

Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting[☆]

Michael H. Birnbaum^{*}

*Department of Psychology, California State University, Fullerton and Decision Research Center, H-830M,
P.O. Box 6846, Fullerton, CA 92834-6846, USA*

Received 28 April 2003
Available online 17 July 2004

Abstract

Four studies with 3440 participants investigated four new paradoxes that refute rank-dependent utility and cumulative prospect theories of risky decision making. All four paradoxes can be interpreted as violations of coalescing, the assumption that branches leading to the same consequence can be combined by adding their probabilities. These studies explored if there is some format in which coalescing and cumulative prospect theory would be satisfied. Three variables were manipulated: probability format, branch splitting, and event-framing. Probability was displayed via text, pie charts, natural frequencies, and lists of equally likely consequences. Probability format and event framing had minimal effects, but branch splitting had large effects. In all 12 conditions of format and framing, splitting created majority violations of stochastic dominance and a second round of splitting reversed preferences, restoring majority satisfaction of stochastic dominance. Two cumulative independence properties were violated in 47 of 48 new tests. Birnbaum's TAX model, in which the relative weight of each probability-consequence branch depends on probability and rank of its consequence, correctly predicted the main trends. In this model, splitting the branch with the lowest consequence can make a gamble worse, and splitting the branch with the highest consequence can make a gamble better.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Allais paradoxes; Branch weighting; Coalescing; Configural weighting; Cumulative independence; Cumulative prospect theory; Decision making; Descriptive decision theory; Event-splitting; Expected utility; Format; Paradoxes; Rank- and sign-dependent utility; Rank-dependent utility; Risk; Stochastic dominance

Recent studies reported new paradoxes of choice that violate a number of models proposed as descriptive theories of choice (Birnbaum, 1999a, 1999b; Birnbaum, Patton, & Lott, 1999; Birnbaum & Navarrete, 1998). Evidence is rapidly accumulating that refutes a class of models including Rank-Dependent Expected Utility (RDU) theory (Camerer, 1992; Quiggin, 1985, 1993), Rank- and Sign-Dependent Utility (RSDU) theory

(Luce, 1998, 2000; Luce & Fishburn, 1991, 1995), and Cumulative Prospect (CPT) theory (Chateauneuf & Wakker, 1999; Diecidue & Wakker, 2001; Tversky & Kahneman, 1992; Tversky & Wakker, 1995; Wakker & Tversky, 1993; Wu & Gonzalez, 1996, 1998).

Among these new paradoxes is a recipe by Birnbaum (1997) that creates majority violations of stochastic dominance. If the probability of winning x or more in gamble A is greater than or equal to this same probability in gamble B for all x , it would be a violation of stochastic dominance if people systematically preferred B over A . After the recipe for creating violations had been proposed, Birnbaum and Navarrete (1998) tested it empirically, and found that this recipe produced significantly more than 50% violations.

[☆] Support was received from National Science Foundation Grants, SBR-9410572, SES 99-86436, and BCS-0129453. I thank R. Duncan Luce and Jerome Busemeyer for helpful suggestions on an earlier draft; thanks are also due Sandra Schneider, Christof Tatka, and Peter Wakker for useful discussions of these issues.

^{*} Fax: 1-714-278-7134.

E-mail address: mbirnbaum@fullerton.edu.

Empirical violations of stochastic dominance not only violate the class of RDU/RSDU/CPT theories, but also Becker and Sarin's (1987) Lottery-Dependent Utility (LDU), Lopes and Oden's (1999) Security Potential-Aspiration Level (SP-A) theory, and Machina's (1982) generalized utility (GU), among others.

This body of new evidence runs against a current trend in Economics whereby the rank-dependent models have been gaining increasing acceptance as empirical descriptions of choices between risky and uncertain prospects (Starmer, 2000). Indeed, a share of the 2002 Nobel Prize in Economics was awarded to Daniel Kahneman in recognition of his work on prospect and cumulative prospect theory and for the empirical tests of these theories against expected utility (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992).

Because the implications of these new paradoxes run counter to the current trend in Economics and are so important to descriptive theory, it is vital to determine if they occur only with specific methods or if they are more generally descriptive of human decision making. If results are contingent on procedure, it not only allows certain theories to survive in a limited domain, but it also provides a better understanding of the causes of violations. Classic paradoxes that refuted EU were tested in this way (e.g., see the chapters in Allais & Hagen, 1979; Kahneman & Tversky, 1979; Keller, 1985), and these new paradoxes that refute CPT require the same scrutiny before they will be accepted.

It has been suggested that people are better able to reason with probability when the information is formatted in terms of natural frequencies rather than as probabilities or percentages (Gigerenzer & Hoffrage, 1995). These authors noted that Kahneman and Tversky's (1973) apparent evidence for base rate neglect was undone when probabilities are expressed as frequencies. The fragile case of the so-called "base rate fallacy," a result that crumbles in within-subjects designs and with other variations of procedure stands as a warning to investigators not to invest in theory building until the facts and artifacts have been sorted out (Birnbaum & Mellers, 1983).

Studies showing significant majority violations of stochastic dominance displayed gambles with decimal probabilities. Therefore, it seems important to determine whether violations are reduced when choices are represented by natural frequencies, by enumerated lists of equally likely consequences, by graphical charts, or by other variations that might help people "see" dominance. Three variables of procedure and display will be manipulated in these studies, *branch splitting* (gambles may be presented with certain branches in split or coalesced form), *probability format* (how probabilities are represented to the participant), and *event-framing* (how the outcomes of the random event of a gamble are labeled). The distinctions among these variables will be made clear below.

Other tests of stochastic dominance (Diederich & Busemeyer, 1999; Loomes, Starmer, & Sugden, 1992; Tversky & Kahneman, 1986) reported "high" rates of violations; however, none of these cases reported significantly more than 50% violations by a conventional two-tailed statistical test at the .05 level of significance. Therefore, a skeptic could argue that those manipulations merely confused some participants so that they reverted to choosing randomly, which would cause choice proportions to regress toward 50% violations. I have not yet found any publication previous to Birnbaum and Navarrete (1998) that reported significantly more than 50% violations.

Four new paradoxes of choice

Kahneman and Tversky (1979) criticized expected utility theory based on variations of the two Allais paradoxes (Allais, 1953, 1979). This paper explores four new paradoxes that violate the class of RDU/RSDU/CPT in the same way that the Allais paradoxes contradict expected utility. That is, these paradoxes are self-contradictions within a theory that result from attempting to apply that theory to empirical data. In this section, these four "new" paradoxes will be illustrated by example; formal definitions are presented in Appendix A, and proofs are presented in Birnbaum and Navarrete (1998). Two definitions, however, are needed here for the exposition.

First, a *branch* of a gamble is a probability-consequence pair that is distinct in the gamble's presentation to the participant. Second, *coalescing* is the assumption that if a gamble has two or more branches that lead to the same consequence, these branches can be combined by adding their probabilities without affecting the utility (subjective value) of the gamble. For example, consider gamble *A*, presented as follows:

A: .10 probability to win \$100
 .10 probability to win \$100
 .80 probability to win \$10

Gamble *A* has three distinct branches, but two of these branches lead to the same consequence (\$100). According to coalescing, gamble *A* should be indifferent to a two branch gamble, *A'*, which is (objectively) the same as *A*, except that in *A'* the two branches of *A* that yield \$100 have been combined as follows

A': .20 probability to win \$100
 .80 probability to win \$10

Gamble *A'* is called the *coalesced* form of *A*. Gamble *A* will be described as a *split* form of *A'*. It should be clear that there is only one way to coalesce, from *A* to *A'*, but there are many ways to split *A'*, and *A* is only one of the split forms of *A'*. Gambles will also be denoted as

follows: $A = (\$100, .1; \$100, .1; \$10, .8)$ and $A' = (\$100, .2; \$10, .8)$.

Kahneman and Tversky (1979) assumed that “most decision makers will spontaneously transform the former prospect into the latter and treat them as equivalent in subsequent operations of evaluation and choice.” (Kahneman, 2003, p. 727). Coalescing follows from RDU/RSDU/CPT with or without this editing principle of combination (Birnbaum & Navarrete, 1998; Luce, 1998).

For example, in the model of cumulative prospect theory fit to previous data (see Appendix B), the calculated values of A and A' both have the same value, 20.6. In gamble A , the two branches to win \$100 and \$100 have weights of .186 and .075, respectively. In gamble A' , the weight of the coalesced branch to win \$100 is .261, which is the sum of the weights of the branches that were coalesced. The property of coalescing follows from the representation (Expression (B.1) in Appendix B), so it holds not only for the specific model fit by Tversky and Kahneman (1992); rather, it holds for any RDU/RSDU/CPT model.

Violations of coalescing: Event-splitting effects. The first of the new paradoxes are violations of coalescing. According to the class of RDU/RSDU/CPT models, coalescing must be satisfied, aside from random error. Event-splitting effects, studied by Starmer and Sugden (1993) and by Humphrey (1995), are violations of coalescing, if one assumes transitivity. With $A \succ B$ representing that A is preferred to B , coalescing and transitivity imply that $A \succ B \iff A' \succ B \iff A' \succ B'$, where A' and B' are coalesced versions of A and B , respectively. Luce (1998, 2000) criticized these studies, however, noting that they were based on between-subjects comparisons. The studies in this paper use Birnbaum’s (1999b) design, which demonstrated large violations of coalescing within subjects (see also Birnbaum & Martin, 2003).

Violations of stochastic dominance. The second of the new paradoxes to be tested in this study is Birnbaum’s (1997) recipe for constructing choices between three branch gambles in which significantly more than half of the participants violate stochastic dominance. If the probability of winning x or more in gamble A is greater than or equal to the probability of winning x or more in gamble B for all values of x , and if this probability is strictly higher for at least one value of x , we say that A stochastically dominates B .

Birnbaum (1997) designed the following choice:

- I : .05 probability to win \$12
 .05 probability to win \$14
 .90 probability to win \$96
- J : .10 probability to win \$12
 .05 probability to win \$90
 .85 probability to win \$96

Gamble I stochastically dominates J because the probabilities of winning \$14 or more and of \$96 or more are greater in gamble I than J , and the probabilities of winning \$12 or more and of \$90 or more are the same in both gambles. According to Birnbaum’s (1997) models with parameters estimated from previous data, however, people will choose J over I , violating dominance. Birnbaum’s (1997) prediction had been set in print before it was tested. For example, the prior TAX model (see Appendix B) predicts that J has a value of 63.1, higher than that of I , 45.8.

