Tests of Stochastic Dominance and Transitivity

Variation	$G^- > G^+$	Preference pattern							Mean	
		---	--+	-+-	-++	+-	++-	+++		
1	90	12	33	7	38	0	2	2	16	-\$14.33
2	76	10	31	12	23	3	8	3	20	-\$4.88
3	86	11	42	7	26	2	3	2	17	-\$11.21
4	67	10	32	11	14	5	6	6	26	-\$2.98
5	86	5	43	3	35	0	4	1	19	-\$10.75
Totals	405	48	181	40	136	10	23	14	98	-\$8.83

Note. Column $G^- > G^+$ shows the number of judges (out of 110) who violated stochastic dominance by choosing G^- over G^+ for each variation of the recipe. Minus signs (-) represent preference for G^- over G^+ , G^- over G_0 , and G_0 over G^+ , respectively, contrary to stochastic dominance, and plus signs (+) are used to designate preferences for the dominant gambles. Patterns - + + and + - - are intransitive. Pattern + + + indicates three satisfactions of stochastic dominance. In Variations 2 and 4, left-right positions of the gambles were reversed.

all five choices). Only 22 subjects had 0, 1, or 2 violations (only 5 satisfied stochastic dominance on all five choices).

How much do people offer to get the dominated gamble? When the data are coded so that negative numbers are assigned to violations of dominance and positive numbers to satisfactions of dominance, the mean amount offered for G^+ over G^- was -\$8.83, indicating that more money was offered for dominated (G^-) than dominant gambles. Of 110 subjects, 85 offered more money for dominated gambles, compared to only 25 judges whose means were positive. Mean judgments were negative for all five choices in Table 2. These results are consistent with those of Birnbaum and Navarrete (1998), and extend their results to five new variations of the gambles and new judges.

Minus signs in the column labels of Table 2 represent violations of stochastic dominance in the choices of G^- over G^+ , G^- over G_0 , and G_0 over G^+ , respectively. Each entry in the table shows the number of judges (out of 110) who showed each combination of preferences for each variation of the choice triad. For example, in Row 1, 90 judges violated stochastic dominance in the comparison of G^- and G^+ , 12 violated stochastic dominance on all three comparisons (- - -), and 38 satisfied stochastic dominance on both comparisons with G_0 (- + +), violating transitivity. There were only 16 who satisfied stochastic dominance on all three choices (+ + +) in this triad.

Averaged over variations of the choice (rows), stochastic dominance is violated in comparisons of G^- against G_0 in 47.6% of the trials, and it is violated in the comparison of G_0 against G^+ in 20.4% of the trials. Recall that no single judge is included who violated transparent dominance on more than 2 per 32 tests (6%) and that the average rate of violation was 2.2%. Thus, although violations in choices against G_0 are less frequent than for G^- versus G^+ , they are substantially more frequent than violations of transparent dominance.

Table 2 shows that among those 405 cases (out of 550) where stochastic dominance was violated in the comparison of G^- and G^+ , there are 136 cases where choices satisfied stochastic dominance in both comparisons with G_0 , creating a violation of transitivity ($- + +$). Thus, the conditional probability of violating transitivity, given violation of stochastic dominance in G^- versus G^+ , is .336. However, among the 145 cases of satisfaction of stochastic dominance in G^- versus G^+ , there are only 10 cases of violation of transitivity ($+ - -$), for a conditional probability of .069. The overall percentage of choice triads violating transitivity was 26.5%.

According to WST, if $P(A, B) > 1/2$ and $P(B, C) > 1/2$ then $P(A, C) > 1/2$. Averaged over rows, $P(G^+, G_0) = .796$ and $P(G_0, G^-) = .524$. Therefore, by WST, $P(G^+, G^-)$ should exceed .5; by strong stochastic transitivity, $P(G^+, G^-)$ should exceed .796. Instead, $P(G^+, G^-)$ is only .264. The conditional probability of violating stochastic dominance on G^- versus G^+ given satisfaction of dominance in both comparisons with G_0 is .581, which is still quite high. The conditional probability of violating stochastic dominance on G^- versus G^+ given satisfaction of transitivity in the triad is .67, which is even higher. Therefore, these results suggest that violations of stochastic dominance can produce but are not explained by violations of transitivity.

There were two modal patterns of data among judges who violated stochastic dominance in G^- vs G^+ : the second most frequent pattern is to obey stochastic dominance on both simpler choices, thereby violating transitivity (33.6%). However, the most frequent pattern was also to violate stochastic dominance in the comparison of G^- vs G_0 (181 out of 405, or 44.7%). Additionally, some subjects violated stochastic dominance in the comparison of G_0 vs G^+ (40 of 405 cases) or both of these choices (48 of 405). By violating stochastic dominance more than once, about two-thirds of choice triads satisfy transitivity.

Analyzing individuals, we found 37 judges whose modal preference pattern was the transitive pattern, $- - +$, 28 whose modal pattern was the intransitive combination, $- + +$, and 17 who had the modal pattern, $+ + +$. There were 16 who had multiple modes (of whom 14 had at least one of their modes on $- - +$ or $- + +$); 5 had the modal pattern $- - -$; 4 had the modal pattern $- + -$; 2 had the modal pattern $+ + -$; 1 had the modal pattern $+ - +$; and no one had as their most frequent pattern the intransitive sequence, $+ - -$. Although modal patterns are counted in a mutually exclusive way, choices are not mutually exclusive; for example, 35 judges had some choices in both of these patterns: $- - +$ and $- + +$. In summary, individual patterns echo the group analyses: a minority of 28 judges appear to violate transitivity by violating stochastic dominance in the comparison of G^- against G^+ and by satisfying it in comparisons against G_0 . However, a larger number violated stochastic dominance at least twice in each choice triad, thereby satisfying transitivity.

Tests of Cumulative Independence and Branch Independence

Tests of cumulative independence are shown in Tables 3 and 4. Lower cumulative independence, $S > R \Rightarrow S' > R'$, is refuted by instances of $S > R$ and $S' <$

TABLE 3

Test of Lower Cumulative Independence: $S > R \Rightarrow S' > R'$

x	y	x'	y'	SS'	SR'	RS'	RR'
52	56	11	97	45	31*	17	17
48	52	10	98	42	25	17	26
45	49	11	97	33	33*	15	29
40	44	10	98	20	31*	15	44
35	39	11	97	15	23	27	45
30	34	12	96	16	14	23	57
Totals:				171	157*	114	218

R' , designated SR' in Table 3. Each entry shows the number of subjects who displayed each preference pattern for each variation. Violations of lower cumulative independence, SR' (bold), are more frequent than RS' (consistent with the property), in four of six variations, three of which are significant (asterisks).

