





RDU/RSDU/CPT theories, but it can be deduced from them.

The configural weight, RAM and TAX models imply that both coalescing and restricted branch independence are systematically violated (except in special cases where these models reduce to EU); like the rank-dependent models, these models satisfy restricted comonotonic branch independence.

This paper will compare the theories in Table 1 by separating tests of branch independence from those of coalescing in Allais common consequence paradoxes.

### 1.3. SWU and OP models

One way to describe Allais paradoxes is to replace objective probabilities with subjective weights (Edwards, 1962) as follows:

$$SWU(G) = \sum_{i=1}^n w(p_i)u(x_i), \tag{3}$$

where  $SWU(G)$  is the subjectively weighted utility (SWU) of gamble  $G$ . In this model, the weight of a given objective probability is a function of its probability. In this model, there is no contradiction in choosing  $A$  over  $B$  and  $D$  over  $C$ . In particular, if  $w(p)$  is an inverse-S function of  $p$ , paradoxes of Allais can be described by Eq. (3).

However, Eq. (3) implies that people will violate transparent dominance in ways that few humans would do (Fishburn, 1978). For example, with parameters chosen to fit the Allais paradoxes, Eq. (3) implies that people should prefer  $E = (\$103, 0.98; \$102, 0.01; \$101, 0.01)$  over  $F = (\$120, 0.5; \$110, 0.5)$ , despite the fact that the lowest consequence of  $F$  is better than the best consequence of  $E$  (see Birnbaum, 1999a).

In their Original Prospect (OP) model, Kahneman and Tversky (1979) restricted Eq. (3) to gambles with no more than two nonzero consequences (which puts Gamble  $E$  and the four gambles of Expression 2 outside the domain of OP). They also added editing rules to OP in order to avoid certain implausible implications of Eq. (3). For example, people are assumed to detect and conform to transparent dominance. Three other editing principles of OP are relevant to this paper: (1) *Combination* assumes that people combine branches with identical consequences by adding their probabilities (which implies coalescing). (2) *Cancellation* postulates that people will cancel elements that are identical in two gambles of a choice, implying restricted branch independence. (3) *Simplification* is an editing rule where people round off and ignore small differences, which facilitates cancellation or combination of nearly equal branches. Starmer and Sugden (1993) refer to “stripped” prospect theory as Eq. (3) without the editing rules of OP, which can be extrapolated to include three branch gambles.

### 1.4. Rank-dependent expected utility models

Quiggin (1985, 1993) proposed RDU theory, which accounts for the Allais paradoxes without violating stochastic dominance. Luce and Narens (1985) developed a dual bilinear representation for two branch gambles that generalizes the original form of Quiggin, which required the weight of 1/2 to be 1/2. RSDU was later proposed, which generalized the rank-dependent approach to allow different weightings for positive and negative consequences (Luce & Fishburn, 1991, 1995; Luce, 2000). Tversky and Kahneman (1992); Tversky and Wakker (1995); Wakker and Tversky (1993) proposed CPT, which combined rank- and sign-dependent weighting with the editing principles of OP (see also Starmer & Sugden, 1989). All of these rank-dependent theories have the same representation for gambles composed of strictly positive consequences:

$$RDU(G) = \sum_{i=1}^n [W(P_i) - W(Q_i)]u(x_i), \tag{4}$$

where the consequences are ranked, such that  $x_1 > x_2 > \dots > x_i > \dots > x_n > 0$ ,  $P_i = \sum_{j=1}^i p_j$ , the (decumulative) probability that a consequence is greater than or equal to  $x_i$ , and  $Q_i = \sum_{j=1}^{i-1} p_j$  is the probability that a consequence is strictly greater than  $x_i$ .

Eq. (4) satisfies stochastic dominance, avoiding the need for the editing principle of dominance detection (Tversky & Kahneman, 1992). It also automatically satisfies coalescing, eliminating the need for the editing rule of combination (Birnbaum, 1999a; Birnbaum & Navarrete, 1998; Luce, 1998). CPT generalizes OP to gambles with more than two nonzero consequences. The

Table 1  
Comparison of decision theories

	Branch Independence	
	Satisfied	Violated
Coalescing	Satisfied	Violated
Satisfied	EU (OP*/CPT*)	RDU/RSDU/CPT*
Violated	SWU/OP*	RAM/TAX

Notes: Expected Utility (EU) theory satisfies both properties. OP = Original Prospect theory and CPT = Cumulative Prospect theory; these theories make different predictions with and without their editing rules. The editing rule of combination produces satisfaction of coalescing and the editing rule of cancellation implies branch independence. CPT has the same representation as Rank Dependent Expected Utility (RDU). With or without the editing rule of combination, CPT satisfies coalescing. The Rank Affected Multiplicative (RAM) and Transfer of Attention Exchange (TAX) models are configural weight models that violate both branch independence and coalescing.

SP/A theory of Lopes and Oden (1999) also satisfies coalescing and stochastic dominance.

The rank-dependent theories attribute the Allais paradox to violations of restricted branch independence. Eq. (4) links the pattern of violation of branch independence to the Allais paradox, because both phenomena are (in theory) produced by the same weighting function (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997).

Wu and Gonzalez (1998) presented an illuminating analysis of three distinct types of common consequence paradoxes, among which the original versions of Allais represent only one type. These correspond to changing the consequence on the common branch from the lowest to middle, from middle to highest, and from lowest to highest consequence in the choice. They showed that if the weighting function has an inverse-S shape, the observed paradoxical choices in these three types of common consequence paradoxes can be fit by Eq. (4). Birnbaum (2001b) replicated all three types of common consequence paradoxes with chances at real but modest monetary prizes.

The studies of Wu and Gonzalez (1998) and of Birnbaum (2001b) investigated violations of Allais independence, which confounds branch independence and coalescing. All three types of constant consequence effects can be predicted equally well by the CPT model (which attributes the paradoxes to violations of branch independence) and by Birnbaum's configural weight models (which attribute constant consequence paradoxes mainly to violations of coalescing). The present paper will dissect these two properties in order to distinguish the models in Table 1; therefore, the design allows a comparison of these four classes of rival theories.

In this paper, CPT will be tested both with and without its editing principle of cancellation, which implies branch independence. Allowing CPT both its equation and its contradictory editing principles is a *very lenient* standard, since it allows CPT to handle two of three possible outcomes of a test of branch independence, including mixtures of those two. The standard is as follows: Either the Allais paradox and violations of branch independence will be linked by the same weighting function in Eq. (4), or branch independence will hold in "transparent" tests, or the data will be intermediate between these two patterns.

Note that OP and CPT with or without these editing rules lay claim to three of the four cells in Table 1. There is only one possible outcome of the experiment that would refute both prospect theories with or without their editing rules.

This experiment is designed to test predictions of prospect theories against the configural weight models of Birnbaum (1997, 1999a). The configural models were fit to data of Birnbaum and McIntosh (1996) for

violations of restricted branch independence in gambles with three equally likely branches, and to Tversky and Kahneman's (1992) data for certainty equivalents of binary gambles with nonnegative consequences. Calculations from those parameter estimates are termed here the "prior" predictions, and should not be confused with "predictions" based on a post hoc fit of a model to the same data being "predicted."

### 1.5. Configural weight, RAM and TAX models

Birnbaum (1974); Birnbaum and Stegner (1979) proposed configural, branch weighted averaging models in which the weight of a branch "depends in part on its rank within the set." Birnbaum employed this configural weighting to explain interactions in judgment data (Birnbaum, 1973, 1974), risk aversion and risk seeking in buying and selling prices (Birnbaum & Sutton, 1992), and violations of branch independence (Birnbaum & McIntosh, 1996). Although these models have some similarities to rank-dependent utility models (Weber, 1994), it is important to keep in mind that the definition of *rank* in Birnbaum's models applies to consequences on discrete *branches*, and not to cumulative probability. In the class of models that have subsequently come to be known as "rank-dependent" models (including RDU, RSDU, CPT, and SP/A), rank refers to cumulative probability.

In a risky gamble, the term "branch" refers to each probability-consequence pair that is distinct in the gamble's presentation. In this notation, gambles that represent the same prospect may be subjectively distinct. Event-splitting produces extra branches, whereas coalescing reduces the number of branches in a gamble. For example,  $G = (\$98, 0.8; \$2, 0.2)$  is a two branch gamble that is distinct from the three branch gamble,  $H = (\$98, 0.4; \$98, 0.4; \$2, 0.2)$ , even though they are the same objectively, and represent the same prospect.

These two classes of representations (configural versus rank-dependent) cannot be distinguished when applied separately to certain types of experiments, such as experiments of Birnbaum and McIntosh (1996), Tversky and Kahneman (1992), or Wu and Gonzalez (1998). However, these two classes of models can be distinguished by other tests (Birnbaum, 1997), including new tests used in this paper (Table 1).

The Rank-Affected Multiplicative Weights Model (RAM) and Transfer of Attention Exchange (TAX) models (Birnbaum, 1999a, b; Birnbaum & Navarrete, 1998) are two configural weight models that make identical predictions for modal choices in the present study, but which can be distinguished by other tests (Birnbaum, 1997; Birnbaum & Chavez, 1997). Both RAM and TAX models are special cases of branch-weighted configural expected utility models in which the weight of each distinct branch of a gamble gets a weight

that is affected by its probability, the rank of its consequence, and the weights of other branches, as follows:

$$CWU(G) = \frac{\sum_{i=1}^n w(p_i, G)u(x_i)}{\sum_{i=1}^n w(p_i, G)} \quad (5)$$

where  $CWU(G)$  is the configurally weighted utility of gamble  $G$ , and  $w(p_i, G)$  is the configural weight of the branch with consequence  $x_i$  in Gamble  $G = (x_1, p_1; x_2, p_2; \dots; x_i, p_i; \dots; x_n, p_n)$ , where the consequences are ranked such that  $x_1 > x_2 > \dots > x_n$ . This expression is quite general and includes CPT, RAM and TAX as special cases.