Birnbaum and Navarrete (1998) tested this prediction and found that about 70% of 100 undergraduates violated stochastic dominance in this choice and three others like it. Birnbaum et al. (1999) replicated these results with five new variations and 110 new participants. In all nine tests, significantly more than half of the participants violated stochastic dominance.

Birnbaum’s (1997) RAM and TAX models violate coalescing, and this property was used to construct the recipe for violations of stochastic dominance. According to those models, splitting the lower branch increases the total weight given consequences on those splinters and makes a gamble worse, whereas splitting branches that lead to better consequences increases their weight and thus makes a gamble better. The choice between I and J was constructed from the following “root” gamble:

- H : .10 probability to win \$12
 .90 probability to win \$96

To create gamble J , split the higher-valued branch of H (\$96, .9) into two splinters (\$96, .85) and (\$96, .05); next, lower the consequence on the .05 splinter from \$96 to \$90. Gamble J is therefore dominated by H . Now, construct I from the root gamble by splitting the lower-valued branch (\$12, .10) into two splinters, and increasing the consequence on one (\$12, .05) splinter from \$12 to \$14. Gamble I thus dominates H which dominates J .

If branch splitting can be used to create the violations, then it should be possible to use a second round of splitting to undo them. Note that the following choice is (objectively) equivalent to the choice between I and J , except for coalescing/splitting:

- I'' : .05 probability to win \$12
 .05 probability to win \$14
 .05 probability to win \$96
 .85 probability to win \$96
- J'' : .10 probability to win \$12
 .05 probability to win \$12
 .05 probability to win \$90
 .85 probability to win \$96

Although RDU/RSDU/CPT imply the same decision between I'' and J'' as between I and J , Birnbaum’s prior TAX model implies that people will satisfy stochastic

dominance in this split form of the choice, since the value of I'' is 53.1, which is greater than the calculated value of J'' , 51.4. In contrast, any RDU/RSDU/CPT model implies that people should satisfy stochastic dominance in both cases. For example, the prior CPT model (see Appendix B) implies that the values of I and I'' are both 42.2, better than the values of J and J'' , which are both 41.9. One can also consider these choices in reverse order to understand violations of stochastic dominance as a violation of at least one of three properties: coalescing, consequence monotonicity, or transitivity (see Appendix A).

Violations of lower and upper cumulative independence. Birnbaum (1997) derived two other theorems from CPT that were predicted to be violated by RAM and TAX, creating two new paradoxes. These predictions had also been set in print before Birnbaum and Navarrete (1998) conducted their experiment to test them. As shown in Birnbaum et al. (1999, pp. 79–80), violations of these properties create contradictions in RDU/RSDU/CPT: if paradoxical choices are observed, there is no cumulative weighting function and utility function that will reproduce the results in these models.

Upper cumulative independence is illustrated by the following choices:

- S' : .80 probability to win \$108
 .10 probability to win \$44
 .10 probability to win \$40
- R' : .80 probability to win \$108
 .10 probability to win \$98
 .10 probability to win \$10

and

- S''' : .80 probability to win \$98
 .20 probability to win \$40
- R''' : .90 probability to win \$98
 .10 probability to win \$10

According to the class of RDU/RSDU/CPT models, if a person chooses R' over S' , then she or he should choose R''' over S''' . Here, we have reduced the consequence on the common branch from \$108 to \$98 on both sides, but we have reduced the consequence on the branch to win \$44 to \$40 on the left side only. That means that the utility difference favoring the right side increases; i.e., $U(R''') - U(S''') > U(R') - U(S')$. For example, in the prior CPT model, the values are $U(R''') = 42.4$, $U(S''') = 44.4$, $U(R') = 46.1$, and $U(S') = 48.3$, which satisfy the inequality, $42.4 - 44.4 > 46.1 - 48.3$.

According to Birnbaum's models, however, we have coalesced the two lower branches in S' , making the left side better, and we have coalesced the two higher branches in R' , making the right side relatively worse.

Birnbaum's prior TAX model implies that these changes in weight overcome the small reduction in value from \$44 to \$40. For this model, $U(S') = 65.0 < U(R') = 69.6$ and $U(S''') = 68.0 > 58.3 = U(R''')$, so people should switch preferences in the opposite way from that allowed by any version of RDU/RSDU/CPT.

A test of *lower cumulative independence* is illustrated by the following choices:

- S : .10 probability to win \$44
 .10 probability to win \$40
 .80 probability to win \$2
- R : .10 probability to win \$98
 .10 probability to win \$10
 .80 probability to win \$2

and

- S'' : .20 probability to win \$44
 .80 probability to win \$2
- R'' : .10 probability to win \$98
 .90 probability to win \$10

We have increased the consequence on the common branch from \$2 to \$10, which should not change the preference order. Next, we increase the consequence on the (\$40, .10) branch of S , and coalesce the two upper branches, yielding S'' . Finally, coalesce the two branches in R that now yield \$10. According any parameterization of RDU/RSDU/CPT, therefore, if a person prefers S to R , then he or she should prefer S'' to R'' , apart from error, because we have increased one of the consequences of S without making a corresponding change in R . These models imply $U(S'') - U(R'') > U(S) - U(R)$. For the prior CPT model, for example, $12.9 - 16.7 > 8.5 - 12.5$.

In Birnbaum's models, however, we have coalesced the two higher-valued branches of S , making S relatively worse, and we have coalesced the two lower branches in R , making it better. According to the prior TAX model, the effects of coalescing outweighs the increase in value from \$40 to \$44; $U(S) = 11.4 > 10.9 = U(R)$ and $U(S'') = 16.2 < 20.4 = U(R'')$, so people should be more likely to switch in the opposite direction from that allowed under any version of RDU/RSDU/CPT.

Goals of the present studies

Because violations of stochastic dominance can be considered violations of rationality as well as of descriptive theory, it seems important to determine how to manipulate their incidence for both practical and theoretical purposes. If we can find formats for presenting choices that substantially reduce violations of stochastic

dominance, it could be useful in aiding people to make better decisions, and if one finds a way of maximizing them, it could provide a way to make money.

For the theoretician, being able to control (i.e., to reduce or increase) violations means that we understand something about their causes.

According to Birnbaum’s RAM and TAX models, violations of coalescing produce violations of stochastic dominance and produce the paradoxical violations of lower and upper cumulative independence. Intuitively, splitting a branch increases the relative weight of the splinters, and coalescing reduces the weight of a branch. Thus, splitting the branch of a gamble with the lowest consequence makes that gamble worse, and splitting the upper branch of a gamble makes it better. This intuition explains all four of the new paradoxes studied here.

In contrast, the family of RDU/RSDU/CPT models imply that the sum of the weights of split branches equals the weight of the coalesced branch (Birnbaum & Navarrete, 1998; Luce, 1998).

A number of people proposed manipulations that they thought would reduce violations of these new paradoxes. For example, Sandra Schneider (personal communication, November, 1999) suggested that with graphical displays, people might be able to visualize the combination of split branches and therefore perceive stochastic dominance. Christof Tatka (personal communication, June, 2000) conjectured that branch weights may depend on order of presentation rather than rank, and he proposed that if the order of branches were reversed, results theorized to depend on branch ranks would also reverse. The first study of this paper therefore tested whether the use of pie charts to display probability or reversing the presentation order of

branches would affect violations of RDU/RSDU/CPT models.

Fig. 1 shows how the pie chart format differs from text format, illustrated above. In the pie format in Fig. 1, one might be able to “see” that the sum of the two 0.05 splinters in *I* has the same area as the .10 splinter in *J*, and that the consequence in *I* in any slice is either the same or better than that of *J*, and therefore, that gamble *I* dominates *J*.

The second study was designed to explore the idea that people might reason better with natural frequencies than they do with probabilities. Perhaps humans have a better understanding of counts, lists, or frequencies (ideas taught to young children) than they do of the decimal ratios of probabilities defined on sets (ideas taught to university students).

Among ways of representing probabilities as frequencies, one might provide the number of tickets in an urn with a given payoff (as in Fig. 2), or the number of marbles of a given color, where color of marble drawn determines the prize (Fig. 6). Alternately, one might list the equally likely consequences themselves, so that natural perception of frequency is available instead of a numerical summary. In the *list* formats used here, there were always exactly 20 tickets in each urn, and the payoffs of the 20 tickets were listed, as illustrated in the example of Fig. 3.

The split form of the list format placed each branch’s consequences on a different line (Fig. 4). The four types of displays in this second study were the *tickets* format (Fig. 2), the *list* format (Fig. 3), a *small list* format (like the list format, but with a smaller font, not shown), and the *semi-split list* format, in which no more than 10 ticket values would be presented on each line (Fig. 5). In

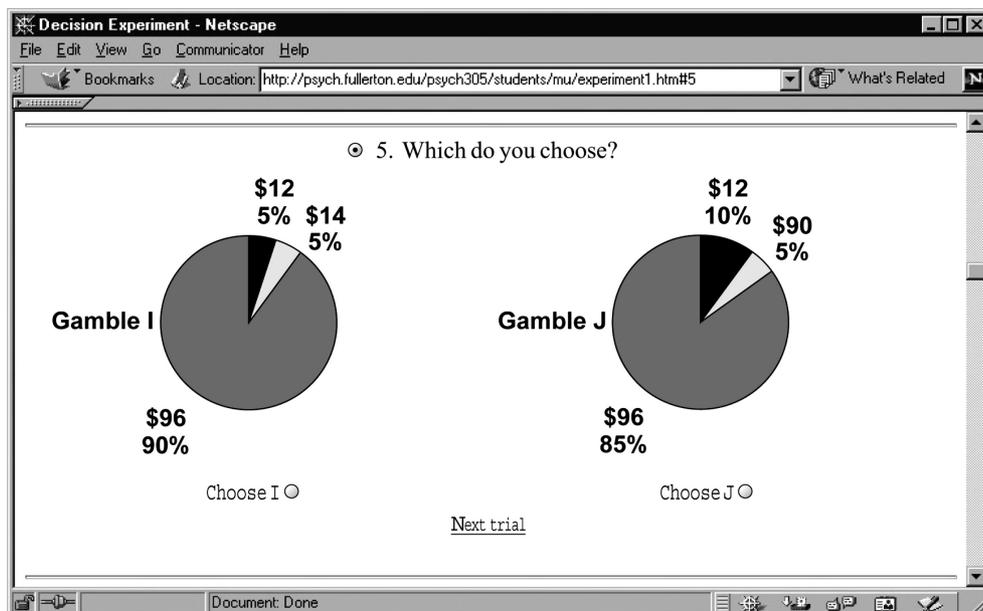


Fig. 1. Example of display of probability by pie charts (pies and pies* conditions).