Upper cumulative independence, $S' < R' \Rightarrow S'' < R''$, is refuted by cases where $S' < R'$ and $S'' > R''$, denoted $R'S''$ in Table 4. Violations (bold type) are more frequent than satisfactions in all six variations of the gambles used, as well as the total, summed over rows. Four of six are statistically significant (asterisks), tested individually.

Branch independence requires $S > R \Leftrightarrow S' > R'$, and can be tested by inequality of SR' and RS' . Consistent with previous results with other gambles (Birnbbaum & McIntosh, 1996; Birnbbaum & Chavez, 1997; Birnbbaum & Navarrete, 1998), SR' choices are more frequent than RS' : there were 129 SR' choices and 97 RS' choices ($z = 2.13^*$).

Tests of Interval Independence

In six subdesigns that tested interval independence, judges received pairs of gambles that were identical, except that one gamble (A) had a .2 probability to win \$96 instead of \$12. Thus, all differences in expected value (EV) were \$16.80 in these designs. To test interval independence, the interval between \$12 and \$96 was placed in different configurations of outcomes that were

TABLE 4

Tests of Upper Cumulative Independence: $S' < R' \Rightarrow S'' < R''$

x	y	x'	y'	$S'S''$	$S'R''$	$R'S''$	$R'R''$
52	56	11	97	61	14	19	16
48	52	10	98	53	15	20	22
45	49	11	97	45	9	23*	33
40	44	10	98	30	11	34*	35
35	39	11	97	25	4	42*	39
30	34	12	96	21	8	36*	45
Totals:				235	61	174*	190

common to both gambles. Assuming SWU and Eq. (10), judgments in these subdesigns should all be equal (Birnbaum et al., 1997).

Figure 3 shows mean judgments for tests of comonotonic interval independence, in subdesigns where the better gamble (A) had a comonotonic (but sometimes coalesced) .20 probability to receive \$96 instead of \$12 (in B). Mean judgments are shown as a function of the cumulative probability of the contrasting branch in the better gamble, where $P(x < \$96 | A) = 0$ indicates improvement in the lowest outcome and $P(x < \$96 | A) = .8$ indicates improvement of the highest outcome. Separate curves show results with 2, 3, or 5 outcome gambles. EU implies that the curves should coincide in a horizontal line. RDU allows judgments to change as a function of rank, but the curves should coincide (because coalescing should have no effect). CPT with the inverse-S weighting function (Tversky & Kahneman, 1992) implies that the curves coincide and the (single) curve in Fig. 3 should have a U-shape, first decreasing, then increasing.

Figure 3 shows that the data have three features that refute these predictions. First, contrary to EU, the curves are not horizontal, but systematically decrease as the rank of the contrast increases. Improving the worst outcome

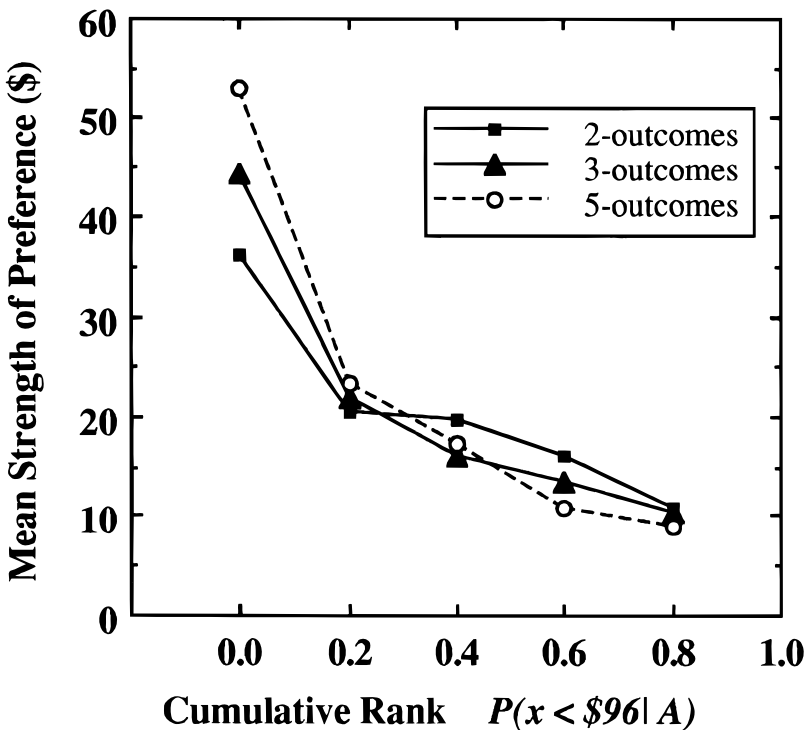


FIG. 3. Tests of interval independence. Mean judgments of strength of preference between gambles A and B , which are identical except the outcome on a .2 branch is \$96 in A instead of \$12 in B . The abscissa shows the cumulative probability that an outcome is less than \$96 in the dominant gamble, A . Interval independence, which is implied by EU theory combined with the subtractive model, requires that all judgments be equal. RDU implies that judgments need not be equal, but should be the same for all rank positions, so the curves should coincide. Instead, the curve for five-outcome gambles has a steeper slope than curves for three- or two-outcome gambles.

of a gamble makes the greatest difference. For example, when four common outcomes are all lower than \$12, improving the highest outcome from \$12 to \$96 is worth an average of only \$8.98, about half of the difference in EV. However, when all four common outcomes exceed \$96, improving the *worst* outcome from \$12 to \$96 is judged to be worth \$52.98, more than twice the EV difference.

Similarly, (\$12, .8; \$96, .2) is judged to be worth (on average) \$10.76 more than \$12 with certainty; however, a sure win of \$96 is judged to be worth \$36.12 more than (\$12, .2; \$96, .8).