**RAM model.** In the RAM model, each configural weight is a product of a function of branch probability and a function of the rank and augmented sign of the branch's consequence (Birnbaum, 1997). For gambles with strictly positive consequences, the weight of each branch is assumed to be a product as follows:  $w(p_i, G) = a(i, n)s(p_i)$ , where  $s(p_i)$  is a function of branch probability and  $a(i, n)$  is the (positive) weight of the branch having the  $i$ th ranked consequence in a gamble with  $n$  discrete branches. Substituting this assumption for the weights in Eq. (5) yields the RAM model:

$$RAM(G) = \frac{\sum_{i=1}^n a(i, n)s(p_i)u(x_i)}{\sum_{i=1}^n a(i, n)s(p_i)} \quad (6)$$

The rank weights describe how much weight is applied to each discrete branch depending on the rank of the consequence of that branch in the gamble. In practice, for  $n = 2, 3$ , and 4 branches, the estimated rank coefficients in the RAM model are approximately equal to their ranks; i.e.,  $a(i, n) = i$ , with 1 = highest, 2 = second highest, 3 = third highest. In practice,  $s(p)$  and  $u(x)$  are approximated by power functions,  $s(p_i) = p_i^\gamma$  with  $0 < \gamma < 1$ , and  $u(x) = x^\beta$  where  $0 < \beta \leq 1$ . It has been found that for positive cash prizes in the domain of pocket money ( $\$1 < x < \$150$ ), the approximations,  $u(x) = x$ , and  $s(p) = p^{0.6}$  gives a good fit to choices made by undergraduates. These will be termed the "prior" parameters of RAM, as they were selected to approximate previous data of Birnbaum and McIntosh (1996) and Tversky and Kahneman (1992). The use of a linear utility function reveals that risk aversion in this RAM model is ascribed entirely to the rank weights, with lower ranked branches receiving more weight.

The Certainty Equivalent (CE) of Gamble  $G$  is the value of cash for which a person would be equally happy to accept the cash or Gamble  $G$ ; i.e.,  $CE(G) \sim G$ , where  $\sim$  represents indifference. Predicted certainty equivalents of gambles of the form,  $G = (x, p; y, 1 - p)$ , based on the prior parameters of RAM are shown in Fig. 1. These are an inverse-S function of probability:

$$CE(G) = \frac{a(1,2) \cdot t(p)}{a(1,2)t(p) + a(2,2)t(1-p)}$$

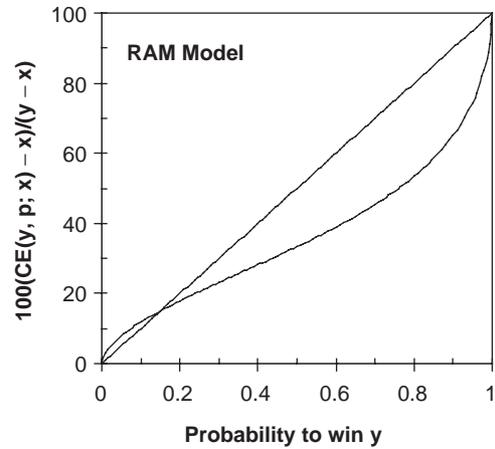


Fig. 1. Predictions of the RAM model with prior parameters for certainty equivalents (CE) of two-branch gambles of the form,  $F = (x, p; y)$ , plotted as a function of  $p$ . Each CE has been linearly transformed to a 0–100 scale. The height of the curve at  $p = 1/2$  can be used to solve for branch rank weights, and the value of  $\gamma$  can be estimated from the curvature.

The height of the curve at  $p = 1/2$  can be used to estimate the ratio of rank weights; for example, in Fig. 1,  $a(2, 2)/a(1, 2) = 2 : 1$ . If  $\gamma < 1$ , the curve will have an inverse-S shape in which people are risk-seeking for small  $p$ ; and with  $\gamma > 1$ , it will have an S-shape in which people can be risk averse for small  $p$ . The prior RAM model, illustrated in Fig. 1, agrees with data of Tversky and Kahneman (1992) and with the Tversky and Wakker (1995) model of CPT for two branch gambles.

Like CPT, the RAM model violates restricted branch independence and satisfies comonotonic restricted branch independence. Unlike CPT, which violates distribution independence, RAM satisfies distribution independence (Birnbaum & Chavez, 1997). If all of the rank weights were equal to each other (e.g.,  $a(i, n) = 1 \forall i, n$ ), the RAM model would imply no violations of restricted branch independence, but would still violate coalescing and properties derived from coalescing, such as stochastic dominance. When the branch rank weights are all equal and  $s(p) = p$ , then RAM reduces to EU.

It is important to keep in mind that although CPT and RAM can both account for violations of restricted branch independence, they make opposite predictions for the types of violation, given their prior parameters (which as we will see below, are needed to account for Allais paradoxes).

**TAX model.** The TAX model is also a special case of Expression 5. In the TAX model, a portion of the probability weight of each branch is transferred among branches according to the ranks of the consequences on the branches. To explain risk aversion in the TAX model, it is assumed that weight is taken from branches with higher valued consequence and given to branches with lower valued consequences.

Birnbaum and Chavez (1997) represented branch weights in the TAX model for a risk-averse subject by a model that can be rewritten as follows:

$$\mathbf{w}_n = \mathbf{t}_n \cdot \Omega_n, \tag{7a}$$

where  $\mathbf{w}_n$  is a  $1 \times n$  vector containing the relative weights of branches 1– $n$ , respectively,  $\mathbf{t}_n$  is a  $1 \times n$  vector containing transformed probabilities normalized to sum to 1 in each gamble. In practice, the normalized probability weighting is approximated as follows:  $t_i =$

$\frac{p_i^\gamma}{\sum_{j=1}^n p_j^\gamma}$ , where  $\gamma$  is the parameter of the probability transformation. Each entry of the  $n \times n$  matrix,  $\Omega_n$ ,  $\omega_{ijn}$  represents the proportion of weight transferred from branch  $i$  to branch  $j$  in a gamble with  $n$  branches. The sum of each row of this matrix is 1. Birnbaum and Chavez (1997) further simplified this model by the assumption that in choice, weight transfers are given by  $\omega_{ijn} = \omega_n = -\delta/(n + 1)$  for  $j > i$ , and  $\omega_n = 0$  for  $j < i$ . For three branch gambles, this model can be written as follows:

$$[w_1 \ w_2 \ w_3] = [t(p_1) \ t(p_2) \ t(p_3)] \times \begin{bmatrix} 1 - 2\omega & \omega & \omega \\ 0 & 1 - \omega & \omega \\ 0 & 0 & 1 \end{bmatrix}, \tag{7b}$$

where  $\omega = -\delta/4$ , for three branch gambles.

If  $\omega = 0$ , then all weight transfers are zero and there would be no violations of restricted branch independence. In this case, each branch’s weight would be a simple function of branch probability. If  $\delta = 0$  and  $t(p) = p$ , this TAX model reduces to EU. When  $\delta \neq 0$ , the proportion of weight transferred is assumed to be a fixed proportion  $[\delta/(n + 1)]$  of the branch giving up the weight. Therefore, the sum of the weights is constant, so that weight is neither created nor destroyed, but only transferred from one branch to another.

When  $\delta < 0$ , weight is transferred from branches with higher ranked consequences to branches with lower ranked consequences. Intuitively, the transfer of weights in the TAX model represents a transfer of attention from branches with higher valued consequences to branches with lower valued consequences. The TAX model, like RAM, can imply risk aversion without postulating a nonlinear utility function. With the assumption that utility is linearly related to cash value, risk averse behavior for binary gambles with  $p = 1/2$  holds iff  $\delta < 0$  in this model.

Birnbaum and Chavez noted that if  $\delta = -1$ , this model would reproduce violations of branch independence reported by Birnbaum and McIntosh (1996). Birnbaum and Navarrete (1998), who tested three new properties not examined in the previous work, estimated the median value of  $\delta$  for 100 undergraduates to be

–1.09, close to this prior value of –1. With  $\delta = -1$ , the transfers are as follows: In two branch gambles, one-third of the probability weight of the branch with the higher consequence is transferred to the lower valued branch. In three branch gambles, one-fourth of the weight of each higher branch is transferred to each lower valued branch. Birnbaum and Navarrete (1998) reported that the median estimate of  $\gamma$  was 0.74, close to the prior value of 0.7 that was chosen to mimic data of Tversky and Kahneman (1992). Birnbaum (1999b) reported median estimates for a more highly educated sample of people tested via the Internet to be  $\delta = -0.33$  and  $\gamma = 0.79$ .

Predictions for the prior TAX model in this paper are computed with the following parameters:  $u(x) = x$ ,  $t(p_i) = \frac{p_i^{0.7}}{\sum_{j=1}^n p_j^{0.7}}$ , and  $\delta = -1$ . These parameters were chosen to approximate data of Tversky and Kahneman (1992) and Birnbaum and McIntosh (1996), but they also work well for predicting other phenomena (Birnbaum, 1999a). Predictions for two branch gambles for this TAX model are shown in Fig. 2, plotted as in Fig. 1. The value of the configural parameter,  $\delta$ , can be estimated from the relative height of the curve at  $p = 1/2$ , and the curve will be inverse-S or S-shaped when  $\gamma$  is less than or greater than 1, respectively.

When there is a fixed number of positive-valued branches with a fixed probability distribution, RDU/RSDU/CPT, RAM and TAX are all equivalent to each other and reduce to what Birnbaum and McIntosh (1996) called the “generic rank-dependent configural weight model,” which Luce (2000) calls the rank weighted additive model.