I:	\$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96,
	\$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96
	\$14
	\$12
OR	
J:	\$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96, \$96,
	\$96, \$96, \$96, \$96, \$96, \$96, \$96
	\$90
	\$12, \$12

Fig. 5. Example of display in “semi-split” format of choice in Fig. 3, in which no more than 10 tickets were listed on each line. The split form used additional lines as in Fig. 3.

5. Which do you choose?	
I:	90 red marbles to win \$96
	05 blue marbles to win \$14
	05 white marbles to win \$12
OR	
J:	85 red marbles to win \$96
	05 blue marbles to win \$90
	10 white marbles to win \$12

Fig. 6. Example of display in urns condition FU, illustrating a “framed” choice in which colors match on corresponding branches. In the “unframed” version of each choice, marble colors are all different in both gambles.

consequences. If this works, it should be possible to increase or decrease incidence of violations of stochastic dominance by this trick.

The term “framing” has been used to refer to many different manipulations, usually in reference to manipulations that do not change the objective situation, but which produce different psychological responses. “Framing” might be used to refer to wordings of consequences as apparent gains or losses (when in fact they are all gains), to different wordings of the same instructions, to different formats for presenting the same underlying gambles, and to other manipulations. In this paper, however, it is important to distinguish these various manipulations, so three terms will be used as follows. The term “format” will be used to refer to different ways to display probabilities and consequences in a gamble (text, pie charts, lists, etc.). The terms “coalesced” or “split” will refer to forms of presentation of the branches of a gamble. The term “event framing” or “framing” will be used only for the manipulation of Tversky and Kahneman (1986) as described below. [The term “juxtaposition” has also been used to describe procedures that create or seem to create correlation in the consequences (Loomes et al., 1992).]

Tversky and Kahneman (1986) argued that event-framing could be used to “mask” a dominance relation or to make it “transparent.” In their framing manipulation, for any color of marble drawn from an urn, the

dominant gamble gave the same or a lower consequence. Because the numbers of marbles of each color were not identical in the two urns, the so-called “events” of drawing different colors were not really the same. In their choice, 58% of 124 people violated stochastic dominance. Although this percentage was not significantly different from 50% ($z = 1.78$), it was greater than the percentage of violations observed among 88 others for whom the gambles had been both differently split and differently framed. In the other version of the same objective choice, the dominant gamble always gave an equal or better consequence for any color of marble drawn, which presumably made dominance “transparent.” In addition, the splitting was such that the number of marbles of each color was identical in the two urns. Event splitting and event framing were therefore confounded in this single test.

Tversky and Kahneman (1986) argued that according to their editing principle of combination (Kahneman & Tversky, 1979), coalescing should not affect choice. They proposed instead that the higher rate of violation of stochastic dominance in their coalesced and framed choice was due to event framing. As Tversky and Kahneman put it (p. S265), “Dominance is masked by a frame in which the inferior option yields a more favorable outcome in an identified state of the world (e.g., drawing a green marble.)” In a sense, participants were “tricked” if they treated the so-called “events” as the same, since the number of marbles of each color differed in the two urns. Tversky and Kahneman (1986) acknowledged that several variables had been confounded in their test, including not only event-splitting and event-framing, but also the *number* of framed “events” in each gamble giving positive or negative consequences in their two choices.

The third study below will employ a “pure” test in which event framing and event splitting will be unconfounded, and all consequences will be positive. A factorial design is used in which each choice will be either framed or unframed and either choice will be presented either coalesced or appropriately split. This design permits us to assess the relative effects of event framing (as

opposed to coalescing) on rates of violation of stochastic dominance.

Fig. 6 shows an example of a choice that is “framed” because the colors of marbles are the same on corresponding ranked branches. According to the framing notion, the judge sees that the number of red marbles to win \$96 is about the same in both alternatives, as is the number of white marbles to win \$12. If these nearly equal branches are cancelled, it leaves an “event” of drawing a blue marble that yields \$90 in one case and \$14 in the other. Event framing should thus increase violations of stochastic dominance in the coalesced form and reduce them in the split form in which the numbers of marbles on corresponding branches are identical. In the unframed version, colors of marbles on all branches of both urns were different.

One theory for violations of stochastic dominance is that people do not properly coalesce branches by adding probabilities or see how to split them to reveal dominance. Indeed, the definition of stochastic dominance is usually written in terms of decumulative probability (the probability of winning x or more). If a person wanted to check for stochastic dominance by comparing decumulative probabilities, that person would have to add. Would some other format for presenting the gambles (aside from coalescing or splitting branches) allow people to obey coalescing (i.e., to add or to see that they should add) and therefore perceive dominance?

The fourth study investigated decumulative probability format—each gamble was described in terms of the probabilities of winning prize x or greater. Decumulative probabilities do not require a person to add to find the probability of winning x or more. An illustration of a trial in this format is presented in Fig. 7. It should be easy to see in Fig. 7 that the probability of winning \$96 or more is higher in the first gamble than the second, that this probability in the first gamble equals the probability of winning only \$90 or more in the second gamble, that probability of winning \$14 or more is higher in the first gamble, and that for other values of x , the probabilities are the same. Furthermore, because CPT is usually written in terms of decumulative probabilities for positive consequences, such a format might facilitate use of CPT, if we think that decision

makers are doing something akin to the equations. Peter Wakker (personal communication, 2003), who still considers CPT the “best” model despite certain contrary evidence, concurred that this format would be an interesting one in which to test CPT.

Method

This paper summarizes new results from four studies, including 3440 people recruited from the subject “pool” of undergraduates and by links on the WWW.

In each condition of four new studies, participants made 20 choices between gambles. Deciders viewed the materials via a browser on the WWW, and clicked a “radio button” (a small circle that fills when clicked) beside the gamble in each pair that they would rather play. Participants were informed that three lucky participants (in each study) would be chosen at random to play one of their chosen gambles for money, with prizes as high as \$110, so they should choose wisely. Prizes were awarded as promised.

The 20 choices included 12 choices described in Tables 1 and 3 that were the same in all conditions. These 12 choices provide two tests each of stochastic dominance (Choices 5 and 7), coalescing (Choices 5 vs. 11 and 7 vs. 13), branch independence (Choices 6 vs. 10 and 17 vs. 12), lower cumulative independence (Choices 6 vs. 8 and 17 vs. 20), and upper cumulative independence (Choices 10 vs. 9 and 12 vs. 14). The two variations of each test were counterbalanced for position (first or second gamble). In addition, there were four warm up trials and four filler trials. These were the same for each condition, except in Study 3 where two fillers were converted to two additional tests of stochastic dominance (repeats of Choices 5 and 7 that were either framed or unframed).

Study 1: Text and pies formats

The probability mechanism in the text conditions of Study 1 was described in the same way as in Birnbaum (1999b). Gambles were described in terms of urns containing tickets that were otherwise identical, except for the prize values printed on them. Participants were told that there were always 100 tickets in an urn. Probability was explained as the ratio of the number of tickets with a given prize to the total number of tickets in the urn. An example of a choice between a “50-50” gamble to win either \$100 or \$0 and another “50-50” gamble to win \$25 or \$35 was presented. The first gamble was described as an urn that contains 100 tickets of which 50 have “\$100” printed on them and 50 have “\$0” printed on them. The other gamble was described as an urn containing 50 tickets that said “\$25” and 50 that said “\$35.” Once an urn was selected, a ticket would be selected blindly and at random from the chosen urn and

5. Which do you choose?

I: .90 probability to win \$96 or more
 .95 probability to win \$14 or more
 1.00 probability to win \$12 or more

OR

J: .85 probability to win \$96 or more
 .90 probability to win \$90 or more
 1.00 probability to win \$12 or more

Fig. 7. Example of display in decumulative probability condition of Study 4.

Table 1
Violations of stochastic dominance and coalescing

No.	First gamble	Second gamble	Birnbaum 99b		Study 1				Study 2				Study 3		Study 4		
			Web A 1224	Lab 124	Text 172	Rev 169	Pies 353	Pies 305	Tick 342	List 377	Small 267	Semi 693	FU 74	UF 83	Cum 445	Con 160	
5	90 (18) of \$96	85 (17) of \$96	58*	73*	63*	59*	66*	65*	71*	61*	64*	56*	70*	76*	80*	59*	
	05 (1) of \$14	05 (1) of \$90											72*	77*			
	05 (1) of \$12	10 (2) of \$12															
11	85 (17) of \$96	85 (17) of \$96	10*	15*	13*	08*	12*	13*	06*	09*	13*	13*	12*	20*	18*	12*	
	05 (1) of \$96	05 (1) of \$90															
	05 (1) of \$14	05 (1) of \$12															
	05 (1) of \$12	05 (1) of \$12															
7	91 (17) of \$99	94 (18) of \$99	46*	64*	56	67*	65*	64*	67*	60*	62*	50	73*	78*	86*	65*	
	03 (1) of \$96	03 (1) of \$8										73*	76*				
	06 (2) of \$6	03 (1) of \$6															
13	91 (17) of \$99	91 (17) of \$99	05*	08*	07*	14*	08*	06*	13*	16*	14*	15*	14*	17*	14*	16*	
	03 (1) of \$96	03 (1) of \$99															
	03 (1) of \$6	03 (1) of \$8															
	03 (1) of \$6	03 (1) of \$6															

Notes. (a) The first gamble is dominant in Choices 5 and 11 and the second gamble is dominant in Choices 7 and 13. (b) Numbers in parentheses show the actual number of values displayed in the list format, which had 20 equally likely values in each gamble. (c) Choice percentages in bold show results for framed choices in which the marble colors on corresponding branches were identical. Asterisks indicate choice percentages that are significantly different from 50%. Bold entries in Study 3 results for framed choices (in which marble colors of corresponding branches were identical).

Table entries are percentages of violation of stochastic dominance.

the value printed on the ticket would determine the prize awarded.