This first feature, a decrease in strength of preference as the contrast is increased in rank, is consistent with all three models (RDU, RAM, and TAX); it extends previous results with binary gambles (Birnbbaum et al., 1997). This decrease is characteristic of the majority of individual data; e.g., for 5 outcome gambles, 102* out of 110 judges showed a decrease in judgments as $P(x < \$96 | A)$ is increased from 0 to .8, 5 showed an increase, and the rest were tied. Similar results were obtained with two- and three-outcome gambles, and the decrease was also characteristic of the majority when $P(x < \$96 | A)$ is increased from .2 to .6, averaged over data in Fig. 3.

The second feature of Fig. 3 is that the curves do not coincide, contrary to RDU. The interval between (\$2, \$4, \$6, \$7, \$96) and (\$2, \$4, \$6, \$7, \$12) is judged less than the interval between (\$12, .8; \$96, .2) and \$12 for sure. The interval between (\$96, \$108, \$111, \$113, \$115) and (\$12, \$108, \$111, \$113, \$115) is judged more than the interval between \$96 for sure and (\$12, .2; \$96, .8). The slopes for three-outcome gambles are intermediate between two and five. Comparing $P(x < \$96 | A) = 0$ versus $P(x < \$96 | A) = .8$, 76 judges showed a steeper negative slope for five-outcome gambles than for two-outcome gambles, 29 showed the opposite, and 5 showed no difference. Comparing $P(x < \$96 | A) = .2$ versus .6, 67* showed a greater decrease for five-outcome gambles than for two-outcome gambles, 28 showed the opposite, and 15 showed no difference.

Third, mean judgments do not increase between $P(x < \$96 | A) = .6$ and .8, contrary to the inverse-S weighting function of Tversky and Kahneman (1992). Averaged over choices with comparisons of $P(x < \$96 | A) = .6$ against .8, there were 65* subjects who showed a decrease, 23 who showed an increase, and 22 who showed no change. A significant majority show continued decrease, contrary to the U prediction of the inverse-S weighting function.

The TAX model allows that the curves in Fig. 3 can differ for different numbers of outcomes. With $\gamma = .7$ and $\delta = -1$, the predicted intervals for $n = 5$ should strictly decrease, consistent with the data; however, the curve for $n = 2$ should decrease and then increase again at the upper end, similar to the predictions for CPT with $n = 2$. Therefore, although the TAX model is in better agreement with the data than CPT, the failure of the curve for $n = 2$ to increase at the upper end is not in accord with the TAX model, if $\gamma < 1$. The effect of n is also not consistent with $\gamma < 1$.

Figure 4 shows mean judgments between three-outcome gambles as a function of rank position (cumulative probability), with separate curves for different

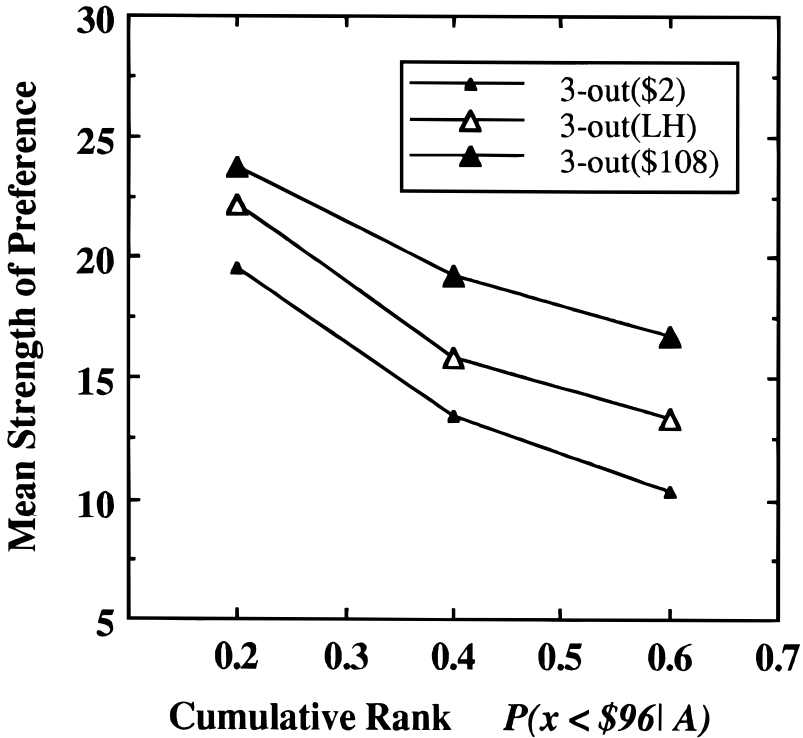


FIG. 4. Tests of interval independence with three-outcome gambles, plotted as in Fig. 3. Large filled triangles show judgments when common outcome is the highest (\$108), and the interval represents an improvement in the lower outcomes. Small filled triangles show judgments when the common outcome is the lowest (\$2). Unfilled triangles show judgments when there are common outcomes that are both higher and lower than the contrast.

common outcomes. The upper curve, shown as large filled triangles, shows data when the common outcome was \$108 (highest). The lowest curve (small filled triangles) shows data when the common outcome was \$2 (lowest). The middle curve (unfilled triangles), shows results when there are common outcomes that are both lower and higher, and the dominant gamble has a .2 branch with an outcome of \$96 instead of \$12. According to EU theory, the curves should be horizontal and they should all coincide. According to RDU and CPT, the curves need not be horizontal, but all three curves should coincide.

Instead, the curves in Fig. 4 decrease in all cases, and they do not coincide. Vertical gaps between the curves, which refute RDU, are representative of the majority. For example, out of 110 individuals, 85 gave greater mean judgments [averaged over $P(x < \$96 | A) = .2, .4, \text{ and } .6$] when the common outcome was \$108 than when it was \$2, 20 showed the opposite, and 5 showed no difference. Other vertical gaps in the means of Fig. 4 were also representative of the majority of individuals.