In both RAM and TAX models, the approximation  $u(x) = x$ , gives a good fit to data in studies involving risky decisions with monetary consequences in the domain of pocket money, i.e., when  $\$1 < x < \$150$ . With this assumption, RAM and TAX have fewer parameters

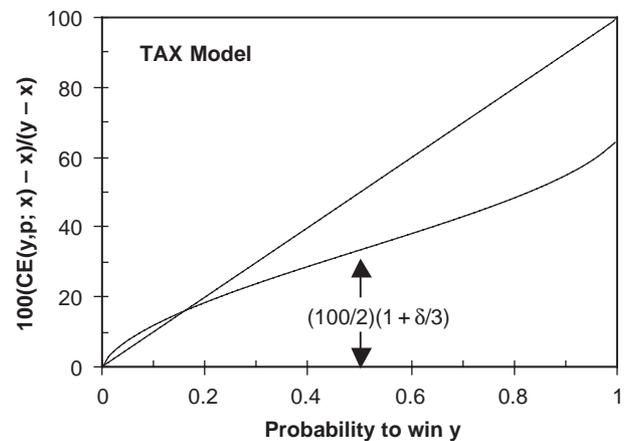


Fig. 2. Predictions of the prior TAX model for binary gambles, plotted as in Fig. 1. Height of the curve at  $p = 1/2$  can be used to estimate  $\delta$ ; the curvature can be used to estimate  $\gamma$ .

than the comparable CPT model for strictly positive consequences. But in this paper, RAM and TAX will use prior parameters to predict the new results, requiring nothing to be estimated from the present data. Predictions of CPT with parameters of Tversky and Kahneman (1992) will be calculated for comparison. Thus, both models are tested on equal footing.<sup>3</sup>

RAM and TAX models violate both restricted branch independence and coalescing, except when they reduce to EU. Because they violate coalescing, these models violate properties that can be derived from coalescing including stochastic dominance and Allais independence.

Unlike stripped prospect theory, RAM and TAX models do not, however, violate transparent dominance (Birnbaum, 1999a): improving the consequence of a given branch (holding everything else constant) improves a gamble. Similarly, moving probability from a branch with a lower valued consequence to a branch with a higher valued consequence (holding everything else constant) improves the gamble.

Although they satisfy transparent dominance, both RAM and TAX violate first order stochastic dominance for specially constructed choices (Birnbaum, 1997).

1.6. Violations of stochastic dominance

Because RAM and TAX models violate coalescing, they violate stochastic dominance. Birnbaum’s (1997) recipe for creating violations of stochastic dominance in configural weight models is based on splitting the lower or higher-valued branch of a root gamble. For example, let the root gamble be  $G = (\$98, 0.9; \$12, 0.1)$ , and construct the following:

$G-$ : 0.85 to win \$98     $G+$ : 0.90 to win \$98  
           0.05 to win \$90        0.05 to win \$14  
           0.10 to win \$12        0.05 to win \$12

According to the configural weight RAM and TAX models, splitting the higher branch of  $G$  gives greater total weight to higher consequence(s). Thus, even though the consequence on the 0.05 splinter has been reduced from \$98 to \$90 ( $G-$  is dominated by  $G$ ),  $TAX(G-) > TAX(G)$ . Splitting the lower branch of  $G$  creates  $G+$ , which dominates  $G$ , but in theory, now the lower consequences get greater weight, making  $G+$  worse than  $G$ , even though the 0.05 splinter has been increased in value from \$12 to \$14. So, according to RAM and TAX models, people should choose  $G-$  over  $G+$ , even though  $G+$  dominates  $G-$ .

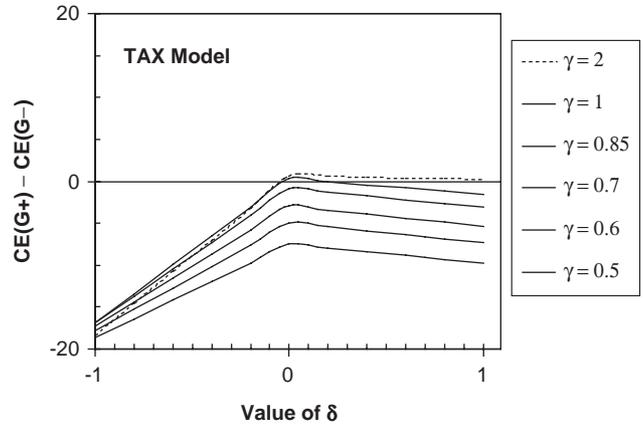


Fig. 3. Analysis of violation of stochastic dominance in the TAX model. The values calculated are  $CE(G+) - CE(G-)$  for Choice 5 in Table 4. Negative values show violations of stochastic dominance, which occur for  $\delta < 0$  (risk averse for  $p = 1/2$ ) and  $\gamma < 1$  (inverse-S in Fig. 2). Violations of stochastic dominance can also occur for risk-neutral ( $\delta = 0$ ) and even risk-seeking ( $\delta > 0$ ) people if  $\gamma$  is small enough.

An analysis of predicted violations of stochastic dominance in the TAX model is shown in Fig. 3. The ordinate plots the value of  $CE(G+) - CE(G-)$ , as a function of the value of  $\delta$ , the configural weighting parameter of TAX. Separate curves are used for different values of  $\gamma$ . Negative values on the ordinate represent violations of stochastic dominance.

Fig. 3 shows that the TAX model always violates stochastic dominance in this case if  $\delta < 0$  and  $\gamma < 1$ . Given the data of Tversky and Kahneman (1992), one concludes that  $\delta < 0$  because people are risk-averse for two-branch, 50–50 gambles to win positive consequences. In addition, one concludes that  $\gamma < 1$  because people are simultaneously risk-seeking for 2-branch gambles with very small probabilities to win (for a fit of this model to their data, see Birnbaum, 1997, Fig. 9). Therefore, given the data of Tversky and Kahneman (1992), the TAX model of Eq. (7b) is forced to predict violations of stochastic dominance in this recipe. Birnbaum (1997, p. 94) put this prediction in print, specifying both  $G-$  and  $G+$  and stating (p. 94), “It seems worthwhile to test such predictions...”

Birnbaum and Navarrete (1998) tested this interesting prediction and found that about 70% of 100 undergraduates tested violated stochastic dominance in this choice and three others like it. Birnbaum, Patton, and Lott (1999) found similar results with a new group of 110 subjects and five new choices constructed from the same recipe.

In a subsequent study, Birnbaum (1999b) found that 72% of a new group of 124 undergraduates violated dominance on the above choice, but only 15% of the same people violated stochastic dominance when the

<sup>3</sup>To explore predictions of TAX, RAM, CPT, and EV, visit the following URL for links to free on-line calculators: <http://psych.fullerton.edu/mbirnbaum/calculators/>

same (objective) choices are presented in split form, as follows:

$GS-$ : 0.85 to win \$98	$GS+$ : 0.85 to win \$98
0.05 to win \$90	0.05 to win \$98
0.05 to win \$12	0.05 to win \$14
0.05 to win \$12	0.05 to win \$12

Because the choice between  $G-$  and  $G+$  is the same as that between  $GS-$  and  $GS+$ , except for coalescing, people should make the same choices, if coalescing holds. Approximately 62% of undergraduates tested (significantly more than half!), however, switched from  $G-$  to  $GS+$ , whereas fewer than 5% switched in the opposite direction; these systematic preference reversals indicate that coalescing is not descriptive of human choice (Birnbaum, 1999b, 2000; Birnbaum & Martin, 2003). The prior RAM and TAX models predicted both results.

One might express reservations about these previous tests of event-splitting, however, based on the following argument. The choice between  $GS-$  and  $GS+$  might invoke the editing mechanism of dominance detection, so the apparent violation of coalescing might be produced by comparison processes such as editing, rather than by the evaluation function. The present study provides new tests of coalescing that do not involve dominance.

### 1.7. Event framing versus coalescing

Tversky and Kahneman (1986) presented a case in which more violations of stochastic dominance were observed in a framed and coalesced choice than in a differently framed and split form of the same choice. They noted that their theory assumed coalescing (see also Kahneman, 2003, p. 727), and they emphasized instead the importance of the event framing used to “mask” the dominance relationship. In their framing, the dominated gamble was made to appear as if for any named event (for any color of marble drawn from an urn) the dominated gamble gave either an equal or better consequence. Because there were (slightly) different numbers of marbles of each color in the two urns, however, the so-called “events” were not really the same. Because the numbers of marbles were nearly equal, Tversky and Kahneman theorized that judges would simplify the choice by canceling the nearly equal branches produced by the “same” named events. They conceded, however, that in their test, several different interpretations were confounded, including a comparison of the number of branches with positive or negative consequences in each gamble.

A second purpose of the present study is therefore to assess the importance of event framing (as opposed to event-splitting/coalescing) in tests of branch independence and Allais independence as well as in tests of

stochastic dominance. Event framing would be expected to *reduce* violations of branch independence in the split forms. Such choices might be termed “transparent” tests of branch independence in the framed form, because both gambles would clearly share a common event–consequence branch. In such a framed format, a decision-maker should find it easy to cancel branches that are identical in two choices and to make a choice based strictly on what is left.

The event framing manipulation is illustrated in Choice 16 of Table 2, which is “framed” as opposed to Choice 14 in Table 3, which is “unframed”. If people attend to framing and cancel common branches, they would presumably show greater *conformance* to branch independence in framed than unframed tests. Choices 9 and 16 (Table 2) should more likely yield the same decisions in the FU condition, where the common branch has the same color than they would in condition UF, where the colors of marbles on corresponding branches are different.

Similarly, people should be *more* likely to *violate* stochastic dominance on Choice 5 in Table 4 in the FU condition, with common color framing, than in the UF condition where it is unframed. The reasoning here uses the editing principles (Kahneman & Tversky, 1979) of simplification and cancellation: the common color branches to win \$96 and \$12 in Choice 5 are nearly equal (in probability), which if cancelled from both sides, leaves a branch that favors the (dominated) gamble.