Choices were displayed as in the following example:

1. Which do you choose?
 A: .50 probability to win \$0
 .50 probability to win \$100
 OR
 B: .50 probability to win \$25
 .50 probability to win \$35

Notice that branches were listed in ascending order of their consequences. Complete materials for this and all other conditions of this article can be viewed via the WWW from the following URL: <http://psych.fullerton.edu/mbirnbaum/archive.htm>

There were four conditions in Study 1 with a total of 999 participants who received either a text display (with decimal probabilities and branches in either ascending or descending order of consequences) or a graphic display in which probabilities were represented by slices on a pie chart, as illustrated in Fig. 1. In two *pies* conditions, participants were instructed to imagine a spinner mounted at the center of each pie that was equally likely to land in any two equal sized slices of the pie. Prizes would be determined by where the spinner landed.

The *text* display listed branches in ascending order with respect to value of consequences, as in Birnbaum (1999b). The condition in which branches are presented in descending order of consequences is termed the *reversed* display. The *pies* condition was paired with the ascending text display, and the *pies** condition was paired with the reversed text display. Within each half-study, participants were randomly assigned to text or pie charts, with a 1/3 probability of assignment to text and 2/3 to pies.

Study 2: Frequency and list formats

In the *frequency* format (*tickets*), each branch consisted of the number of (equally likely) tickets for each prize in an urn containing 100 tickets from which one ticket would be drawn blindly at random to determine the prize. The choice between *I* and *J* in this case would be presented as in Fig. 2.

There were three *list* formats, with different participants in each. In each case, the list formats represented each gamble as a list of 20 equally likely prizes. Gambles had exactly 20 tickets in each urn, with each .05 of probability represented by 1 ticket. The values of each of the 20 tickets were listed. Thus, this way of presenting frequencies allows the participant to “see” each instance instead of seeing a number that represents the total frequency. Each “branch” was a list presented on a separate line. The translation from decimal probabilities to lists of equally likely prizes required rounding in Choices 7 and 11 in Table 1, where .03 in probability was rounded to one ticket. The numbers of tickets in this

format are shown in parentheses in the descriptions of the choices in Tables 1 and 3.

A coalesced choice in the *list* condition was displayed as in Fig. 3, with a new line for each distinct value. The “split” form of the same choice was presented by placing the split branch on a new line, as shown in Fig. 4.

Small font: a second variation employed a smaller font, so that less (or zero) scrolling would be required of those using certain browsers, monitors, and display settings.

Semi-split: a third variation conformed to the following rules: no more than 10 tickets could be displayed on one line. This “semi-split” version of the choice between *I* and *J* is displayed in Fig. 4. The fully split choices in this third variation followed the same rules (no more than 10 to a line), except that new lines were again used to split branches.

Study 3: Event framing vs. event splitting

Study 3 described probability in terms of the number of marbles of each color, from which one marble would be drawn at random, and the color of the drawn marble would determine the prize. An example of this format is shown in Fig. 6.

There were two conditions, with random assignment to conditions. Each choice was either “framed” by having the same colors of marbles on corresponding branches, or “unframed” by having different colors of marbles in the urns of the two gambles. Framing was counterbalanced both within and between subjects. Each coalesced choice was presented in both framed and unframed versions to each person. Choices that were framed in condition FU (same color marbles on corresponding branches) were unframed (different colors on corresponding branches) in UF and vice versa.

Study 4: Decumulative probabilities

Participants were randomly assigned with odds of 3:1 to receive the decumulative probability format (as in Fig. 7) or the reversed text format (as in Study 1), which served as its control.

Participants

In Study 1, there were 999 who were randomly assigned to receive either text displays or pie charts. Of these, 172 received the *text* display (ascending order of consequences), 169 saw *reversed* text (with branches presented in descending order of consequences), 353 viewed *pie* charts, and 305 were in the *pies** condition (which was yoked with the reversed text condition).

In Study 2, there were 1679 people of whom 342 received the *tickets* (i.e., *frequency*) format, 377 viewed the *list* format. Participants were randomly assigned to tickets or list formats. In addition, there were two

separate follow up studies in which 267 viewed lists displayed in *small* font, and 693 received the *semi split* format in which no more than 10 prizes could be listed on a line.

In Study 3, 157 undergraduates were randomly assigned to two counterbalanced conditions with *marbles* in urns, of which 74 and 83 were in the FU and UF framing conditions, respectively.

In Study 4, 605 participants were randomly assigned to two conditions with odds of 3:1 of being assigned to the decumulative probability condition and otherwise to the control condition, which was the same as the reversed text condition. Of these, 445 completed the decumulative condition and 160 served in reversed text condition.

Included for comparison in Table 1 are previous results from 1224 additional participants recruited by Birnbaum (1999b) using methods intended to reach highly educated people (Sample A), and 124 undergraduates who made each choice twice (Lab).

Results

Violations of stochastic dominance and coalescing

Table 1 shows percentages of violation of first-order stochastic dominance in each condition for two choices that were either split or coalesced, with the dominant gamble first or second. Columns in the right side of the table represent studies and conditions within studies. For each condition, the number of participants is listed in italics under the short name for that condition. Consider first the rows labeled 11 and 13. These are choices between four-branch gambles in the “split” form. The incidence of violations of stochastic dominance is never more than 20% when the choice is presented in these split forms. Asterisks on these 28 entries signify that all 28 percentages are significantly less than 50% by individual binomial tests. [In this paper, the term “significant” and asterisks are used throughout to describe results of conventional two-tailed tests with .05 level of significance]. Averaged over the new conditions only, there were 12.2% violations of stochastic dominance in these split forms.

In sharp contrast, however, consider the rows representing Choices 5 and 7. These are simply the coalesced forms of Choices 11 and 13, respectively. The incidence of violation of stochastic dominance in these rows is always greater than or equal to 46%. In 29 of the 32 cases, the percentage significantly exceeds 50% by individual tests. Averaged over the 12 new conditions, there were 64.3% violations of stochastic dominance in the coalesced form.

To test the hypothesis that the percentage of violations of stochastic dominance exceeds 50% averaged over two

choices in each condition, one can compare the number of people who had two violations against the number who had zero violations (the rest have exactly 50% violations). By this binomial test of correlated proportions, there are significantly more than 50% violations in 13 of the 14 conditions tested separately, including all 12 of the new conditions of Studies 1–4. (This statistical test refutes stochastic dominance under any of the error models (“error stories”) described in Appendix C).

The standard that if *A* dominates *B* by first-order stochastic dominance, then the majority should not choose *B* is a conservative standard; however, it can be rejected at least once in every one of the 14 conditions tested separately.

The binomial test of correlated proportions can also be used to test for event-splitting effects (violations of coalescing). This test compares the number of people who violated stochastic dominance in the coalesced form *and* satisfied it in the split form against the number who reversed preferences in the opposite direction. These tests of event-splitting are statistically significant in all 28 comparisons.

The incidence of violations of stochastic dominance appears lowest (among the new conditions) in the semi-split format, where on average 53% of the choices violated stochastic dominance in the semi-split coalesced form. This figure still significantly exceeds 50%, but it is significantly lower than the rate of violations in other new conditions studied here, where rates vary from 60 to 83%. This reduced magnitude agrees with predictions of the prior TAX model.

Study 3 allows a comparison of the effects of event framing and event splitting. The columns of Table 1 for this study show the results for framed (bold font) and unframed (regular font) choices. Averaging between- and within-subjects, Table 2 summarizes the effects of framing and splitting. It is rare to see such a clear result in empirical research: The difference between coalesced and appropriately split events is great whereas the difference between having events framed or not framed is trivial.

In Study 4, violations of stochastic dominance were greater in the decumulative probability condition (83% violations averaged over Choices 5 and 7) than in its control condition (62% violations) with reversed text (both listed branches in decreasing order of conse-

Table 2
Effects of event framing versus event splitting (Study 3, *n* = 157)

	Coalesced	Split
Framed	74.6	14.6
Unframed	74.1	16.6

Each entry is the percentage of violations of stochastic dominance, averaged over objectively identical choices that are presented in either coalesced or split format and either framed by the same “events” (marble colors) or different “events” (different marble colors).

Table 3
Tests of upper cumulative independence and lower cumulative independence

No.	Type	First gamble	Second gamble	Birnbaum 99b		Study 1				Study 2				Study 3		Study 4	
				A 1224	Lab 124	Text 172	Rev 169	Pies 353	Pies* 305	Tick 342	List 377	Small 267	Semi 693	FU 74	UF 83	Cum 445	Cont 160
10	$S' \circ R'$	10(2) of \$40 10(2) of \$44 80(16) of \$110	10(2) of \$10 10(2) of \$98 80(16) of \$110	73*	72*	78*	70*	68*	70*	74*	68*	71*	69*	62*	73*	60*	68*
9	$S''' \circ R'''$	20(4) of \$40 80(16) of \$98	10(2) of \$10 90(18) of \$98	47	34	47	46	38	42	33	34	30	40	31	33	28	42
12	$R' \circ S'$	05(1) of \$12 05(1) of \$96 90(18) of \$106	05(1) of \$48 05(1) of \$52 90(18) of \$106	50	49	47	40	55	57	46	47	45	43	51	44	47	52
14	$R''' \circ S'''$	05(1) of \$12 95(19) of \$96	10(2) of \$48 90(18) of \$96	74*	81*	69*	77*	72*	68*	78*	66*	68*	66*	72*	72*	79*	67*
6	$S \circ R$	80(16) of \$2 10(2) of \$40 10(2) of \$44	80(16) of \$2 10(2) of \$10 10(2) of \$98	69	58	62	65	65	64	63	63*	64	62	62	53	51	65
8	$S'' \circ R''$	80(16) of \$10 20(4) of \$44	90(18) of \$10 10(2) of \$98	75*	69*	70	78*	78*	76*	70*	48	66	68*	71	65	64*	66
17	$R \circ S$	90(18) of \$3 05(1) of \$12 05(1) of \$96	90(18) of \$3 05(1) of \$48 05(1) of \$52	49*	61*	54*	46*	55*	62*	54*	44	40	42	57*	54*	63*	51*
20	$R'' \circ S''$	95(19) of \$12 05(1) of \$96	90(18) of \$12 10(2) of \$52	28	31	28	35	29	26	31	41	39	39	36	26	50	40

Note. Each entry is the percentage of participants in each condition who chose the second gamble, shown on the right. Asterisks indicate percentages significantly violating upper or lower cumulative independence by test of correlated proportions. Bold entry shows only case where satisfactions significantly exceeded violations.

quences). The differences between the decumulative condition and its control are statistically significant for both Choices 5 and 7, $\chi^2(1) = 24.4$, and, $\chi^2(1) = 28.2$, respectively, but these significant effects are opposite the direction anticipated from the notion that the decumulative probability format would help people satisfy stochastic dominance or CPT.