Subdesigns 8 and 9 yielded similar conclusions. For example, judges evaluated the gamble (\$10, .85; \$99, .15) to be worth an average of \$11.70 more than (\$10, .95; \$99, .05); however, they judged the gamble (\$10, .05; \$99, .95) to be worth \$25.03 more than (\$10, .15; \$99, .85). When the same contrast (a .10

probability to win \$99 instead of \$10) was expressed as a change in outcome value in a three-outcome gamble, the effect was more extreme, consistent with the change of slopes in Fig. 3. For example, (\$3, .85; \$99, .10; \$107, .05) was judged to be worth \$9.08 more than (\$3, .85; \$10, .10; \$107, .05); however, (\$3, .05; \$99, .10; \$107, .85) was judged to be worth \$26.75 more than (\$3, .05; \$10, .10; \$107, .85).

Fit of TAX and CPT Models

Following Birnbaum and Chavez (1997), we fit TAX and CPT models to each judge's choices, excluding choices with transparent or "translucent" dominance (comparisons with G_0). (CPT and TAX both satisfy transitivity and transparent dominance). Models were fit to a compromise of the negative log likelihood of the choices given the model and the sum of squared discrepancies between the model and judgments. Both models were fit with $u(x) = x$. Median parameters for the TAX model are $\gamma = .793$ and $\delta = -.751$. For CPT, median parameters are $\gamma = .962$ and $c = .207$. The negative log likelihood was significantly lower for the TAX model than the CPT model, $t(109) = 3.47^*$, and the mean sum of squared deviations was also lower, but not significantly. Fitting the mean judgments, the comparison of fit was more extreme. The parameters fitting mean judgments are $\gamma = 1.06$ and $\delta = -.796$ for TAX and $\gamma = .852$ and $c = .434$ for CPT. The negative log likelihood for TAX was .01 with a sum of squared deviations of 435.1, much better than the values for CPT, 16.54 and 4389.6, respectively.

DISCUSSION

Violations of stochastic dominance and of cumulative independence are to the class of RDU/RSDU/CPT models as the Allais paradoxes are to EU. Even with freedom to select any $u(x)$ and $W(P)$ functions, no member of this class of models can account for violations of these three properties. The present study adds five new variations of the recipe that produces violations of stochastic dominance. This study also adds to the evidence concerning violations of cumulative independence by using six variations not previously tested. These data reinforce and extend the findings of Birnbaum and Navarrete (1998).

Because these results are predicted by the RAM and TAX models with previously estimated parameters, our conclusion is not merely that configural weight models are flexible enough to be consistent with odd patterns of data. Instead, the conclusion is that these models with their previously estimated parameters made dramatic predictions for new experiments—and when these experiments were carried out, these unusual predictions were confirmed by the empirical choices.

The present study also investigated whether judges would correctly obey stochastic dominance when choosing between three-outcome gambles and two-outcome gambles. We found that in about two-thirds of triads in which judges

violated stochastic dominance in the comparison between three-outcome gambles, judges also violated stochastic dominance on at least one of the two simpler comparisons, thereby satisfying transitivity. Although violations of transitivity represent a minority of choices (about one-fourth of all triads), they suffice to produce violations of both strong and weak stochastic transitivity.

Tests of interval independence indicated systematic violations. Judgments of intervals were greatest when the lowest outcome of a gamble was improved. Contrary to the inverse-S weighting function, improvement of the middle outcome was not the smallest interval; instead, judged intervals decreased monotonically as the rank of the (contrasting) outcome increased.

Summary of Evidence against RDU/RSDU/CPT

Table 5 analyzes five tests that refute the class of RDU/RSDU/CPT models. These five complex tests have in common that they can be deduced from coalescing, combined with other simple assumptions. Although the class of RSDU/RDU/CPT models must satisfy these tests, RAM and TAX models violate coalescing and predict systematic violations.

Cumulative independence. The properties of lower and upper cumulative independence can be derived from transitivity, monotonicity, coalescing, and comonotonic branch independence. Both properties were systematically violated in Tables 3 and 4, as predicted by RAM and TAX models.

Tail independence. The property of upper tail independence (ordinal independence) tested by Wu (1994) can be derived from comonotonic branch independence, coalescing, and transitivity. Wu (1994) noted that his results, which

TABLE 5

Analysis of Five Tests that Distinguish Configural Weight Models from RDU models

	Simpler properties				Models	
	T	C	M	CBI	RDU	CWT
Complex tests						
Event-splitting	X	X			S	V
Stochastic dominance	X	X	X		S	V
Tail independence	X	X		X	S	V
Lower cumulative independence	X	X	X	X	S	V
Upper cumulative independence	X	X	X	X	S	V
Models						
RDU	S	S	S	S		
CWT (RAM and TAX)	S	V	S	S		

Note. T, transitivity; C, coalescing; M, monotonicity; CBI, comonotonic branch independence. S, satisfied; V, violated. X indicates that the simpler property can be used to derive the complex property; in each case the complex property can be derived from the combination of simpler properties marked with X. For example, stochastic dominance should be satisfied if choices satisfy transitivity, coalescing, and monotonicity. RDU satisfies stochastic dominance, and CWT violates it.

violated tail independence, were inconsistent with the family of RDU/RSDU/CPT models. Wu combined editing principles with the CPT model to describe his data. He proposed that when common branches are “transparent” (i.e., not coalesced with other outcomes), judges cancel common branches; in contrast, when common branches are coalesced with other branches, judges do not cancel them. Thus, although his explanation is different, Wu focused on the role of coalescing in producing the violations from the RDU models. The model of Wu (1994), however, implies that choices should satisfy branch independence and distribution independence, contrary to data of Birnbaum and McIntosh (1996) and Birnbaum and Chavez (1997).

Branch independence and the inverse-S. Research on branch independence and distribution independence (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997) found violations of these properties that were opposite those predicted by the inverse-S weighting function of Tversky and Kahneman (1992). The present data also extend the pattern reported previously to new gambles.

Stochastic dominance. Stochastic dominance follows from transitivity, coalescing, and monotonicity. Violations refute any theory that assumes these three principles. This study tested whether violations of stochastic dominance can be explained by violations of transitivity in choice triads including G^- , G^+ , and G_0 . The data show systematic violations of transitivity by a minority of the judges, but they also show that the conditional probability of violation of stochastic dominance in G^- versus G^+ given satisfaction of transitivity is .67, which is still quite high. This finding suggests that violations of stochastic dominance were not produced by violations of transitivity.