## 2. Method

Deciders made 20 choices between pairs of gambles. They viewed the materials on-line via the Internet, clicking the button beside the gamble they would rather play in each choice. They were informed that 3 lucky participants would be selected at random to play one of their chosen gambles for money, with prizes as high as \$110, so they should choose carefully. Prizes were awarded as promised. Each choice appeared as in the following example:

### 1. Which do you choose?

- A: 50 red marbles to win \$100
- 50 white marbles to win \$0
- OR
- B: 50 blue marbles to win \$35
- 50 green marbles to win \$25

Instructions read (in part) as follows:

“Think of probability as the number of marbles in one color in an urn (container) containing 100 otherwise identical marbles, divided by 100. Gamble A has 50 red marbles and 50 white marbles; if a marble drawn at random from urn A is red, you win \$100. If a white

Table 2  
Dissection of Allais paradox (Series A) (each entry is the percentage in each condition choosing the “safe” gamble, *S*)

No	Relation to previous row	Choice as in condition FU		Condition			Prior TAX model		Prior CPT model	
		First gamble, <i>R</i>	Second gamble, <i>S</i>	FU	UF	Rep	<i>R</i>	<i>S</i>	<i>R</i>	<i>S</i>
6		10 black marbles to win \$98 90 purple marbles to win \$2	20 black marbles to win \$40 80 purple marbles to win \$2	41	37	36	13.3>	9.0	16.9>	10.7
9	Split # 6	10 red marbles to win \$98 10 blue marbles to win \$2 80 white marbles to win \$2	10 red marbles to win \$40 10 blue marbles to win \$40 80 white marbles to win \$2	69	66	60	9.6	<11.1	16.9>	10.7
12	RBI #9	10 red marbles to win \$98 80 blue marbles to win \$40 10 white marbles to win \$2	10 red marbles to win \$40 80 blue marbles to win \$40 10 white marbles to win \$40	62	55	47	30.6	<40.0	38.0	<40.0
16	RBI #9, 12	80 red marbles to win \$98 10 blue marbles to win \$98 10 white marbles to win \$2	80 red marbles to win \$98 10 blue marbles to win \$40 10 white marbles to win \$40	47	45	39	62.6>	59.8	67.6	<74.5
19	Coalesce #16	90 red marbles to win \$98 10 white marbles to win \$2	80 red marbles to win \$98 20 white marbles to win \$40	81	82	72	54.7	<68.0	67.6	<74.5

Notes: The common branch was 80 marbles to win \$2 in Choices 6 and 9, \$40, in Choice 12, and \$98 in Choices 16 and 19, respectively. RBI = Restricted Branch Independence, *R* = Risky Gamble, *S* = Safe Gamble, In condition FU, these choices were all “framed” by having the same color of marbles on corresponding branches; in condition UF, these were unframed; Rep = replication study with *n* = 150. TAX model shows calculated certainty equivalent values of each gamble with prior parameters. CPT model shows calculated certainty equivalents under model of Tversky and Kahneman (1992).

Table 3  
Dissection of Allais paradox into branch independence and coalescing (Series B)

No	Relation to previous row	Choice as in condition FU		Condition			Prior TAX model		Prior CPT model	
		First gamble, <i>S</i>	Second gamble, <i>R</i>	FU	UF	Rep	<i>S</i>	<i>R</i>	<i>S</i>	<i>R</i>
10		15 red marbles to win \$50 85 black marbles to win \$7	10 blue marbles to win \$100 90 white marbles to win \$7	86	74	82	13.6	<18.0	15.9	<22.1
17	Split #10	10 red marbles to win \$50 05 blue marbles to win \$50 85 white marbles to win \$7	10 black marbles to win \$100 05 purple marbles to win \$7 85 green marbles to win \$7	53	45	46	15.6>	14.6	15.9	<22.1
20	RBI #17	10 red marbles to win \$50  85 white marbles to win \$50 05 blue marbles to win \$50	10 black marbles to win \$100  85 purple marbles to win \$50 05 green marbles to win \$7	44	49	52	50.0>	40.1	50>	49.2
14	RBI # 17, 20	85 red marbles to win \$100 10 white marbles to win \$50 05 blue marbles to win \$50	85 black marbles to win \$100 10 yellow marbles to win \$100 05 purple marbles to win \$7	63	63	60	68.4	<69.7	82.2>	79.0
8	Coalesce #14	85 black marbles to win \$100 15 yellow marbles to win \$50	95 red marbles to win \$100 05 white marbles to win \$7	25	16	34	75.7>	62.0	82.2>	79.0

Notes: Gambles unframed, as in condition FU; in condition UF, each choice was framed. Each entry is the percentage of people in each condition who chose the second, “risky” gamble, *R*. The common branch is 85 marbles to win either \$7 (in Choices 10 and 17), \$50 (Choice 20), or \$100 (Choices 14 and 8), respectively. RBI = restricted branch independence.

marble is drawn, you win \$0. So, the probability to draw a red marble and win \$100 is 0.50 and the probability to draw a white marble and get \$0 is 0.50. If someone reaches in urn A, half the time they draw red and win \$100 and half the time they draw white and win \$0. But in this study, you only get to play a gamble once, so the

prize will be either \$0 or \$100. Gamble B’s urn has 100 marbles also, but 50 of them are blue, winning \$35, and 50 of them are green and win \$25. Urn B thus guarantees at least \$25, but the most you can win is \$35. Some will prefer A and others will prefer B. To mark your choice, click the button next to A or B...”

Table 4  
Violations of stochastic dominance and coalescing linked to event framing and event-splitting

No.	Choice as in condition FU		Condition			Prior TAX		Prior CPT	
	G+	G–	FU	UF	Rep	G+	G–	G+	G–
5	90 red marbles to win \$96 05 blue marbles to win \$14 05 white marbles to win \$12	85 red marbles to win \$96 05 blue marbles to win \$90 10 white marbles to win \$12	<b>73</b>	85	76	45.8	<63.1	70.3>	69.7
11	85 red marbles to win \$96 05 blue marbles to win \$96 05 green marbles to win \$14 05 white marbles to win \$12	85 red marbles to win \$96 05 blue marbles to win \$90 05 green marbles to win \$12 05 white marbles to win \$12	<b>15</b>	11	11	53.1>	51.4	70.3>	69.7
15	90 red marbles to win \$96 05 yellow marbles to win \$14 05 pink marbles to win \$12	85 black marbles to win \$96 05 blue marbles to win \$90 10 white marbles to win \$12	77	<b>74</b>	78	45.8	<63.1	70.3>	69.7
7 <sup>a</sup>	94 black marbles to win \$99 03 yellow marbles to win \$8 03 purple marbles to win \$6	91 red marbles to win \$99 03 blue marbles to win \$96 06 white marbles to win \$6	78	<b>74</b>	70	46.0	<66.6	76.2>	75.9
13 <sup>a</sup>	91 black marbles to win \$99 03 pink marbles to win \$99 03 yellow marbles to win \$8 03 purple marbles to win \$6	91 red marbles to win \$99 03 blue marbles to win \$96 03 green marbles to win \$6 03 white marbles to win \$6	10	<b>16</b>	13	54.2>	53.2	76.2>	75.9
18 <sup>a</sup>	94 red marbles to win \$99 03 blue marbles to win \$8 03 white marbles to win \$6	91 red marbles to win \$99 03 blue marbles to win \$96 06 white marbles to win \$6	<b>75</b>	72	70	46.0	<66.6	76.2>	75.9

Notes: Choices 5, 11, and 18 were framed. In Choices 5, 11, and 15, the dominant gamble was presented first and in Choices 7, 13, and 18, G+ was presented second. Each entry is the percentage of people in each condition who violated stochastic dominance (bold type shows violations in framed choices).

2.1. Allais paradoxes: coalescing and branch independence

Choices for Series A and B of Allais paradoxes are shown in Tables 2 and 3, respectively. Each choice is created from the choice in the row above by either coalescing/splitting or by restricted branch independence. Within each series, choices should be the same in every row, according to EU, except for random error. In Series A, the common branch is 80 marbles to win \$2 (first two rows), \$40 (middle row), or \$98 (last two rows). In Series B, the common branch is 85 marbles to win \$7 (first two rows), \$50 (third row), or \$100 (fourth and fifth rows). Note that the positions (First or Second) of the S or “safe” gamble with higher probability to win a smaller prize and the R, or “risky” gamble are counterbalanced between Series A and B.

2.2. Framing manipulation

Each choice was either framed or unframed. In the framed version, the same marble colors are used for each ordered branch. A framed and coalesced test of stochastic dominance is shown in Choice 5 of Table 4,

and the unframed version of the same (objective) choice is shown in Choice 15. The framed and split form of this choice is shown as Choice 11 of Table 4.

There were two conditions to which participants were randomly assigned by means of a JavaScript routine (Birnbaum, 2001a, p. 211). In the FU condition (shown in Tables 2, 3, and 4), all choices in Series A (Table 2) were framed, all in Series B (Table 3) were unframed; Choices 5, 11, and 18 of Table 4 were framed and Choices 7, 13, and 15 were not. In the UF condition, framing was reversed from that of FU.

The first four choices, which served as a “warm-up,” were the same as those of Birnbaum (1999b), formatted in terms of the marbles. Complete materials can be viewed at URLs: [http://psych.fullerton.edu/mbirnbaum/Exp2\\_urnsUF\\_A.htm](http://psych.fullerton.edu/mbirnbaum/Exp2_urnsUF_A.htm) [http://psych.fullerton.edu/mbirnbaum/Exp2\\_urnsFU\\_A.htm](http://psych.fullerton.edu/mbirnbaum/Exp2_urnsFU_A.htm)

2.3. Participants

Participants were 200 people recruited by links on the Web and from the usual “subject pool” in the psychology department of California State University,

Fullerton. When each condition had 100 participants, the study was deemed complete.

#### 2.4. Replication study

An additional 150 participants, recruited entirely from the Web, were randomly assigned to Conditions FU or UF, and were tested in a simple replication of the entire study.

### 3. Results

#### 3.1. Allais paradoxes

Tables 2 and 3 show the percentage of participants in each condition who chose the second gamble in each choice of Series A and B. Separate columns show choice percentages for each framing condition of the main study and for the combined results of the replication study. According to EU, choices should be the same in every row within Table 2 and within Table 3, except for error; therefore, the choice percentages should not change systematically from row to row. The original type of the Allais paradox involves comparison of Choices 6 and 12 in Table 2, but any systematic change in preference from row to row in Tables 2 and 3 would be a violation of Allais independence. The data show systematic reversals, demonstrated by the finding that choice percentages change significantly from row to row in both tables.