A model (see Appendix C) that allows people to make “errors” (responses that differ from their “true,” or latent preferences), was applied to the data to estimate the proportion of participants who “truly” violated stochastic dominance and the proportion who did so only by random “error.” Estimates from these models indicate that the theoretical percentages who “truly” violated stochastic dominance on both Choices 5 and 7 are even greater than the observed percentages who actually violated it twice.

For example, in the decumulative probability condition of Study 4, there were 441 participants who completed both Choices 5 and 7, of whom 314 (71%) violated stochastic dominance on both choices. When the true and error model is fit to these data, the estimate is that people make “errors” 12% of the time on these choices, and that 92% in this condition would have violated stochastic dominance twice, apart from errors. For comparison, in the control condition of Study 4 (with reversed text) there were 70 of 158 participants who violated dominance on both choices (44%). Here, the model fit indicates that the error rate is 23% and that 72% of participants “truly” violated stochastic dominance twice.

Violations of upper and lower cumulative independence

Table 3 summarizes tests of cumulative independence. The column labels representing the different studies and conditions are organized as in Table 1. Entries in the lower right portion of Table 3, however, show the percentage of people in each condition that chose the *second* gamble in each case.

According to upper cumulative independence, the choice percentages in Choice 9 should exceed those in Choice 10 (first two rows in Table 3). If a person prefers the *second* gamble in Choice 10 (R'), then he or she should prefer the *second* gamble in Choice 9 (R'''). It would be consistent with upper cumulative independence to switch from S' to R''' , but not from R' to S''' . However, in all 14 conditions, the percentage is reduced from significantly greater than 50% in Choice 10 to values below 50% in Choice 9 (and significantly less than 50% in most cases). If the property were as likely to be violated as satisfied, one would expect at least half of the 14 comparisons to be consistent with the property; instead, there are 14 out of 14 cases with more switches that are violations than switches that are compatible with the property. The TAX model with its prior parameters predicts that people will violate upper cumu-

lative independence by choosing R' over S' in Choice 10 and S''' over R''' in Choice 9, so it correctly predicts the majority choices in the first two rows of Table 3.

The test of correlated proportions, which compares the number who switch in violation of the property against the number who reverse in the direction compatible with the property, is significant in all 14 tests. Asterisks in Table 3 denote significant results by this test. As shown in Appendix C, this is the appropriate statistical test under different models of error. The numbers of individuals showing each choice pattern are presented in that appendix.

Similarly, Choices 12 and 14 also test upper cumulative independence, except that positions (first and second gambles) are reversed. According to upper cumulative independence, if a person prefers the first gamble in Choice 12 (R') then that person should prefer R''' in Choice 14; therefore, the percentage who choose the second gamble should be lower in Choice 14 than in Choice 12. Instead, all 14 choice percentages shift in the opposite direction. Again, the test of correlated proportions shows significant violations in all 14 cases, tested separately.

Lower cumulative independence implies that if one prefers S in Choice 6, then one should prefer S'' in Choice 8. Therefore the percentage choosing the second gamble (R or R'') should have decreased from Choice 6 to 8. Instead, these percentages increased in 13 of 14 tests, of which 8 individual tests were significant, in violation of the property. One significant result, in the list format, was consistent with the property.

A second test of lower cumulative independence (with order of gambles reversed) is provided in Choices 17 and 20. Here, S and S'' are the second gambles, so the percentage choosing the second gamble should be greater in Choice 20 than in Choice 17. Data shift in the opposite direction in all 14 tests, of which 11 were significant, tested separately.

In summary, choice percentages in Table 3 shift in the opposite direction from that predicted by the class of RDU/RSDU/CPT models in 55 of 56 tests of lower and upper cumulative independence. Clearly, one should not bet on the predictions of models that imply lower or upper cumulative independence.

Violations of branch independence

The choices in Table 3 can also be analyzed with respect to violations of branch independence. According to the prior CPT model with its inverse-S weighting function, people should switch from preferring R over S to preferring S' over R' . According to prior TAX, people should switch in the opposite direction. In Choices 6 and 10, 984 people switched from S to R' , against only 649 who switched from R to S' , $z = 8.29$. Similarly, between Choices 17 and 12, 916 people switched from S to R' and

only 782 made the opposite switch in preference, $z = 3.25$. The tests of branch independence in this paper are not optimal, since the detection of violations depends on either knowing parameters for each participant in advance or using a fishnet design with many graded comparisons to catch people who may have slightly different parameter values (Birnbaum & Chavez, 1997). The data here suggest that these participants are less risk averse than expected from prior TAX; for example, those parameters predicted that the majority should have chosen *S* over *R* in Choice 6. Nevertheless, violations of branch independence are significant and more switches are opposite those predicted by the published model of CPT than are consistent with it. Additional information on tests of branch independence is included in Appendix C.

Discussion

These studies tested whether violations of RDU/RSDU/CPT models would be eliminated or at least strongly attenuated by changes in the format for displaying probability. No format was found in which these models provided acceptable descriptions of these new paradoxes. Instead, significant violations of stochastic dominance, coalescing, lower and upper cumulative independence were observed in all of the conditions. All four of these systematic violations are consistent with (and predicted by) the RAM and TAX models with prior parameters.

Effects of display formats and framing

Although all conditions refute the class of RDU/RSDU/CPT theories, rates of violation do change a bit from condition to condition. Conditions with natural frequencies, represented by the number of tickets (or of marbles) for each prize, had slightly higher rates of violations of stochastic dominance than did those with text presentation of numerical probabilities. However, rates with frequency displays are not much higher than those reported in previous research (Birnbaum, 2000; Birnbaum & Navarrete, 1998; Birnbaum et al., 1999).

Recall that the mathematics required to “see” dominance is addition—a computation that was theorized to be facilitated by use of natural frequencies, lists, or graphical displays. There appears no apparent advantage of presenting probabilities as natural frequencies for reducing violations of stochastic dominance. Nor does the pie chart graphical format reduce violations, despite the argument that a person should be able to visually and graphically add events to “see” dominance.

Even listing each of 20 equally likely prizes does not eliminate violations of stochastic dominance when those tickets are grouped with a separate line showing the list for each value. If a person had simply cancelled equal-

valued tickets from both choices, and chosen the gamble with the better remainder, she or he would have satisfied stochastic dominance without having to do any arithmetic. The list format could be viewed (objectively) as a completely split display, where each ticket constitutes a branch, so the manipulation of placing each branch on a separate line may seem subtle. As in the other conditions, when new lines are introduced so that the two gambles in a choice have branches with equal numbers, the majority satisfies stochastic dominance.

The semi-split list reduced the proportion of violations of stochastic dominance to “only” 53%; the lowest rate among the new variations tested here. However, this figure still significantly exceeds 50% and is still much higher than rates observed in any of the formats when branches are fully split with the splitting designed to make comparable branches of equal probability. Apparently each line, rather than each ticket, forms a branch in this format. With this definition of branch, the prior TAX model correctly predicted that the magnitude of violation of stochastic dominance should be reduced but not reversed in semi-split format.

The decumulative display is one that was conjectured to promote satisfaction of stochastic dominance and CPT. Surprisingly, the decumulative format yielded the highest rates of violation of stochastic dominance.

A study by Harless (1992) is sometimes cited as evidence that display format affects preferences. However, that study unintentionally confounded format with event-splitting. Harless compared a matrix format for presentation of choices with a text (“tickets”) format that described how tickets would be converted to prizes. In the matrix format, the juxtaposition of branches was confounded with event-splitting: when branches were juxtaposed, the larger prize of one choice was also split, and when they were not juxtaposed, the two branches yielding the same prize were coalesced (Harless, Fig. 1). However, in the text (“tickets”) format used by Harless (1992, Fig. 3), branches were always coalesced, whether juxtaposed or not. Because the juxtaposition effect was theorized to be a “regret” effect, Harless reached the conclusion that “regret” effects depend on problem representation. However, the results in Harless (1992) can be attributed to event-splitting effects, rather than to either format or juxtaposition.

Such an interpretation of Harless (1992) fits with that of Starmer and Sugden (1993), who noted that event-splitting may have been the variable that produced the apparent juxtaposition (“regret”) effects in earlier studies. Luce (2000) expressed reservations concerning the kinds of tests run by Starmer and Sugden (1993), however. Terry Connolly (personal communication, March, 2004) tested some of the present choices using matrix format. Consistent with the present conclusions, he found no violations of stochastic dominance in a small pilot study when branches were appropriately split.

Research is presently under way to assess the effect of branch splitting (versus coalescing) in the matrix format, to test if it was the splitting or some other feature of matrix format that reduced violations in Connolly's test.

There is little solace in the present results for those who theorized that CPT might work in some restricted domain of presentation formats. The main problem for CPT is that it must satisfy the property of coalescing, so it cannot explain splitting effects, which were very powerful in all of the conditions tested here, nor can it explain paradoxes that can be deduced as implications of coalescing. Violations of CPT are observed whether probability is presented as text or accompanied by pie charts, whether by lists or by frequencies, whether by marbles or by tickets, and using either branch or decumulative probabilities. Reversing the order of branches with respect to the consequences appears to have minimal effect as did the event framing manipulation of using the same or different colors of marbles on corresponding branches. Majority violations of stochastic dominance have also been observed whether branches are juxtaposed or not, when participants are tested in the lab, in the classroom, or via the Web (Birnbaum & Martin, 2003).

In this study, there were modest financial incentives. The incentive to participate in these studies had an expected value approximately equal to that of a California State lottery ticket that costs \$1. Apparently, this incentive, perhaps combined with intrinsic interest or desire to help, was sufficient to induce thousands of volunteers to participate in the study. Why might these people systematically choose dominated gambles? Violations of stochastic dominance do not appear to depend on mere presence of financial incentives. Birnbaum and Martin (2003) found that rates of violation did not differ much between undergraduates who had zero financial incentives and those who had a 1 in 15 chance of winning a gamble with prizes as high as \$220 and an EV of about \$50. Previous studies of CPT used no financial incentives (e.g., Tversky & Kahneman, 1992), and despite incentives to find such effects, there is remarkably little evidence that cash prizes influence decision strategy (Camerer & Hogarth, 1999).