Event-splitting effects. Event-splitting effects refer to reversals of preference when the same choice is presented in split or coalesced format (Starmer & Sugden, 1993; Humphrey, 1995). Event-splitting effects violate RDU/RSDU/CPT models, but can be explained by configural weight models. Birnbaum (1998b) tested for violations of both stochastic dominance and event-splitting effects, using undergraduates who were given a chance to play one of their chosen gambles for real money. One of the choices was between G^- and G^+ as follows:

G^+ : .05 probability to win \$12	G^- : .10 probability to win \$12
.05 probability to win \$14	.05 probability to win \$90
.90 probability to win \$96	.85 probability to win \$96

Consistent with the present data, most favored G^- over G^+ , violating dominance. However, a majority of the same judges reversed preferences in the following choice:

GS^+ : .05 probability to win \$12	GS^- : .05 probability to win \$12
.05 probability to win \$14	.05 probability to win \$12
.05 probability to win \$96	.05 probability to win \$90
.85 probability to win \$96	.85 probability to win \$96

GS^+ is the same as G^+ , and GS^- is the same as G^- , except for coalescing. These event-splitting effects refute the combined assumptions of coalescing

and transitivity. The TAX model (with $u(x) = x$, $\gamma = .7$, and $\delta = -1$) implies the following CEs for these gambles: $CE(G+) = \$45.77 < CE(G-) = \63.10 but $CE(GS+) = \$53.06 > CE(GS-) = \51.38 , consistent with the data.

According to any special case of RDU/RSDU/CPT, the dominant gamble should have the same higher value whether it is split or not. These results with event-splitting, combined with the present exoneration of transitivity as the cause of violations of stochastic dominance, are consistent with the argument that violations of stochastic dominance and cumulative independence are most likely due to violation of coalescing.

Critical Evaluation of Editing Principles of Prospect Theory

Both versions of prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) used editing principles in addition to evaluation rules. Wu (1994), Leland (1994), and Starmer (in press) have made more precise specifications of editing principles, to make these notions testable. The present data bear on three of the editing principles. Evidence is inconsistent with the principles of cancelation and combination, but there may be some evidence for a dominance detector.

Cancelation. The editing principle of cancelation implies that common components should have no systematic effect on preferences. Therefore, this principle cannot explain systematic violations of branch independence and distribution independence.

Violations of interval independence also constitute evidence against cancelation. About half of all choices in this study involved comparisons in which judges could have edited out and canceled common branches. If subjects had canceled common branches, then the curves in Figs. 3 and 4 would have coincided in a horizontal line. Instead, the data show that it is worth more to improve the worst outcome than to improve the best outcome. If CPT gives up the cancelation principle, then the inverse-S weighting function predicts that the (single) curve in Fig. 3 should have been U-shaped. The failure of the curves in Figs. 3 and 4 to coincide shows that judged intervals depend on features of the gambles that are supposed to have no effect according to RSDU/RDU/CPT models (with any $W(P)$ function). These violations may be due to violations of coalescing.

Perhaps there is a way to modify the idea of cancelation to make it more compatible with data (e.g., perhaps common branches cause distinct branches to make a greater difference), but the simple theory of cancelation cannot explain violations of distribution independence, branch independence, and interval independence. It is also hard to see how one could explain violations of branch independence in both judgment and choice with editing principles.

Combination. The editing principle of combination implies coalescing. It predicts that judges should combine the equal outcomes of $GS-$ and $GS+$ before choosing, which would convert that choice into the choice between $G-$ and $G+$. Event-splitting effects contradict this editing principle (Starmer & Sugden, 1993).

Dominance detection. Kahneman and Tversky (1979; Tversky & Kahneman, 1986) proposed that subjects will conform to dominance when the relation is “transparent,” but not when it is “masked” by the problem frame. The present results show that violations of stochastic dominance can occur even without any framing manipulation.

The concept of transparency is not completely clear, but presumably it is easier to see dominance relationships between the two-outcome gamble, G_0 , and the three-outcome gambles than between G^+ and G^- . If tests of outcome monotonicity and probability monotonicity are termed “transparent,” and if the choice between G^- and G^+ is termed “opaque,” then perhaps we should call choices between G^- and G_0 and between G^+ and G_0 “translucent,” since their rates of violation are intermediate between the other two. This “visibility” metaphor may help us name and remember the results, but it does not really explain them.

The results can be summarized as follows: (1) There are 2.2% violations of transparent dominance (outcome monotonicity or probability monotonicity). (2) Comparing G^+ and G_0 , there are 20.4% violations. (3) Comparing G^- and G_0 there are 47.6% violations. (4) Comparing G^+ and G^- there are 73.6% violations.

All of the models compared here are transitive, yet the data show violations of transitivity. We need to explain why people violated stochastic dominance so often in the comparison of G^+ and G^- , and why they did *not* violate stochastic dominance more often than they did in the comparisons of G^- and G^+ against G_0 . The TAX model predicts that $G^- > G_0 > G^+$. If we theorize that choice probabilities are a function of utility differences, then the TAX model correctly implies that the greatest choice proportion should be between G^- and G^+ , since they are most different in utility; however, the model also implies that the choice proportions for $P(G_0, G^+)$ and $P(G^-, G_0)$ should have both exceeded 1/2, which they did not.

To make an editing notion predictive, we need to specify its mechanism (Leland, 1994; Wu, 1994; Starmer, in press). We need a clearer definition of “translucence.” Suppose a dominance detector works by first comparing the values of corresponding outcomes (“corresponding” means the outcomes have the same ranks, when ranked by discrete values) and then comparing probabilities of equal outcomes. In cases where these two comparisons do not conflict, the choice is “transparent,” and the judge chooses the dominant gamble. However, when one gamble has at least one higher outcome, but the other gamble has a larger probability for an equally high outcome, the case seems “mixed,” and the judge chooses the gamble with the higher $U(G)$. Because the comparison of G^+ and G^- is “mixed” to this detector, the judge will violate stochastic dominance when $U(G^-) > U(G^+)$, which it can be under either RAM or TAX model.

In the case of choices between two- and three-outcome gambles, the middle outcome of the three-outcome gamble has no corresponding middle outcome in G_0 . Suppose the judge compares the middle outcome half the time to the higher outcome of G_0 and half the time to the lower outcome of G_0 ; if so, then half of

these “translucent” cases will be detected as dominant, and half of these cases will appear as mixed. Such a mechanism would partially satisfy stochastic dominance in these “translucent” cases.