For example, Choice 6 in Table 2 (averaged over framing) shows that only 39% chose the “safe” gamble (with 20 marbles to win \$40) over the “risky” gamble (with 10 marbles to win \$98 and 10 to win \$2) when the common branch was 80 marbles to win \$2. However, in Choice 19 (where the common branch was 80 marbles to win \$98), 82% chose the “safe” gamble. Similarly, Table 3 shows that the percentage choice changed from 20% (for the “safe” gamble in Choice 10) to 80% (for the “safe” gamble in Choice 8), as the common consequence on the branch with 85 marbles was increased from \$7 to \$100. These are large violations of Allais independence. The replication study (column labeled “Rep”) yielded very similar results, which are averaged over two framing conditions.

These violations of Allais independence are statistically significant, even by the conservative standard that the modal choice had to be significantly reversed. For a sample of  $n = 200$ , the binomial distribution with  $p = 1/2$  has a mean of 50% and a standard deviation of 3.5%; therefore, observed percentages outside the interval from 43% to 57% deviate significantly from 50% by a two-tailed test with  $\alpha = 0.05$ . Combining over conditions, the percentage choosing the “risky” gamble changed from significantly less than 50% to significantly

greater than 50% as the common consequence was increased from the lowest value to the highest value in both Tables 2 and 3. The replication study ( $n = 150$ ) also shows significant reversals by the same test in both tables.

The binomial test of correlated proportions is a more sensitive test of significance of within-subject changes in choice proportions. This test checks for equality of choice proportions, rather than requiring a significant reversal of the mode. Comparing Choices 6–19, for example, this test compares the number who switched from choosing the (“risky”) first gamble in Choice 6 to the (“safe”) second gamble in Choice 19 against the number who switched preferences in the opposite direction. In the FU condition of Series A, 47 switched from “risky” to “safe”, against only 7 who switched in the opposite direction ( $z = 5.44^*$ ); in the UF condition, the numbers were 50 versus 5 ( $z = 6.07^*$ ). The critical value of  $|z|$  for a two-tailed test with  $\alpha = 0.05$  is 1.96. The term “significant” and asterisks (\*) are used throughout this paper to denote significant differences by this test.

For Series B (Table 3), the first gamble was the “safe” one; hence, the same directional change is observed in both tables, counterbalanced for position. The numbers who switched from the “risky” second gamble in Choice 10 to the “safe” first gamble in Choice 8 were 64 and 60 in the FU and UF conditions, respectively, against only 4 and 4 who switched in the opposite directions ( $z = 7.28^*$  and  $7.00^*$ ), respectively.

#### 3.2. Tests of coalescing in Allais paradoxes

Tables 2 and 3 separate coalescing/splitting from branch independence. Note that in Table 2, Choice 9 is the same as Choice 6, except for coalescing. The first gamble in Choice 9 is the same as the first gamble in Choice 6, except the lower branch of 90 marbles to win \$2 has been split into 80 marbles to win \$2 and 10 marbles to win \$2. According to the class of RDU/RSDU/CPT models, this manipulation should have no effect; however, the TAX model (with its prior parameters) implies that splitting the lower branch increased the relative weight of winning only \$2 and thus made the first gamble worse.

Similarly, the second gamble of Choice 9 is the same as in Choice 6, except that the 20 marble branch to win \$40 was split, which should have no effect according to the RDU/RSDU/CPT models, but makes the second gamble better according to RAM or TAX. If people obeyed coalescing, they should make the same decisions in Choices #6 and #9. However, according to the prior TAX model, this split makes the second gamble in Choice 9 better because more weight is transferred to the higher consequence.

Choice 19 in Table 2 is the result of coalescing the upper branches (to win \$98) in the first gamble of Choice 16 and coalescing the lower branches (to win \$40) in the second Gamble. Table 3 (Series B) is based on the same plan.

According to the class of RDU/RSDU/CPT models (apart from editing), a person should make the same decisions in Choices 6 and 9, 16 and 19, 10 and 17, and 14 and 8, since each of these comparisons involves only event coalescing/splitting (and transitivity). The values labeled CPT in the tables show calculated certainty equivalent (cash value) of each gamble, based on the model and parameters of Tversky and Kahneman (1992). Although a particular model was used to calculate these predictions, the invariance with respect to coalescing/splitting holds for any functions with any parameters in CPT.

According to the TAX model, however, coalescing/splitting affects the weights of the branches and hence the values of the gambles. The three tables show calculated cash equivalents for each gamble based the TAX model with prior parameters (see Introduction). From Choice 9 to 12, the lower branch of the first gamble has been split. In theory, this split increases the weight of \$2, thereby making the “risky” first gamble (*R*) seem worse (in Choice 9) than it does in the coalesced form of Choice 6 (the calculated certainty equivalent of *R* drops from \$13.3 to \$9.6). In addition, the higher-valued branch of the “safe” second gamble (*S*) has been split, which increases the weight of \$40, making the Second gamble seem relatively better in #9 compared to #6 (its CE increases from \$9 to \$11.1). Therefore, the prior TAX model correctly predicts that there will be a reversal in the modal choice from *R* in Choice 6 to *S* in Choice 9.

Similar predictions in Tables 2 and 3 can be understood from these implications of branch weighting in RAM and TAX: splitting the lower branch makes a gamble worse and splitting the higher branch makes a gamble better.

Testing separately in each framing condition, all eight tests of coalescing in the main design are significant by the test of correlated proportions, with all eight shifts in the direction predicted by the configural weight TAX model with its prior parameters. For example, from Choice 6 to 9, 68\* people switched from picking *R* in Choice 6 to *S* in Choice 9, compared with only 6 who reversed preferences in the opposite direction. From Choices 16 to 19, the higher valued branches of *R* have been coalesced, making *R* seem worse in Choice 19 than it did in Choice 16, whereas the lower consequences of *S* have been coalesced, making *S* seem better in Choice 19. In this case, 77\* switched from *R* to *S* compared with only 8 who switched in the opposite direction. For Choices 10 and 17 in Series B (Table 3), the results are similar: 73\* reversed preferences in the direction

predicted by the TAX model compared to only 12 who switched in the opposite direction. For Choices 14 and 8, 96\* reversed preferences in the predicted direction against only 13 who switched in the opposite direction.

All eight significant changes due to coalescing/splitting are predicted by the configural weight model with its prior parameters, and all eight results are inconsistent with the class of RDU/RSDU/CPT models with any set of parameters, because those models require no systematic effects of coalescing or splitting. The replication study confirms the same conclusions: in all four cases, the shifts are significant by the test of correlated proportions. In three of the four cases (all except Choice 10 versus 17), the shift is even significant by the conservative standard that coalescing significantly reverses the mode.

### 3.3. Tests of branch independence in Allais paradoxes

Choices 9, 12, and 16 in Table 2 differ only in that the common branch of 80 marbles to win \$2 in Choice 9 has been changed to 80 marbles to win \$40 in Choice 12, and to 80 marbles to win \$98 in Choice 16. Therefore, if choices obeyed restricted branch independence, decisions in Choices 9, 12, and 16 would be the same. Similarly, Choices 17, 20, and 14 in Table 3 are the same, except that the consequence on the common branch of 85 marbles was either \$7, \$50, or \$100.

Consider, the following defense of CPT for the violations of coalescing in the previous section. Suppose that in transparent tests of branch independence, people used the editing rule of cancellation at least part of the time. Such a strategy would produce greater satisfaction of restricted branch independence. For example, Choices 9 versus 16 might be called a “transparent” test of branch independence because in each case, the judge could simply cancel the common branch (80 marbles to win \$2 or \$98), which leaves the same remainder in Choices 9 and 16. Similarly, Choices 17 and 14 would be the same if one canceled the common branch of 85 marbles to win \$7 or \$100. Such editing might give an explanation for significant shifts (between the coalesced and split forms in Tables 2 and 3) observed with the tests of correlated proportions.

The problem with this argument, however, is that the violations of branch independence are statistically significant and in the *opposite* direction from that observed in the coalesced versions of the same choices. Summed over framing conditions, 57 switched from the “safe” to the “risky” gamble from Choice 9 to Choice 16 compared to only 17 who switched in the opposite direction ( $z = 4.65^*$ ). Similarly, 55 reversed from “safe” to “risky” from Choices 17 to 14, compared with only 28 who made the opposite switch ( $z = 2.96^*$ ).

In the replication study, significantly more than half chose *S* in Choice 9 and significantly less than half chose

$S$  in Choice 16. Similarly, significantly less than half chose  $S$  in Choice 6 but significantly more than half chose  $S$  in Choice 9. In addition, there was a significant reversal in the majority choice between 16 and 19. These significant reversals, even by the conservative standard, show that CPT, even with its editing principle of cancellation, can be rejected. If we theorize that cancellation is used some of the time by some of the judges, we must conclude that the results would have been even more devastating for CPT.

According to CPT, RDU, and RSDU models, Choice 6 is the same as Choice 9 and Choices 16 and 19 are the same. Similarly, Choice 10 is the same as 17 and 14 is the same as Choice 8 in Table 3. Assuming people sometimes do and sometimes do not use cancellation, the results might have been intermediate between no violations of branch independence and the pattern needed to explain the Allais paradox. For example, had the percentages of choosing the second gamble in Choices 9 and 16 been 47% and 69%, respectively, instead of 69% and 47%, respectively, it might have been argued that the apparent violations of coalescing are caused by partial use of cancellation. However, the results in Tables 2 and 3 are *opposite* those required to save the CPT/RDU/RSDU models even with this editing argument.

Even by the conservative standard of significantly reversing the mode (but lenient to CPT), there is at least one significant switch in each series of the modal choice from significantly less than 50% to significantly greater than 50%. (These significant switches are observed in Choices 6 versus 9 in Series A, and Choices 14 versus 8 in Series B). In the replication study, there are three reversals that significantly reverse by this conservative test (Choices # 6 versus 9, 16 versus 19, and 14 versus 8). It is hard to see how to reconcile such results with CPT, even with the editing principle of cancellation, because the violations are not merely reduced by splitting (as might be expected from the editing principle of cancellation) but significantly reversed.