So, there is an answer to the practical questions: to minimize violations of stochastic dominance, present gambles in an appropriately split format with equal probabilities on corresponding branches. To create violations, follow Birnbaum's recipe: split lower branches of superior gambles (to make them seem worse) and split the upper branches of inferior gambles (to make them seem better).

Violations of RDU/RSDU/CPT: Coalescing

The identification of coalescing/splitting as the culprit provides a theoretical answer as well as a practical one.

The assumption made both in original prospect theory and cumulative prospect theory that most people spontaneously perform coalescing (Kahneman, 2003, p. 727) is apparently not descriptive of empirical choices. It is possible that some people on some occasions might do this, of course, but as a descriptive scientific theory, this core idea of both original and cumulative prospect theories can be rejected.

Event-splitting effects, which can be interpreted as violations of coalescing combined with other premises, appear to be at the root not only of violations of the four paradoxes studied here, but also of others. Wu (1994) tested a property of upper tail independence that incorporated the assumption of coalescing, as did Birnbaum (2001b), who tested a variant of Wu's property. Violations of this property observed by Wu and by Birnbaum can be interpreted as violations of coalescing, as can the classic "common consequence" paradoxes of Allais and variants on those paradoxes (Birnbaum, 2004). Violations of stochastic dominance and cumulative independence, interpreted here as violations of coalescing, occur not only in choice (as in the present studies), but also in judgment (Appendix D), where gambles are presented one at a time.

One of the points in favor of rank-dependent models (Quiggin, 1993) over models such as "stripped prospect theory" (Starmer & Sugden, 1993) was that the rank-dependent models do not violate transparent dominance. The extended version of "stripped" prospect theory [$V(G) = \sum_{i=1}^n \pi(p_i)u(x_i)$, without editing rules] can violate dominance even when the worst consequence in the dominant gamble is better than the best consequence in the dominated gamble (Birnbaum, 1999a; Fishburn, 1978). Because few people would violate dominance in such cases, behavioral scientists resisted theories that implied such violations.

It is important to note, however, that RAM and TAX models, which can violate stochastic dominance, do not violate it in such cases. They only violate it in special circumstances such as in the special recipe used in this paper. Birnbaum (2004) sampled "random" choices between "random" three-branch gambles (with uniformly distributed probabilities and uniformly distributed consequences). Stochastic dominance relations are found in 1/3 of random choices among gambles constructed that way; however, in fewer than 2 cases per 10,000 did TAX with its prior parameters imply a violation of stochastic dominance. So, these models produce violations only rarely in such an environment.

Concluding comments

In summary, the observed effects of format and event framing were minimal. However, the effect of splitting versus coalescing of branches had large effects.

It would be interesting to explore methods for training people to detect stochastic dominance, either by giving explicit rules for proper comparison of decumulative probabilities or by instruction in methods for finding appropriate splits that reveal dominance. If such education were effective, it would be interesting to study what effect the training might have on choices between nondominated alternatives. For example, would such training transfer to tests of lower and upper cumulative independence?

When EU was considered descriptive, its utility function explained risk aversion, sales and purchase of gambles, and sales and purchase of insurance. However, EU could not explain the Allais paradoxes and it gave an unsatisfactory account of risk seeking and risk aversion by the same person. Prospect theory and CPT provided better ways to handle these phenomena. The RDU/RSDU/CPT models were more complicated than EU, since they used a weighting function to account for Allais paradoxes and other violations of EU (Tversky & Kahneman, 1992; Tversky & Wakker, 1995; Wu & Gonzalez, 1996, 1998).

If we take the present results (establishing the generality of the four new paradoxes) seriously, it means that the class of RDU/RSDU/CPT models can no longer be regarded as descriptive models of choice. Violations of lower and upper cumulative independence can be viewed as direct contradictions in the weighting function (Birnbaum et al., 1999, pp. 79–80), under any version of RDU, RSDU, or CPT. This suggests that the weighting functions of RDU/RSDU/CPT are merely artifacts of wrong theory and do not represent decumulative weighting at all. These paradoxes help explain why violations of branch independence and distribution independence, which should have been predictable from the CPT weighting function, yielded results that were opposite those predicted by the fitted model of CPT (Birnbaum & Chavez, 1997; Birnbaum & Navarrete, 1998).

The TAX model resolves the apparent contradictions. This point deserves emphasis: the apparent contradictions between studies are not necessarily contradictions in results. The contradictions occur when one assumes coalescing, as in RDU/RSDU/CPT. Once we replace cumulative weighting with branch ranked weighting (as in RAM or TAX) these paradoxes can be resolved. Violations of stochastic dominance, event-splitting effects, lower and upper cumulative independence can all be explained by models that violate coalescing.

Appendix A. Four “new paradoxes”

Basic definitions

Let $G = (x, p; y, q; z, r)$ represent a gamble to win x with probability p , y with probability q , and z with

probability $r = 1 - p - q$. Let \succ represent the preference relation, where $G \succ F$ means that G is preferred to F . Let $G \sim F$ denote that G is indifferent to F .

A *branch* of a gamble is a probability-consequence pair that is distinct in the gamble’s presentation to the participant.

Transitivity is the premise that if $A \succ B$ and $B \succ C$ then $A \succ C$.

Consequence monotonicity is the assumption that if one consequence of a gamble is improved, holding everything else in the gamble the same, the gamble with the better consequence should be preferred. For example, supposing $x \succ y$, then $F = (x, p; z, 1 - p) \succ G(y, p; z, 1 - p)$. Systematic violations of this property have been reported in judgment, but not in choice (Birnbaum & Sutton, 1992).

Coalescing is the assumption that if a gamble has two branches that yield the same consequence, these branches can be combined by adding their probabilities. For example,

$$A = (x, p; x, q; z, r) \sim A' = (x, p + q; z, r).$$

The class of RDU/RSDU/CPT theories satisfies coalescing (Birnbaum & Navarrete, 1998), as do a number of other theories (for a list, see Luce, 1998), but the configural weight models of Birnbaum and his associates do not (Birnbaum, 1997; Birnbaum & Navarrete, 1998). Marley and Luce (2001) have recently proposed Gains Decomposition Utility, which can also violate coalescing.

New paradoxes

1. *Event-splitting effects* (branch splitting) are systematic violations of coalescing combined with transitivity (Starmer & Sugden, 1993). From coalescing and transitivity it follows that

$$\begin{aligned} A = (x, p; x, q; z, 1 - p - q) \succ B = (w, 1 - r - s; y, r; y, s) \\ \iff A' = (x, p + q; z, 1 - p - q) \succ B' = (w, 1 - r - s; y, r + s) \end{aligned}$$

because $A' \sim A \succ B \sim B'$. When $A \succ B$ and $A' \prec B'$, the finding is termed an “event-splitting effect.” Note that A' is the same as A and B is the same as B' , except for coalescing, so if we assume transitivity, event-splitting effects are violations of coalescing.

2. *Stochastic dominance* (also known as First-Order Stochastic Dominance) is the relation between two nonidentical gambles, F and G , such that if the probability of winning x or more in gamble F exceeds or equals that probability in G for all x , one says that F stochastically dominates G .

The statement that *preferences satisfy stochastic dominance* means that if F stochastically dominates G , then people should not systematically prefer G to F . First-order stochastic dominance, as tested in this study, should not be confused with other special types of “dominance,” for example such as in Levy and Levy

(2002). Although violations of certain special “dominance” relations are actually consistent with the model of CPT (Wakker, 2003), first-order stochastic dominance *must* be satisfied by any member of the class of RDU/RSDU/CPT models. These models are equivalent for gambles consisting strictly of gains, so they are the same in these studies. CPT is sometimes said to include some additional restrictions on functions and parameters. This class of models satisfies coalescing and properties that follow from it with any functions and parameters (Birnbaum & Navarrete, 1998).

Birnbaum’s (1997) *recipe for violations of stochastic dominance*. Start with a “root” gamble, $G_0 = (x, p; y, 1 - p)$, where $x > y$. Next, split the lower branch, $(y, 1 - p)$, into two splinters and increase the consequence on one splinter $(y, 1 - p - r; y^+, r)$; where $y^+ > y$. This creates a new gamble, $I = (x, p; y^+, r; y, 1 - p - r)$, which dominates G_0 . Now create gamble $J = (x, p - q; x^-, q; y, 1 - p)$ by splitting the upper branch of G_0 and replacing consequence x with an inferior consequence, x^- . Therefore, I dominates G_0 which dominates J . If choices satisfy consequence monotonicity, coalescing, and transitivity, then they would not violate stochastic dominance in this recipe.

Birnbaum (1997) devised two paradoxes to reconcile what seemed to be contradictions between data of Birnbaum and McIntosh (1996), Tversky and Kahneman (1992), and Wu and Gonzalez (1996). These paradoxes were designed to test if the apparent contradictions in results could result in direct contradictions with CPT within an experiment.

3. *Lower cumulative independence*, with $0 < z < x' < x < y < y' < z'$, can be written:

$$S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q) \Rightarrow S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q).$$

4. *Upper cumulative independence* can be written:

$$S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r) \Rightarrow S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r).$$

Any theory that satisfies restricted comonotonic branch independence, monotonicity, transitivity, and coalescing must satisfy both lower and upper cumulative independence conditions (Birnbaum & Navarrete, 1998).

These studies also include tests of *restricted branch independence*, which can be written:

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

The term “restricted” is used to indicate that the number of branches and probability distribution is fixed (p, q, r) in all gambles. The term, restricted *comonotonic branch independence* refers to the special case in which ranks of corresponding consequences are the same in all four gambles. RAM, TAX, RDU, RSDU, and CPT all sat-

isfy comonotonic branch independence, but can violate restricted branch independence. With their prior parameters, however, they predict different patterns of violation of branch independence.

If people cancelled common components as postulated in original prospect theory (Kahneman & Tversky, 1979), there would be no violations of restricted branch independence. Note that in the first comparison, the branch (z, r) is the same in both S and R , and could thus be cancelled; in the second comparison, the branch (z', r) is identical in both S' and R' . If people cancelled equal valued tickets in the list format, they would not violate stochastic dominance.

Appendix B. RDU and TAX models

Birnbaum (1974) noted that in his “configural weight” models, the weight of a stimulus “depends in part on its rank within the set.” Birnbaum’s models are similar to “rank-dependent” models that were later introduced to decision making (Quiggin, 1985). However, the manner in which weights are affected by rank in Birnbaum’s RAM and TAX models differs from that in RDU/RSDU/CPT models. The “new paradoxes” provide tests between these two classes of models.