One obvious question that remains is as follows: Why are there twice as many violations of stochastic dominance in G^- versus G_0 as in G^+ versus G_0 ? In all cases of $G_0 = (x, p; y, 1 - p)$ studied thus far, the lowest outcome also has the lowest probability; i.e., $p < 1 - p$. Thus, the comparison of G^- versus G_0 involves splitting a small piece off from a relatively large probability. For example, $G_0 = (\$2, .04; \$98, .96)$ and $G^- = (\$2, .05; \$93, .02; \$98, .93)$. In contrast, $G^+ = (\$2, .02; \$6, .02; \$98, .96)$ differs from G_0 by a splinter that is half the probability of the lower outcome. It is possible that the effect of a splinter of probability obeys Weber’s law. However, different variations of the recipe must be investigated empirically to determine if this asymmetry is due to asymmetric effects of splitting the higher or lower outcome, or of splitting the larger or smaller probability.

A recent study by Starmer (in press) reported violations of transitivity in 25% of choice triads that may also be attributable to dominance detection. Starmer used the following recipe: $A = (0, 1 - p; y, p)$, $B = (0, 1 - q; x, q)$, and $C = (0, 1 - q - r; y^-, r; y, q - r)$, where $y > y^- > x > 0$. The comparison of A and C involves event-splitting, since y is split into y and y^- , which was intended to make C more attractive relative to A . Starmer (in press) found that 64.7% preferred A to B , 93.6% preferred the dominant B to C , but only 48.5% preferred A to C . (The rate of satisfaction of dominance is much higher in this study than ours, since B and C were juxtaposed in a format that made the stochastic dominance relation equivalent to transparent outcome monotonicity.) Transitivity is such an important theoretical property that evidence of systematic deviation should be pursued.

Editing versus Configural Weighting Explanations

One might attempt to explain violations of stochastic dominance in choice as the result of simplification and cancelation (Kahneman & Tversky, 1979; Leland, 1994), as follows. Suppose the judge is presented with the following choice:

$$G^+ = (\$12, .05; \$14, .05; \$96, .9) \quad \text{versus} \quad G^- = (\$12, .1; \$90, .05; \$96, .85).$$

Suppose the judge noticed that outcomes \$12 and \$96 are the same in both gambles and canceled these common outcomes (of approximately equal probabilities). That leaves G^- with a 5% chance of \$90 as opposed to G^+ with a 5% chance at \$14. This heuristic might decide that gamble G^- (which is actually dominated by G^+) is better than G^+ .

However, this heuristic would not explain why there should be violations of stochastic dominance in judgment experiments, where G^+ and G^- are presented on separate trials. Birnbaum and Yeary (1997) found that judgments of both buying prices and selling prices violated stochastic dominance. For

example, the median buying prices of $G^+ = (\$12, .05; \$14, .05; \$96, .9)$ and $G^- = (\$12, .1; \$90, .05; \$96, .85)$ were \$30 and \$60, respectively, a \$30 violation! Median judgments of selling prices were \$73.5 and \$83.5, respectively, also violating dominance. Because gambles are presented individually in judgment, separated by many intervening trials, it is difficult to see how this cancellation strategy would account for violations of stochastic dominance in judgment.

It seems simpler to theorize that the combination rule is a configurally weighted average, and that this same model applies to choices and to judgments of buying and selling prices, with different configural weights. This theory predicts violations of stochastic dominance, branch independence, cumulative independence, and event-splitting effects in both judgment and choice.

Intuitions and Implications of Configural Weighting

In both configural weight models, if $S(p) = p^\gamma$, where $\gamma < 1$, then the utility of a two-outcome gamble will be more “sensitive” to changes in probability near zero and near one than near 1/2. This pattern is also implied by the inverse-S weighting function of CPT. All three models can account for risk seeking for low probabilities of positive outcomes, and for risk aversion for medium and high probabilities of positive outcomes. In the configural weight models, these properties are consequences of the relative weight expression in an averaging model and the negatively accelerated psychophysical function for probability.

The equations of configural weighting may appear complex, but they can be understood by very simple intuitions: each discrete outcome carries some weight, and these weights are affected by relationships among the outcomes and the probability distribution over the outcomes. One can think of configural weights as measures of attention given each separate probability–outcome branch of a gamble. In both the RAM and TAX models, the relative weight of an outcome increases as the probability of the outcome increases, and it decreases as a function of the total weight of other outcomes. In this intuition, configural weighting differs from RSDU/RDU/CPT models, which assume that it doesn’t make any difference how branches are presented.

Unlike CPT and RSDU models, both of the configural weight models imply violations of coalescing. For $\delta < 0$ and $\gamma < 1$, splitting an outcome of probability p into two discrete outcomes of probability q and $p - q$ has the effect of increasing that outcome’s relative weight. (Other functions, such as $S(p) = ap + b$, with $b > 0$, would also have this property for small p .) All of the violations of RDU reviewed in Table 5—violations of stochastic dominance, violations of lower and upper cumulative independence, violations of tail independence, and event-splitting effects—can be attributed to violations of coalescing. Intuitively, an outcome can gain weight when it is split into two or more branches. The increase in weight due to event-splitting can even overcome the effects of monotonicity (increasing the value of an outcome), as is the case in violations of stochastic dominance and cumulative independence.

Both RAM and TAX models allow the weight of an outcome to be affected by the rank of the outcome among the other outcomes in the same set; this

effect of ranks on the weights is similar to rank dependence in RDU models. However, unlike RDU models, in which cumulative weight is a monotonic function of cumulative probability, both configural models assume that the ranks of the outcomes are the ranks of the *values* of discrete outcomes, not their cumulative probabilities.

Birnbaum and Stegner (1979) discuss the theory that configural weights are produced by motivational effects of asymmetric costs of overestimation as opposed to underestimation. This account is similar to that used in signal detection theory, in which differential rewards and costs for hits and false alarms can be manipulated to produce different responses to the same stimulus. If it is more costly to overprice (than undervalue) a gamble, then more weight should be placed on outcomes of lower value. Buyers appear to worry more about setting too high a value on a gamble, and sellers worry about judging its value too low. This theory gives an intuitive account, or rationale, for changing configural weights produced by instructions to identify with the buyer or the seller.