### 3.4. Tests of stochastic dominance and coalescing

Table 4 summarizes tests of stochastic dominance and coalescing. In Choices 5, 11, and 15, the first gamble dominates the second. The second gamble dominates the first in Choices 7, 13, and 18. In Table 4, all percentages represent violations of stochastic dominance. Violations of stochastic dominance are significantly greater than 50% in all 12 tests of coalesced choices (values exceeding 60% are significant). There are 12 values in the table exceeding 70%, with an average of 75%.

In contrast, violations of stochastic dominance are significantly less than 50% in all 6 tests of the appropriately split versions of the same choices (average of 13%). Each of the 12 tests of event-splitting/

coalescing is also significant by the test of correlated proportions. The replication study again repeats the pattern of significant violations observed in the main study. These results reinforce those of previous tests of stochastic dominance and coalescing in this recipe.

### 3.5. Event framing

The framing effect of marble color had very small effects. For framed and coalesced choices, violations of stochastic dominance had an average of 74% compared to 78% for coalesced and unframed. In the split forms, there were 15.5% violations in the framed cases and 10.5% in the unframed cases. These are small effects, and they go opposite the directions anticipated by the editing notion, which predicted more violations when framed and coalesced and fewer violations in the framed and split conditions.

For tests of branch independence, summing over Series (Choices 9 versus 16 and 17 versus 14), there were 65 choices with the SR switch compared to 26 who showed the opposite in framed choices. With unframed choices, there were 47 showing the SR switch compared to 19 showing the opposite switch. The editing notion held that there should be *fewer* violations of branch independence in the framed cases, so the data again show small effects that are opposite those predicted. In sum, event framing had minimal effects and did not show the patterns expected from the idea that framing would increase the use of cancellation. The replication study also found that event framing effects were minimal (not shown).

### 3.6. Prior predictions

This experiment was designed based on calculations under CPT and TAX models. As noted by Birnbaum and McIntosh (1996), the prediction of violations of branch independence requires a careful “fishnet” design unless parameters are known in advance. Based on the previous parameters, it was possible to design an experiment that should distinguish these theories, if the prior parameters also work for new choices and new participants. Naturally, fitting models to the same data would give a better fit than using previous data to predict new results. However, because post hoc predictions can take advantage of lack of constraint in an experimental design, I think results are more impressive if one can use a model and its parameters to predict from one study to new properties tested in another study.

Tables 2–4 show calculated certainty equivalents from the TAX model with prior parameters (Birnbaum, 1999a). The RAM model with its prior parameters makes the same (directional) predictions as TAX in this study (not shown). TAX and RAM models correctly

predict the majority choice in 31 of 32 percentages of the main study listed in Tables 2–4 (all except Choice 17 in the FU condition, where the 53% should have been less than 50%). It also predicts the majority choice in all but two of 16 choices in the Replication study (in Choice 12, the 47% should have been greater than 50% and in Choice 20, the 52% should have been less than 50%). In none of these three cases where prior TAX was wrong were the percentages significantly different from 50%.

There are eight choice sets where CPT and TAX make different predictions (Choices 9, 16, 17, 14, 5, 15, 7, and 18). Of the 24 empirical choice percentages for these eight cases, the modal choice agreed with TAX in 23 cases and with CPT in only one choice (Choice 17 in the FU condition, 53%). In 19 of 24 cases, empirical choice proportions were significantly different from 50%; in all of these cases, CPT predicted the wrong choice. If the two models were equally good, one would expect half of these 19 significant cases to favor either model.

The prior TAX and RAM models agree with CPT for Choices 6, 12, 19, 10, 20, 8, 11, and 13. That is, these “configural” models make the same predictions as CPT for coalesced tests of Allais independence and for split tests of stochastic dominance. RAM and TAX models differ from any member of the rank-dependent models in that they predict violations of stochastic dominance in Choices 5, 15, 7, and 18 of Table 4. They also differ from any of the rank-dependent models in predicting reversals in Tables 2–4 due to coalescing/splitting.

#### 4. Discussion

Table 1 shows how the separation of restricted branch independence and coalescing allows one to distinguish theories of decision making. The results are inconsistent with both original and cumulative prospect theories because the data show significant violations of both properties. In addition, the type of violation of restricted branch independence is opposite what is needed by CPT to account for the Allais paradoxes.

These results are instead consistent with the conclusion that the primary cause of Allais common consequence paradoxes is violation of coalescing. In every test of coalescing/splitting in Tables 2 and 3, there are large and significant changes that agree in direction with the Allais paradoxes.

Although “pure” tests of restricted branch independence show systematic violations in Table 2 (Choice 9 versus 16) and Table 3 (Choice 17 versus 14), these “pure” violations go in the opposite direction from what is needed by rank-dependent models RDU/RSDU/CPT to explain the Allais paradoxes (first to last row in Tables 2 and 3). Put another way, the inverse-S function with CPT correctly predicts the results of Choices 6, 19, 10 and 8 in Tables 2 and 3. However, that same

weighting function fails to predict the results of Choices 9, 16, 17, and 14, which are identical prospects according to CPT. Both sets of results are consistent with the pattern predicted by RAM and TAX.

All systematic reversals of preference due to splitting/coalescing reported in Tables 2–4 are in the direction one expects from the RAM or TAX models if splitting a branch gives those splinters greater total weight than they would receive when coalesced. Splitting the higher valued branch of a gamble should make the gamble relatively better and splitting the lower valued branch should make it relatively worse. Note that the configural models imply that *each* splitting *and* coalescing operation (from top to bottom rows of Tables 2 and 3) should improve  $S$  and diminish  $R$ .

Therefore, even if there were no violations of branch independence [i.e., even with  $\delta = 0$  in TAX or with  $a(i, n) = 1 \forall i, n$  in RAM], TAX and RAM models would still imply Allais paradoxes from violations of coalescing produced by the curvature of  $t(p)$  (in TAX) or  $s(p)$  (in RAM) alone.

Violations of coalescing also explain violations of stochastic dominance. Splitting or coalescing (in Table 4) changes the percentage of violations from 76% to 13%, significantly reversing the preference between two equivalent prospects.

The RDU/RSDU/CPT models imply coalescing; therefore, these models cannot account for any of the effects in Tables 2–4 produced by coalescing (or splitting).

The class of RDU/RSDU/CPT models easily violate restricted branch independence. However, the particular observed pattern in previous studies as well as this one is opposite that predicted by the inverse-S form (Birnbaum & Chavez, 1997; Birnbaum & McIntosh, 1996; Birnbaum & Navarrete, 1998). A new feature of the present study is that the experimental tests of branch independence are elegantly coupled precisely with Allais common consequence paradoxes to disentangle branch independence and coalescing in the Allais paradoxes. In this design, CPT must violate branch independence in order to explain the Allais paradox, but it does so in the opposite way from what is observed.

##### 4.1. The case against CPT

The cumulative array of data that violate the class of RDU/RSDU/CPT models has now reached a critical threshold where those theories must be questioned as descriptive of human decision making. The weight of evidence against CPT now exceeds the case against EU theory reviewed by Kahneman and Tversky (1979).

The RDU/RSDU/CPT models can be replaced by a more accurate model that uses no more parameters, but which accounts for seven different results that refute this class of theories. CPT (with any choice of functions and

parameters) cannot account for violations of coalescing (event-splitting effects), violations of stochastic dominance, violations of lower cumulative independence, violations of upper cumulative independence, or violations of 3-branch tail independence. In addition, the CPT model, in order to account for Allais paradoxes, violates both branch independence and distribution independence in the opposite direction from that observed. Each of these seven phenomena has now been well established by systematic experiments, and most have been replicated in more than one study or confirmed in the same study with multiple variations. Each of these seven phenomena are consistent with the TAX model, which predicted five of them in advance, including the dissection of the Allais paradox in the present paper.

Birnbaum (1997) derived the properties of lower cumulative independence and upper cumulative independence to clarify the contradiction between empirical violations of branch independence and empirical evidence that implied the inverse-S weighting function in RDU/RSDU/CPT models. These two theorems can be deduced from transitivity, monotonicity, coalescing, and comonotonic restricted branch independence (Birnbaum, 1997). They were originally deduced directly from the RDU representation (Birnbaum et al., 1999, Appendix). Based on RAM and TAX models, Birnbaum (1997) predicted violations of these properties, which were subsequently confirmed in several studies (Birnbaum, 1999b, 2000; Birnbaum & Navarrete, 1998).

Without modification, the class of rank dependent models does not account for violations of 3-branch upper “tail” independence reported by Wu (1994) and replicated by Birnbaum (2001b), a property that can be deduced from transitivity, coalescing, and restricted comonotonic branch independence. These results are, however, consistent with RAM and TAX.

#### 4.2. RAM and TAX models

Both RAM and TAX models predicted the phenomena that violate the class of CPT and RDU models. Indeed, the TAX model was used to design the empirical tests in this paper. Birnbaum’s (1999a) review showed that with the same set of parameters, RAM and TAX models predict results of tests of stochastic dominance, event-splitting effects, lower and upper cumulative independence, and branch independence.

TAX also implies violations of distribution independence (Birnbaum & Chavez, 1997), which violate RAM. The TAX model, with the same prior parameters, explains both classic and modern variations of the Allais paradoxes (Birnbaum, 2000, 2001b; Birnbaum & Martin, 2003; Wu & Gonzalez, 1996, 1998), as well as other data that can be fit with the inverse-S weighting function (Abdellaoui, 2000; Gonzalez & Wu, 1999;

Tversky & Wakker, 1995; Quiggin, 1993; Luce, 2000; Starmer, 2000). The present study completes the picture by showing that the TAX model with the same (prior) parameters correctly predicts the effects of branch independence and coalescing on new variations of the common consequence paradoxes of Allais.