RDU models

When the consequences are strictly positive, as in the present studies, the RDU/RSDU/CPT models reduce to RDU. The RDU of a gamble, $G = (x_1, p_1; x_2, p_2; \dots; x_i, p_i; \dots; x_n, p_n)$, where $x_1 > x_2 > x_3 > \dots > x_n > 0$; $\sum p_i = 1$, is given by the following:

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

where P_i is the (decumulative) probability of receiving x_i or more, and P_{i-1} is the probability of receiving strictly more than x_i . The function, $u(x)$, is the utility function of the monetary outcome, x . $W(P)$ is a strictly increasing, decumulative weighting function, which assigns $W(0) = 0$ and $W(1) = 1$.

The model fit by Tversky and Kahneman (1992) further specified $W(P)$ to be an inverse-S function of P , with steeper slopes near zero and one than near 1/2. Tversky and Wakker (1995) used the term “S-shaped” for what I call “inverse-S”, and used the following equation for $W(P)$:

$$W(P) = \frac{cP^\gamma}{cP^\gamma + (1 - P)^\gamma}. \tag{B.2}$$

The “prior” CPT model will refer to Eqs. B.1 and B.2, using parameter estimates fit to the data of Tversky and Kahneman (1992), where $c = .724$; $\gamma = .61$; and $u(x) = x^{.88}$ for $x \geq 0$.

RAM and TAX models

Birnbaum and Chavez (1997) described two configural weight models, now called the rank-affected multiplicative weights model (RAM) and transfer of attention exchange (TAX) models. Both models postulate that the value of a gamble is a weighted average of the utilities of the consequences on the branches, where the weight of each branch depends on its probability and the rank of the branch's consequence (Birnbaum & Stegner, 1979). This common ground can be described as general weighted utility models, which Marley and Luce (2001) have shown satisfy the assumptions of conjoint measurement for fixed probability distributions and a commutative property not tested here. Although special forms of RAM and TAX can be tested against each other (Birnbaum & Chavez, 1997), they make the same predictions in these studies, so only the more successful of the models, special TAX, will be described below.

Consider gamble, $G = (x_1, p_1; x_2, p_2; \dots; x_j, p_j; \dots; x_i, p_i; \dots; x_n, p_n)$, with n distinct branches, where the consequences are ordered such that $x_1 \geq x_2 \geq \dots \geq x_i \geq \dots \geq x_n$, and $\sum p_i = 1$. A "special case" of the TAX model (Birnbaum & Chavez, 1997) can be written as follows:

$$\text{TAX}(G) = \frac{\sum_{i=1}^n u(x_i) \left[t(p_i) - \frac{\delta}{(n+1)} \sum_{j=i+1}^n t(p_j) + \frac{\delta}{(n+1)} \sum_{j=1}^{i-1} t(p_j) \right]}{\sum_{i=1}^n t(p_i)}, \quad (\text{B.3})$$

where $\text{TAX}(G)$ is the utility of the gamble; $t(p)$ is a function of probability; $u(x)$ is the utility function of money, and δ is the single configural parameter; $\delta/(n+1)$ is the proportion of weight taken from a branch with a higher consequence and transferred to each branch with a lower consequence. Note that the weight transferred is proportional to the (transformed) probability of the branch giving up weight.

The "prior" TAX model refers to "special TAX" (Eq. B.3), plus the assumptions that $u(x) = x$, $0 < x < \$150$, $t(p) = p^7$, and $\delta = 1$. These parameters were used by Birnbaum (1999a) to describe a variety of phenomena for choices between gambles. Although this model is the equivalent to that in Birnbaum and Chavez (1997), the ordering of the branches is written here to conform to that in CPT, and $\delta = 1$ in this equation corresponds to $\delta = -1$ in Birnbaum and Chavez (1997).

To explore predictions of CPT, TAX, and EV, as well as other models, one can use on-line calculators available from the following URL: <http://psych.fullerton.edu/mbirnbaum/calculators/>.

Appendix C. Models of error and the new paradoxes

When the same choice is presented to different people, not all people make the same decision. When the same choice is presented again to the same person, that person does not always make the same decision. However, algebraic models of utility dictate deterministic preferences, so one must establish some basis for deciding whether a certain incidence of violations should be considered evidence against the theory or attributed instead to such factors as individual differences or unreliability within individuals. The issue of distinguishing systematic from "random" violations, between and within individuals, has received some attention (Birnbaum & Chavez, 1997; Carbone & Hey, 2000; Loomes & Sugden, 1995; Luce, 2000).

In psychophysics, choice models stemming from the ideas of Fechner, Thurstone, and Luce have had success in describing how a person might judge the lighter of two weights to be "heavier" on some proportion of the presentations. What has not yet been achieved is a satisfactory formal representation connecting algebraic structure, individual differences, and probabilistic behavior in a single theory (Iverson & Luce, 1998). Nevertheless, four ideas for analyzing deterministic models with fallible data have been suggested.

Operational definition of preference relation

The first idea is to provide an *operational definition of the preference relation* in terms of a statistical inference. The statement $A \succ B$ is defined to be equivalent to the statement $P(A, B) > 1/2$, where $P(A, B)$ is the probability that A is preferred to B .

To show that $A \succ B$ by this definition, it is necessary to demonstrate that we can reject the hypothesis that A and B are indifferent. If the observed choice percentage exceeds 50% but is not significantly different from 50%, we cannot say that $A \succ B$, nor can we conclude that $A \sim B$, and we must retain even the hypothesis that $A \prec B$. This standard is conservative, and its application forces the investigator in many cases to stand undecided. CPT implies that if A stochastically dominates B , then $A \succ B$. By this definition, it means that $P(A, B) > 1/2$. Therefore, the finding that the dominated choice is chosen significantly more than 50% of the time [rejecting the hypothesis that $P(A, B) \geq 1/2$], means that if we accept this operational definition, then we must reject this implication of CPT.

Choice models

The second idea is to join a choice model, such as Thurstone's (1927), to the utility model. One assumes that the evaluation of each gamble follows the algebraic model of utility, except for some random process that

can be represented by additive error. For example, if RDU is our null hypothesis; we state the model as follows:

$$U(A) = \text{RDU}(A) + \varepsilon_A,$$

$$U(B) = \text{RDU}(B) + \varepsilon_B,$$

where $\text{RDU}(A)$ is the evaluation of A under the RDU representation, and ε is an “error” that causes the momentary utility of gamble A to vary from occasion to occasion. If we assume that errors are normally distributed with zero means, it follows that the probability of choosing A over B is given by the following:

$$P(A, B) = F \left[\frac{U(A) - U(B)}{\sigma_{AB}} \right], \quad (\text{C.1})$$

where F is the cumulative normal distribution function and σ_{AB} is the standard deviation of differences; this standard deviation will depend on the variances of the errors and their covariance.

A special case of this approach in which the standard deviations of differences were all assumed to be equal (and which used the logistic instead of normal for F) was applied by Birnbaum and Chavez (1997), Birnbaum and Navarrete (1998), and Birnbaum (1999b).

According to this model, the standard statistical test of correlated proportions used in the literature is not always appropriate for testing the assumption that $A \succ B \iff C \succ D$, as for example, in the Allais constant ratio paradox (Loomes & Sugden, 1995). Let $A = \$3000$ for sure, $B = (\$4000, .8; \$0, .2)$, $C = (\$3000, .25; \$0, .75)$ and $D = (\$4000, .2; \$0, .8)$. According to EU, people should prefer A to B if and only if they prefer C to D ; however, the number of people who switch from A to C is not equal to the number who switch in the other direction. Is this evidence against EU?

Assuming that the standard deviations were equal, it follows that this statistical test would be justified if and only if the null hypothesis implies that $U(A) - U(B) = U(C) - U(D)$. However, this assumption of equal differences does not follow from EU. For example, suppose that $u(x) = \log(x)$ and EU holds. If so, then $\text{EU}(A) = \log(\$3000) = 3.48$, $\text{EU}(B) = .8 \log(4000) = 2.88$, $\text{EU}(C) = 0.87$, $\text{EU}(D) = 0.72$. With these calculations, EU implies $U(A) - U(B) > U(C) - U(D)$, so it should be the case that more people switch from A to D than vice versa. In other words, the statistical test of correlated proportions is not the appropriate test of EU under this error theory for this property.

A more convincing test in this case is to show that $P(A, B) > 1/2$ and $P(C, D) < 1/2$, both relations of which have been shown to be statistically significant for the “constant ratio” Allais paradox (e.g., Birnbaum, 2001a, 2001b, p. 36), so despite the problem with the statistical test used in some published papers, it appears that the constant ratio paradox is “real.”

This choice model, with or without the assumption of equal standard deviations, justifies the operational definition of $A \succ B$ in the first approach. In addition, if we assume that standard deviations are all equal, then for tests in which the null hypothesis implies equality of utility differences, then the standard statistical test of correlated proportions is justified. When RDU/RSDU/CPT is the null hypothesis, each of the four paradoxes in this paper can be rewritten in terms of utility differences, which justifies the statistical tests used here. In the case of tests of coalescing, the utility differences are equal under the null hypothesis since splitting or coalescing should have no effect on utility. In the other three paradoxes, the observed choice percentages change in the opposite direction of the theoretical utility intervals.

The assumption that the standard deviations of errors are all equal is an oversimplification, however, since scalability can be violated in choices between risky gambles (e.g., Diederich & Busemeyer, 1999). So a skeptic might hold on to the idea that changes in response proportions that do not significantly reverse the majority might be due to differences in the standard deviations, rather than to changes in utility intervals.

Random walk models of the comparison process (Birnbaum & Jou, 1990; Busemeyer & Townsend, 1993; Link, 1992) are models that provide testable implications for response time as well as choice “errors,” providing a process model that can be viewed as the basis for the difference representation of choice proportions. These models can be used as scaling models where decision time provides another theoretical constraint on subjective difference (utility intervals). As shown by Birnbaum and Jou (1990), these random walk models of response time yielded results compatible with configural weight models like RAM and TAX as well as direct judgments of “differences.”