Previous research has shown that configural weight models can account for the differences in preference order between buying and selling prices and between judged prices and choices. The theory assumes that configural weights depend on the judge's point of view (i.e., instructions to identify with the buyer, seller, or a neutral, or the instruction to choose). The theory also assumes that the utility function, $u(x)$, and the probability function, $S(p)$, are independent of viewpoint and task (Birnbaum & Beeghley, 1997; Birnbaum & McIntosh, 1996; Birnbaum & Zimmermann, 1998). The agreement of the estimates of $u(x)$ between different viewpoints provides a test of configural weight models (Birnbaum & Sutton, 1992; Birnbaum et al., 1992). Although judgments of buying and selling prices are not monotonically related, and these also are distinct from the preference order inferred from choice, the configural weight model can reproduce these different preference orders with assumption that utility is proportional to money. The ability of the configural weight models to account for preference reversals, violations of branch independence, distribution independence, stochastic dominance, cumulative independence, and the Allais paradoxes with the same utility function seems an attractive feature of these models (Birnbaum, 1998a).

The effect of the judge's viewpoint in Birnbaum and Stegner (1979; i.e., the discrepancy between buying and selling prices) was later termed the "endowment effect" in the economic literature of the 1980s (reviewed in Kahneman, Knetsch, & Thaler, 1991). In this literature, the notion of "loss aversion" was suggested as a possible explanation of endowment effects. Birnbaum and Zimmermann (1998) compared configural weight theories against "loss aversion" and "anchoring and adjustment" theories of this effect. Their analysis shows that when these notions are stated explicitly and combined with assumptions used in CPT, they do not account for the data in the published literature. Therefore, configural weight models can account for phenomena in both judgment and choice that apparently cannot be reconciled with the model of CPT.

Although both the RAM and TAX models predict violations of stochastic

dominance and cumulative independence in this study, RAM and TAX models differ from each other with respect to the following concept. In the RAM model, weights are a function of the rank and probability of a discrete outcome, and relative weights are weights divided by the sum of these rank-affected weights. In the TAX model, in contrast, each outcome gains weight by taking weight from other outcomes. Each rank position can have some ability to *tax*, or pull weight from other items, and the weight transferred is proportional to the amount of weight that each of the other items has to lose.

The special case of the TAX model, as fit by Birnbaum and Chavez (1997) and used here to predict the results, assumes further that lower outcomes always take weight from higher ranked outcomes by taxing them at the same rate. If this tax rate is $\delta = -1/(n + 1)$, then for $n = 2$, the tax rate is $1/3$; therefore, the lower of two equally likely outcomes ($p = .5$) will take one-third of the weight of the higher outcome, so the relative weights will be $2/3$ and $1/3$, with the lower outcome having twice the weight of the higher. For three equally likely outcomes, the weights would be $3/6$, $2/6$, and $1/6$ for the lowest, middle, and highest outcomes; four equally likely outcomes will have weights of $4/10$, $3/10$, $2/10$, and $1/10$, for the lowest through highest outcomes, respectively.

This special case model implies that the ratio of weights of the lowest to highest outcomes will increase as the number of outcomes increases, and it also implies that configural weights of equally likely outcomes will be a monotonic function of the ranks of the discrete outcomes. Birnbaum and Veira (1998) and Birnbaum and Zimmermann (1998) found evidence against this implication for judgments of value of gambles from the seller's point of view. Sellers appear to place more weight on higher than lower outcomes, but they appear to place most weight on middle outcomes. Therefore, in judgment, a more complex configural model is required to fit the data than the theory that the weight transfers are all at the same rate. The next level of complication is to allow each rank position to have its own tax rate.

The RAM and TAX models are nearly identical in their predictions for this study, but they can be distinguished by other experiments. As noted earlier, the RAM model implies distribution independence, which is violated by choices (Birnbaum & Chavez, 1997). Violation of distribution independence in choice is the strongest argument to date against the RAM model.

The TAX model with $S(p) = p^\gamma$ and $\delta < > 0$, violates asymptotic independence, unlike the RAM model with the same assumptions. Asymptotic independence asserts that as an outcome's probability goes to 0, the value of that outcome should become irrelevant. Violation of this property can be illustrated by a situation in which the value of the lowest outcome does not become irrelevant as p approaches 0, but places an upper bound on the utility of the gamble, as long as that lowest outcome is possible. For example, with $\gamma = 1$ and $\delta = -1$, the TAX model implies that $CE(0, p; \$1000, 1 - p)$ asymptotically approaches \$666.67 as p approaches (but does not equal) 1. In the TAX model, the role of insurance is to increase the lowest outcome. Even with utility proportional to money, insurance can be worth a great deal in the TAX model, so insurance

companies can sell premiums at a profit, and both buyer and seller are pleased with the transaction.

CONCLUSIONS

These data show systematic violations of stochastic dominance and cumulative independence. Violations of these properties are inconsistent with any RSDU model, including RDU and CPT. Violations of interval independence also show evidence against RDU models, and several aspects of the data (violations of branch independence and tests of interval independence) show trends inconsistent with the inverse-S weighting function. Results are not consistent with the editing principles of cancelation (which implies branch independence and interval independence) or combination (which implies coalescing). The major violations observed in this paper may be due to violations of coalescing. The data are better described by the TAX model of configural weighting, which violates coalescing, than by the RDU/RSDU/CPT class of models, which satisfy coalescing. Violations of stochastic dominance appear to produce violations of transitivity in about one-fourth of the choice triads. None of these models can account for intransitivity, which may require an editing principle to explain why choices partially satisfy and partially violate stochastic dominance in “translucent” cases.

APPENDIX: CUMULATIVE INDEPENDENCE

Any theory that satisfies comonotonic independence, monotonicity, transitivity, and coalescing must satisfy both lower and upper cumulative independence (Birnbaum, 1997; Birnbaum & Navarrete, 1998). For upper cumulative independence, if $S' < R'$ then $(x, p; y, q; y', r) < (x', p; y', q; y', r)$, by comonotonic independence. By monotonicity $(x, p; x, q; y', r) < (x, p; y, q; y', r) < (x', p; y', q; y', r)$; hence, $(x, p; x, q; y', r) < (x', p; y', q; y', r)$, by transitivity. Finally, $(x, p + q; y', r) < (x', p; y', q + r)$, by coalescing, which is the same as $S'' < R''$, Q.E.D. The next proof shows that cumulative independence follows directly from Eq. (1). It shows that violation of cumulative independence can be interpreted as a contradiction in the weighting function within RDU.