#### 4.3. Can CPT be saved by changing procedure?

It has been found that violations of CPT are obtained when choices are presented in many different forms and formats (Birnbaum submitted). Majority violations of stochastic dominance and coalescing have been observed whether branches are presented in juxtaposed format or with two other arrangements and whether branches are presented in increasing or decreasing order of consequences (Birnbaum & Martin, 2003). They have been observed whether probabilities are presented numerically or accompanied by pie charts that seemingly reveal dominance. They are found both with and without financial incentives. They are observed with highly educated people as well as students (Birnbaum, 1999b, 2000). They are found with students tested in class, in the lab, or via the Web (Birnbaum & Martin, 2003). They are observed whether probabilities are presented as decimal fractions, as natural frequencies, or as lists of equally likely consequences (Birnbaum submitted). They are even found when gambles are presented with decumulative probabilities, which should help people “see” stochastic dominance, because they do not need to use addition to compute the probability of getting a certain prize or higher. The present data show that violations of stochastic dominance are observed whether choices are framed by the same marble colors or not. This growing collection of null findings represents a waste basket of failed attempts to explain violations of CPT by mechanisms other than coalescing, which has substantial effects in all of these different studies despite surface differences in how splitting/coalescing appears in different formats.

Similar results are also obtained whether people make choices between gambles or judge buying and selling prices of  $G+$  and  $G-$  on different trials (Birnbaum & Beeghley, 1997; Birnbaum & Yeary submitted; Birnbaum, Yeary, Luce & Zhou submitted). Violations of stochastic dominance in judgments of value suggest that one should look to theories of the evaluation of gambles, rather than models that focus on contrasts or comparisons between gambles. In other words, if judgments, which (necessarily) satisfy transitivity, also show violations of CPT, one infers that comparison processes or other mechanisms specific to choice seem less plausible than theories of the evaluation of the gambles.

#### 4.4. Predicting choice percentages

This study used the prior TAX model to accurately predict majority choices; however, one might want to predict exact (numerical) choice percentages. Birnbaum and Chavez (1997) used an approximate model for predicting choice percentages. Their model, like the most restricted case of Thurstone's law of comparative judgment or Luce's choice model, assumed that choice percentages are a function of utility differences between gambles. However, such models are oversimplified, since they do not distinguish utility difference from ease of comparison.

For example, in Choice 5 of Table 4, the TAX model implies a utility difference of  $63.1 - 45.8 = 17.3$ , and the empirical choice percentages range from 73% to 85%. However, in Choice 11, the utility difference is much smaller,  $51.4 - 53.1 = -1.7$ , but the choice percentages are more extreme, ranging from 11% to 15%. Here the smaller absolute utility difference produced the more extreme choice percentage. Similar results can be found in other cases in the tables. To account for such data, one can use a choice model in which the predicted choice percentage is a function of the difference in utility divided by a parameter representing the difficulty of discrimination (see Diederich & Busemeyer, 1999). Despite the small difference, it is "easy to see" that the first gamble is better in Choice 11. The parameter representing the difficulty of discrimination can be thought of as the standard deviation of the difference. Jerome Busemeyer (pers. comm., 2003) is currently working on ideas for calculating this standard deviation for the case of independent gambles.

#### 4.5. Are tests of CPT unfair?

It is sometimes claimed that because CPT has been axiomatized, it must satisfy "axioms" like branch independence. Although branch independence is a clear principle, it is certainly not an axiom of either CPT or TAX. Branch independence should be violated according to both CPT and TAX (given their prior parameters), so it can hardly be an axiom of either class of theories. These prior CPT and TAX models predict opposite types of violations of this property, however, when their parameters have been chosen to explain the Allais paradoxes.

In this paper, CPT was granted its equation (which implies violation of branch independence) and the option of invoking the editing rule of cancellation (which satisfies branch independence). The present data, however, show that even with this extra flexibility CPT can still be rejected. Because CPT was allowed two of three possible outcomes, whereas TAX was granted only one, the test was not equally "fair" to both models. The point is, however, that the theory that was granted the

larger space of compatible outcomes was the one rejected by the data.

Prospect theories were also granted a larger space of possibilities in Table 1, since they were allowed to invoke or not invoke two editing principles and two equations. It is difficult to calculate how many free parameters are consumed by an editing rule that may or may not be exercised, but it should be clear that the editing rules permit more flexibility to prospect theories than to TAX or RAM.

If we treat the editing rules of combination and cancellation as free-standing scientific hypotheses (rather than as post hoc excuses or as partial tendencies that operate only in some of the people some of the time), then they can be rejected by the present results in Tables 2 and 3. Both of these editing rules are systematically violated by significant departures.

Coalescing was not initially stated as an axiom of RDU, RSDU, or CPT (Tversky & Kahneman, 1992; Luce & Fishburn, 1991, 1995; Wakker & Tversky, 1993). Birnbaum (1997) used coalescing to simplify his proofs of lower and upper cumulative independence and considered coalescing as a testable theorem that can be derived from the RDU/RSDU/CPT representation. Luce (1998) subsequently showed that this simple property can be treated as an axiom, and showed that it forces RDU in the context of a fairly general class of rank-weighted utility models. The point, however, is that the property was derived as a theorem implied by the RDU representation before it was used as an axiom to derive the representation. Thus, it was not the axiomatization of CPT that exposed its vulnerability to coalescing.

These tests of coalescing and branch independence are best considered as tests among implications of theories (theorems) rather than as "axiom" testing.

Dominance detection was also stated as an editing rule in original prospect theory (Kahneman & Tversky, 1979). Stochastic dominance is implied by any parameterization of CPT, however, so systematic departures from stochastic dominance refute CPT, with or without this editing principle. This rule also needs revision; clearly, the majority does not conform to stochastic dominance in the simple choices of Table 4. Presumably, those tests are not "transparent," so people do not recognize dominance in these choices. What is needed, to save this editing rule as a theory, are definitions of "transparent", "translucent" and "opaque," that specify how and when such a dominance detector is used (Birnbaum et al., 1999).

RAM and TAX models have not been axiomatized. This does not mean that the theories are not testable. An axiomatization is a proof that one can deduce theory from primitive assumptions. It is not a derivation of predictions from theory, nor is it a deduction of theory from data. Theory tests, on the other hand, are tests of

propositions of the form, “if that theory is true, then this testable implication follows.” Showing that the implication is false refutes that statement, but showing that implications are acceptable for some data does not bear on the truth of the theory, except by virtue of empirical induction.

Suppose all of the axioms of a system are testable and one has shown that all of these axioms are acceptable when tested in isolation. It does not follow that all theorems derived from those axioms can now be assumed as true empirical statements. For example, it seems plausible that highly educated people might not violate coalescing if it was presented in plain form as a simple indifference. However, it has been shown that quite a few highly educated people do violate coalescing when it is tested in concert with transitivity.

Luce and Marley (pers. comm., June 11, 2003) are currently working on the axiomatic analysis of configural weighting models. They have been working with rank-weighted utility models, of which RAM and TAX are both special cases; however, as of this writing, none of us has yet found what additional assumptions force the particular forms of RAM and TAX. Nor is it yet understood how to axiomatize certain other configural weight models (e.g., Birnbaum & McIntosh, 1996, Appendix) that are not special cases of rank weighted utility. See Luce (submitted) for a summary.

RAM implies distribution independence, whereas TAX and CPT do not. This illustrates that there are indeed properties that are satisfied by RAM and violated by CPT. RAM satisfies asymptotic independence (Birnbaum, 1997), whereas TAX does not. Although RAM and TAX make the same directional predictions in this paper, empirical evidence has favored TAX over RAM when the models are directly compared (Birnbaum & Chavez, 1997).

#### 4.6. The domain of predicted violations

When a model violates a principle, it does not mean that people are expected to always violate it or even that they should “frequently” violate it, only that they will violate it in certain special situations, which should be predictable by calculations from parameter estimates fit to previous data.

To assess how “often” TAX violates stochastic dominance, “random” 3-branch gambles were sampled in the following way: First, three probabilities were selected by random numbers uniformly distributed from 0 to 1. These were each divided by their sum so they would sum to 1. Next, prizes were randomly sampled, uniformly distributed between \$0 and \$100. “Random choices” were constructed by independently selecting pairs of such “random” gambles. In 1,000,000 such random choices, it was found that 33.3% had a stochastic dominance relationship, but

only 163 of these one million cases produced a predicted violation of stochastic dominance by the TAX model with its prior parameters. Put another way, if an experimenter randomly picked 1,000 random choices by this scheme, the odds would be more than 7:1 against having even a single case of a predicted violation of stochastic dominance according to the prior TAX model.<sup>4</sup>

In addition, it was found that TAX and CPT with their prior parameters agree in 94.4% of such random choices. Thus, one should not expect these theories to disagree in general. If one wants to compare the models, we need to test their distinguishing predictions rather than explore random or haphazardly devised choices.

#### 4.7. Do RAM and TAX have “unfair” advantages?

It might be argued that RAM and TAX have unfair advantages over CPT. The configural weight models have been studied a bit longer than prospect theory, and they had been developed to describe earlier findings in judgment (Birnbaum, 1973, 1974) that are apparently applicable to risky decision making. These models also have the advantage that they had been developed to provide comprehensive numerical fits to several properties of judgment data, including buying and selling prices (Birnbaum & Stegner, 1979).

The predictions from RAM and TAX models are algebraic derivations from the representations and calculations made with parameters estimated from previous data to make predictions to new experiments. The use of prior parameters and the TAX model to calculate how and where to find deviations from CPT is not trivial, for there are many experiments one could do in which a property will not be violated. An experiment is something like a search for a key that may have been lost at the mall. Even if there is a key, there are many places to search where no key will be found. A theory that tells us where to look can be a great advantage over unguided search.

Just as these models only rarely violate stochastic dominance, TAX and RAM do not always violate branch independence. It takes careful planning to calculate how and where to find violations (Birnbaum & McIntosh, 1996). TAX and its prior parameters informed us where to search for the key. To a person who maintains there is no key, however, it might seem “unfair” to use a rival theory to intentionally search where the key should be found.