Lower cumulative independence: $S > R \Rightarrow S' > R'$. Proof by contradiction. Suppose $S' = (x', r; y, p + q) < R' = (x', r + p; y', q)$. From RDU representation (Eq. (1)),

$$\begin{aligned} W(p + q)u(y) + [1 - W(p + q)]u(x') &< W(q)u(y') + [1 - W(q)]u(x') \\ [W(p + q) - W(q) + W(q)]u(y) + [1 - W(p + q)]u(x') & \\ &< W(q)u(y') + [1 - W(q) + W(p + q) - W(p + q)]u(x') \\ W(q)u(y) + [W(p + q) - W(q)]u(y) + [1 - W(p + q)]u(x') & \\ &< W(q)u(y') + [1 - W(p + q)]u(x') + [W(p + q) - W(q)]u(x') \end{aligned}$$

$$\begin{aligned}
W(q)u(y) + [W(p + q) - W(q)]u(y) &< W(q)u(y') + [W(p + q) - W(q)]u(x') \\
W(q)u(y) + [W(p + q) - W(q)]u(x) &< W(q)u(y') + [W(p + q) - W(q)]u(x') \\
[W(p + q) - W(q)]u(x) - [W(p + q) - W(q)]u(x') &< W(q)u(y') - W(q)u(y) \\
[W(p + q) - W(q)][u(x) - u(x')] &< W(q)[u(y') - u(y)] \\
[W(p + q) - W(q)]/W(q) &< [u(y') - u(y)]/[u(x) - u(x')]. \quad (12)
\end{aligned}$$

Now suppose $S = (z, r, x, p; y, q) > R = (z, r, x', p; y', q)$. From RDU representation,

$$\begin{aligned}
W(q)u(y) + [W(p + q) - W(q)]u(x) + [1 - W(p + q)]u(z) \\
> W(q)u(y') + [W(p + q) - W(q)]u(x') + [1 - W(p + q)]u(z) \\
W(q)u(y) + [W(p + q) - W(q)]u(x) &> W(q)u(y') + [W(p + q) - W(q)]u(x') \\
[W(p + q) - W(q)]u(x) - [W(p + q) - W(q)]u(x') &> W(q)u(y') - W(q)u(y) \\
[W(p + q) - W(q)][u(x) - u(x')] &> W(q)[u(y') - u(y)] \\
[W(p + q) - W(q)]/W(q) &> [u(y') - u(y)]/[u(x) - u(x')], \quad (13)
\end{aligned}$$

which contradicts expression (12), the implication of $S' < R'$, thus proving the proposition. The contradiction between expressions (12) and (13) is analogous to the contradiction between the results of Birnbaum and McIntosh (1996) and of Wu and Gonzalez (1996) under the assumption of CPT. Upper cumulative independence can also be derived from Eq. (1) using the same approach. Violations of cumulative independence can be interpreted as a contradiction between the weighting functions in RDU for the cases of $n = 2$ outcomes and $n = 3$ outcomes.

REFERENCES

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole Americaine. *Econometrica*, **21**, 503–546.
- Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School. In M. Allais & O. Hagen (Eds.), *Expected utility hypothesis and the Allais paradox* (pp. 27–145). Dordrecht, The Netherlands: Reidel.
- 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. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Erlbaum.
- Birnbaum, M. H. (1998a). Paradoxes of Allais, stochastic dominance, and decision weights. In J. C. Shanteau, B. A. Mellers, & D. Schum (Eds.), *Decision Research from Bayesian approaches to normative systems: Reflections on the contributions of Ward Edwards*. Norwell, MA: Kluwer.

- Birnbaum, M. H. (1998b). *Violations of stochastic dominance and cumulative independence with financial incentives*. Manuscript submitted for publication.
- 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., & Jou, J. W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology*, **22**, 184–210.
- 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., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, **45**, 792–804.
- Birnbaum, M. H., & Navarrete, J. (1998). Testing rank- and sign-dependent utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, **17**, 49–78.
- Birnbaum, 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). *Violations of stochastic dominance, cumulative independence, branch independence, coalescing, event-splitting independence, and asymptotic independence in buying and selling prices of gambles*. Unpublished manuscript.
- 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.
- Camerer, C. F. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 207–251). Boston: Kluwer.
- Champagne, M., & Stevenson, M. K. (1994). Contrasting models of appraisal judgments for positive and negative purposes using policy modeling. *Organizational Behavior and Human Decision Processes*, **59**, 93–123.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, **51**, 380–417.
- Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, **69**, 109–135.
- Fishburn, P. C. (1978). On Handa's "New theory of cardinal utility" and the maximization of expected return. *Journal of Political Economy*, **86**, 321–324.
- Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty*, **11**, 263–274.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Experimental tests of the endowment effect and the coarse theorem. In R. H. Thaler (Ed.), *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.
- Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, **21**, 61–72.
- 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 (Ed.), *Acting under uncertainty* (pp. 267–299). Boston: Kluwer.
- 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. (1998). Coalescing, event commutativity, and theories of utility. *Journal of Risk and Uncertainty*, **16**, 87–113.
- 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.
- Miyamoto, J. M. (1989). Generic utility theory: Measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical Psychology*, **32**, 357–404.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, **3**, 324–345.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Starmer, C. (in press). Cycling with rules of thumb: An experimental test for a new form of non-transitive behaviour. *Theory and Decision*.
- Starmer, C., & Sugden, R. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypotheses. *Annals of Operations Research*, **19**, 79–101.
- 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. (1969). Intransitivity of preferences. *Psychological Review*, **76**, 31–48.
- Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, **102**, 269–283.
- 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.
- Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 613–625.
- 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. (1996). The sure-thing principle and the comonotonic sure-thing principle: An axiomatic analysis. *Journal of Mathematical Economics*, **25**, 213–227.
- Weber, E. U. (1994). From subjective probabilities to decision weights: The effects of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, **114**, 228–242.

- 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.

Received: December 29, 1997