<sup>4</sup>A free calculator to make the random choices is available from the following URL: <http://psych.fullerton.edu/mbirnbaum/calculators/RandomGambles4.htm>

#### 4.8. Psychological intuitions of prospects versus branch weighting

Prospect theory and cumulative prospect theory assume that people frame and edit risky gambles as “prospects.” One of the properties of a prospect is that it satisfies coalescing, either by the editing rule of combination (Kahneman & Tversky, 1979; Kahneman, 2003) or by the RDU representation used in CPT (Tversky & Kahneman, 1992), which implies coalescing (Birnbaum & Navarrete, 1998, p. 57–58). Therefore, two ways of presenting the same prospect, with branches split or coalesced, should be evaluated the same, if risky gambles are framed as psychological prospects.

Kahneman (2003) described the history of his collaboration with Amos Tversky on prospect theory. They thought that their original formulation made a number of predictions that they imagined would be disproved. They added editing principles that contradicted implications of the model that they thought were false. Kahneman (2003, p. 727) wrote, “...we proposed that decision makers, prior to evaluating the prospects, perform an editing operation that collects similar outcomes and adds their probabilities. He judged that they “had also made a truly significant advance, by making it explicit that the objects of choice are mental representations, not objective states of the world.”

The configural weight models of Birnbaum (1973) and Birnbaum and Stegner (1979) also use mental representations, but different ones. In those representations, two sources that provide the same message cannot be combined by simply adding their weights. RAM and TAX models treat risky gambles as trees with branches rather than as prospects. The same prospect can be presented in many different trees, by splitting or coalescing branches. In these models, the carriers of weight are probability–consequence branches. One can manipulate the weight of a branch by changing its probability and the relationship between a branch’s consequence and the consequences and probabilities of other branches. Birnbaum and Stegner (1979) also proposed that the relative weights of ranked branches would be affected by changes in the viewpoint of the judge, who might be asked to judge buyer’s, seller’s, or neutral’s prices (see also Birnbaum & Sutton, 1992; Birnbaum, Coffey, Mellers, & Weiss, 1992; Birnbaum, Yeary, Luce, & Zhou submitted).

If people view risky gambles as trees, they are not expected to be indifferent to splitting or coalescing branches. If the probability function is a negatively accelerated function of branch probability (i.e.,  $\gamma < 1$ ), splitting a branch will increase the total relative weight of the splinters and thus reduce relative weights of other branches. Similarly, if risk aversion is represented by transfers of weight from higher to lower ranked branches, splitting branches can also change the relative

weights of other branches (Fig. 3). Both prospect theory and cumulative prospect theory, despite their acknowledgment of various framing effects, require that two different trees representing the same psychological prospect will be evaluated identically.

CPT, RAM, and TAX are all behavioral, psychological models that attempt to describe how people interpret, perceive, evaluate, and choose between risky gambles. These three models share several other principles in common. All three models allow a utility (or value) function defined on changes from a status quo. All three models allow weighting of the consequences to be affected by a consequence’s probability and its position relative to other consequences. The three models differ, however, in how this configural weighting is calculated. The prospect theories assume or imply coalescing, a property that some might call “rational,” since splitting or coalescing branches does not alter the objective probabilities of the consequences. But according to Kahneman (2003, p. 727), the reason this property was assumed was because it was thought to be descriptive of how people edit prospects.

If the probability function [ $s(p)$  or  $t(p)$ ] is negatively accelerated, both RAM and TAX imply that certainty equivalents of binary gambles with fixed consequences will be an inverse-S function of probability to win the higher consequence (Figs. 1 and 2). So all three models agree on this prediction, first reported for bids by Preston and Baratta (1948) and replicated in choices between gambles and sure cash by Tversky and Kahneman (1992).

In CPT, this empirical inverse-S relationship between certainty equivalents and probability in binary gambles is taken as the cumulative probability weighting function (apart from the utility transformation). The psychological “intuition” is that the subjective impact of a change in probability from impossible to possible and a change from nearly certain to certain are larger psychological changes than an equal numerical change near 1/2. In my opinion, however, this “intuition” simply restates this theoretical cumulative weighting function in words.

The inverse-S weighting function can be tested by exploring its implied violations of restricted branch independence and distribution independence (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997), two properties that should be violated in particular ways if the inverse-S cumulative weighting function is correct. The CPT model with its inverse-S weighting function implies violations of both of these properties. Both properties are indeed violated; but the empirical data show the *opposite* patterns of violation from those predicted by the inverse-S weighting function. Because the inverse-S cumulative weighting function yields wrong predictions, either the data, the weighting function, or the CPT model must be wrong. In my opinion, the best interpretation of the evidence is that

the “intuition” of the inverse-S is fine as a description of one empirical result, but the theory behind it is just wrong, because other implications of this theory are contradicted by empirical evidence.

#### 4.9. Final comments on Allais paradoxes

The history of research on the Allais paradox has perhaps been retarded by attention to irrelevant details of the choices as originally constructed by Allais (1953). The original version used hypothetical choices between chances to win very large amounts of money, and it used a choice between a pair of two-branch gambles in one case and a choice between a sure thing and a three-branch gamble in the other.

The consequence of zero (win nothing) was also a prominent feature of early tests. Indeed, the consequence of zero has been shown to behave differently from nonzero consequences in tests of consequence monotonicity (Birnbaum, 1997). These two cases (of gambles with and without the zero consequence) were treated differently by Edwards (1962) and by Kahneman and Tversky (1979).

Many ideas proposed to explain the paradox were based on such details of the original form of the paradox; for example, see various chapters in Allais and Hagen (1979). It was proposed, for example, that the paradox might go away with monetary prizes of modest value, with real consequences (Camerer & Hogarth, 1999), with intellectual arguments (Savage, 1972; Slovic & Tversky, 1974), with different formats or representations (e.g., Keller, 1985), or without sure things. Kahneman and Tversky (1979), for example, described the common consequence paradox as an illustration of the “certainty effect,” which they represented as a discontinuity in the weighting function,  $w(p)$  as  $p \rightarrow 1$ .

The present data and those of others show that none of those features are crucial to the phenomenon. In this study, Allais paradoxes are found with small sums of money, with real prizes at stake, with comparisons between gambles having equal numbers of branches, without the use of the zero consequence, and without the need of sure things. They are found with or without event framing.

Many studies have been conducted in attempts to “avoid” paradoxical behavior by some experimental manipulation. The literature can be summarized as follows: the variable tested failed to eliminate the paradox. This paper follows in that tradition, with one exception. The exception in this study is that a variable has been identified that not only undoes the Allais paradox, it significantly reverses it. That variable is the splitting or coalescing of branches, which appears to give the best explanation of common consequence paradoxes.

## References

- Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, *46*, 1497–1512.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *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.
- Allais, M., & Hagen, O. (Eds.). (1979). *Expected utility hypothesis and the Allais paradox*. 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. (1999a). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B. A. Mellers, & D. A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27–52). Norwell, MA: Kluwer Academic Publishers.
- Birnbaum, M. H. (1999b). Testing critical properties of decision making on the Internet. *Psychological Science*, *10*, 399–407.
- Birnbaum, M. H. (2000). Decision making in the lab and on the Web. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 3–34). San Diego, CA: Academic Press.
- Birnbaum, M. H. (2001a). *Introduction to behavioral research on the Internet*. Upper Saddle River, NJ: Prentice Hall.
- Birnbaum, M. H. (2001b). A Web-based program of research on decision making. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 23–55). Lengerich, Germany: Pabst.
- Birnbaum, M.H. (submitted). Tests of rank dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effect of format, event framing, and branch splitting. 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*(2), 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., & Martin, T. (2003). Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on decision research* (pp. 84–107). New York: Cambridge University Press.
- 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., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and

- cumulative independence. *Journal of Risk and Uncertainty*, 17, 49–78.
- Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes*, 77, 44–83.
- 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., & Yeary, S. (submitted). Tests of stochastic dominance and cumulative independence in buying and selling prices of gambles, submitted for publication.
- Birnbaum, M. H., Yeary, S., Luce, R. D., & Zhou, L. (submitted). Empirical evaluation of theories for buying and selling prices of binary gambles, submitted for publication.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production theory. *Journal of Risk and Uncertainty*, 19, 7–42.
- Chateauneuf, A., & Wakker, P. (1999). An axiomatization of cumulative prospect theory for decision under risk. *Journal of Risk and Uncertainty*, 18, 137–145.
- Diecidue, E., & Wakker, P. P. (2001). On the intuition of rank-dependent utility. *Journal of Risk and Uncertainty*, 23, 281–298.
- Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science*, 10, 353–359.
- Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, 69, 109–135.
- Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75, 643–649.
- 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.
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129–166.
- 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. (2003). Experiences of collaborative research. *American Psychologist*, 58, 723–730.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Karmarkar, U. S. (1979). Subjectively weighted utility and the Allais paradox. *Organizational Behavior and Human Performance*, 24, 67–72.
- Keller, L. R. (1985). The effects of problem representation on the sure-thing and substitution principles. *Management Science*, 31, 738–751.
- Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*, 43, 286–313.
- Luce, R. D. (1998). Coalescing, event commutativity, and theories of utility. *Journal of Risk and Uncertainty*, 16, 87–113.
- Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Luce, R. D. (submitted). Utility representations of gambles: Old, new, and needed results.
- Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty*, 4, 29–59.
- Luce, R. D., & Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty*, 11, 5–16.
- Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology*, 29, 1–72.
- Preston, M. G., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology*, 61, 183–193.
- Quiggin, J. (1985). Subjective utility, anticipated utility, and the Allais paradox. *Organizational Behavior and Human Decision Processes*, 35, 94–101.
- Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model*. Boston: Kluwer.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Savage, L.J. (1972). *The foundations of statistics* (second revised edition), New York: Dover Publications, Inc.
- Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science*, 19, 368–373.
- Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38, 332–382.
- 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.
- 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.
- Wakker, P. P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, 7, 147–176.
- 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.
- 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.
- Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty*, 16, 115–139.