Random utility

A third approach is to assume that variability occurs only in the parameters of the model. This approach is sometimes called the “random utility” model. Assuming that CPT parameters change from choice to choice, it follows that a person should satisfy stochastic dominance 100% of the time (Carbone, 1997; Loomes & Sugden, 1995). This follows because under any parameters, CPT always satisfies stochastic dominance. As Carbone (1997, p. 307) put it, “whenever a pairwise choice is presented to the subject in which one of the pair stochastically dominates the other then the subjects always choose the dominating one. . . it seems to capture well the experimental evidence.” This conclusion is not consistent with the majority violations of stochastic dominance in these studies. There is no contradiction in data, however, since Carbone’s review did not include tests of stochastic dominance constructed

from Birnbaum’s recipe, but rather from a recipe in which RAM and TAX also satisfy dominance.

So, the random utility approach does not allow any violations of stochastic dominance under CPT. However, if we assumed instead that parameters changed from gamble to gamble, rather than from choice to choice, then this approach would imply a more complex version of Thurstone’s model. In this case “errors” would be calculated (or simulated) from the assumed distribution of parameters, correlations among those parameters between gambles within choices, and the numerical structure of the model. This more complex version has not yet been developed, save for situations where these sources of variability are assumed to produce symmetrically distributed errors, leading to Eq. (C.1).

Random errors in responding

The fourth approach is to fit a model in which “errors” cause a person’s observed choice to differ from that person’s “true” preference. A version of this model has been used to argue that evidence against transitivity published by Loomes, Starmer, and Sugden (1991) may have been due to random error rather than to any real intransitivity (Sopher & Gigliotti, 1993). This model is sometimes fit with the assumption that the probability of an error on any given choice is the same for all choices (Harless & Camerer, 1994).

The model will be applied here with the more limited assumption that error rates are equal only in special pairs of choices. The analysis of two choices is illustrated in Fig. 8. This model can be used to deduce the connection between the substantive hypothesis and the statistical test. The remainder of this section analyzes the four “new” paradoxes of this paper within this approach.

Stochastic dominance: Let $A \succ B$ represent satisfaction of stochastic dominance on Choice 5 and let $C \succ D$ represent satisfaction on Choice 7. Let a = the probability that a person is described by a model (such as CPT) that satisfies dominance. Such a person should satisfy dominance on both choices. According to prior TAX, however, a person should violate it both times. Therefore, we can fix $c = 1$ and $d = 1$ in Fig. 8, so that there are only two “true” (latent) choice patterns to be considered: AC (two satisfactions) and BD (two violations).

Table 4 shows the number of participants in each study who satisfied ($G+$) or violated ($G-$) stochastic dominance on Choices 5 and 7 respectively. For example, of the 172 participants in the *Text* condition of Study 1, there were 73 who violated stochastic dominance on both choices, and 40 who satisfied it on both choices.

To fit each row of data in Table 4, there are just two parameters (a and e) to describe the four observed rel-

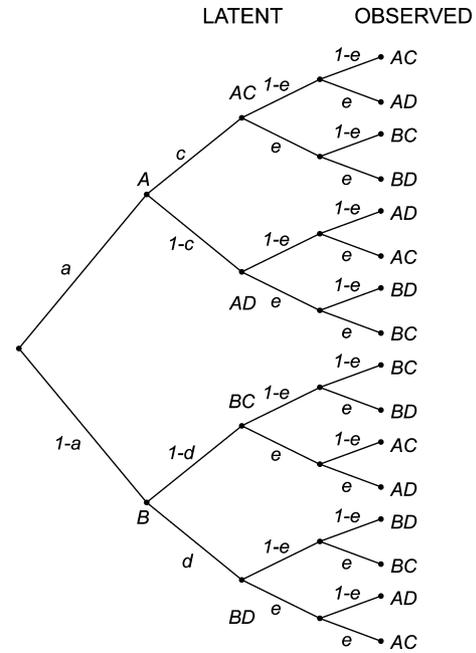


Fig. 8. A true and error model for a pair of choices. The “latent” preferences are constrained by deterministic theory, but due to random “errors,” observed responses may depart from these “true” preferences. The figure represents preference patterns for a pair of choices, between gambles A and B and between C and D . The probability that $A \succ B$ is a ; the probability that $C \succ D$, given that $A \succ B$ is c ; the probability that $D \succ C$ given $B \succ A$ is d . The probability of an error, e , is assumed to be less than $1/2$, and errors are assumed to be independent.

Table 4
Tests of stochastic dominance

Condition	Choice pattern on Choices 5 and 7			
	$G+G+$	$G+G-$	$G-G+$	$G-G-$
Text	40	24	35	73
Reversed	35	34	21	78
Pies	72	47	48	181
Pies*	65	39	44	154
Tickets	57	39	53	188
List	88	57	56	171
Small	58	35	44	128
Semi-split	188	112	150	235
Marbles FU	11	11	9	43
Marbles UF	8	12	10	53
Decumulative	34	59	34	314
Control-rev	32	33	23	70
Birnbaum 99b	384	134	272	431
Birnbaum 99b lab	40	26	49	131
Total	1112	662	848	2250

Choices 5 and 7. Each entry is the number of participants showing each pattern of preferences. $G+G-$ indicates choice of the dominant gamble on Choice 5 and the dominated gamble on Choice 7. Row totals do not equal the number of participants when one or more people skipped one or more items.

ative frequencies of response combinations, AC , AD , BC , BD (i.e., $G+G+$, $G+G-$, $G-G+$, $G-G-$). Table 5 shows the theoretical probabilities of these four choice

Table 5
Analysis of stochastic dominance in Choices 5 and 7 in the true and error model

Choice 5	Choice 7	
	C(G+)	D(G-)
A(G+)	$a(1-e)(1-e)+$ $(1-a)ee$	$a(1-e)e+$ $(1-a)e(1-e)$
B(G-)	$ae(1-e)+$ $(1-a)(1-e)e$	$ae e+$ $(1-a)(1-e)(1-e)$

Notes. This model implies that the probability of response pattern BC equals that of AD. The probability of AC (no violations) will be greater than that of BD (two violations) if and only if $a > 1/2$. When a is 1 or 0, the model implies independence between rows and columns.

Each entry is the theoretical probability of that response category under the assumption that people occasionally make “errors” that cause their overt responses to differ from their true preferences.

combinations. Within this general model, one can compare particular assumptions concerning the parameters, such as $a = 1$ (everyone satisfies stochastic dominance), $a \geq 1/2$ (half or more of the people satisfy stochastic dominance), and $a = 0$ (no one satisfies stochastic dominance, apart from error).

From Table 5, we see that probabilities of the cells, $AC(G+G+) > AD(G-G-)$ if and only if, $a(1-e)^2 + (1-a)e^2 > ae^2 + (1-a)(1-e)^2$, which can be rewritten $2a[(1-e)^2 - e^2] > [(1-e)^2 - e^2]$. Therefore, $a > 1/2$.

This analysis shows that the probability of AC is greater than that of BD if and only if $a > 1/2$. Therefore, the binomial test comparing the frequencies of AC and BD is the appropriate statistical test of the hypothesis that probability of violating stochastic dominance exceeds one-half. In addition, the model makes the testable prediction that the frequency of BC equals that of AD, for any a and e .

Table 6 shows the fit of this model (Table 5) to the data in the decumulative probability condition in Study 4. The solver in Excel was used to find a and e , to minimize the χ^2 comparing observed to fitted frequencies. The observed frequencies of each cell are shown with fitted values in parentheses. The estimated values of a and e are .08 and .12, respectively; these are interpreted to indicate that only 8% of participants “truly” satisfied stochastic dominance and that people make “errors” 12% of the time in expressing their true preferences in this study. The $\chi^2(2)$ for this model is 6.63. The model does not account for the greater frequency in cell AD than BC.

Table 6
Fit of error model to data of Choices 5 and 7 in the decumulative probability condition of Study 4

Choice 5	Choice 7	
	C(G+)	D(G-)
A(G+)	34 (33.7)	59 (47.8)
B(G-)	34 (47.8)	314 (331.7)

When this model is fit with the additional assumption that no one satisfies stochastic dominance (fixing $a = 0$), $\chi^2(1) = 39.6$ (with a predicted frequency of 281 in cell BD); whereas the assumption that at least half of the people satisfy it yielded $\chi^2(1) = 227.5$ (predicting 181.4 in cell BD). The assumption that everyone satisfies stochastic dominance ($a = 1$) yielded $\chi^2(1) = 505.8$, with equal predicted frequencies in all cells (110.25). These results (and similar analyses applied to the other data) indicate that when we fit this model of errors, fewer than half of the participants truly satisfy stochastic dominance.

In the control condition of Study 4 (reversed text), for example, the estimated values of a and e were .28 and .23, respectively, with $\chi^2(2) = 1.77$. The assumption that $a = 0$ yielded a $\chi^2(1) = 11.0$, and the assumption that $a \geq .5$ yielded $\chi^2(1) = 15.8$. The indices of fit are all smaller because the sample size is smaller, but the relative values lead to the same conclusions. Here, the estimate is that 28% satisfy dominance and 72% violate it.

This model can be extended to fit the distribution of violations of stochastic dominance in studies that have four tests of stochastic dominance in the coalesced form. In Study 3, out of 156 people who completed all four choices, there were 79 with four violations, 31 with three violations, 21 with two, 13 with one, and 12 with zero violations. Estimates of the model are $a = .17$ and $e = .15$ with $\chi^2(2) = 10.7$. This indicates that 83% of the participants in Study 3 “truly” violated dominance four times, but only 68.1 (43.7%) were predicted to violate it four times because of errors. There were also four tests in the lab sample of Birnbaum (1999b). In that study, there were 43, 35, 18, 20, and 6 people with 4, 3, 2, 1, or 0 violations, respectively. Estimated parameters are $a = .24$, and $e = 0.20$. Birnbaum and Navarrete (1998) also had four tests of this type, with 41, 24, 18, 8, and 9 people having 4, 3, 2, 1, and 0 violations, respectively. For that study, estimates are $a = .19$ and $e = .18$, with $\chi^2(2) = 4.73$. Here, the model predicted that 36.2 of 100 participants were expected to show four violations, compared to 41 who actually did have four violations. Although the average incidence of violations of stochastic dominance was 74, 68, and 70%, in these three studies, respectively, the model fits indicate that 83, 76, and 81% of people had four “true” violations, apart from errors in these studies, respectively.

This error model allows for individual differences: people are assumed to either violate the property or satisfy it. This model is quite different from a model that assumes that all people are the same within a given study, and have the same, independent probability of violating stochastic dominance. For example, assuming that everyone in Study 3 had the same probability of violating stochastic dominance (equal to the average rate of violation, 0.74), we expect only 47.7 people to have four violations, compared to 79 who actually vio-

lated stochastic dominance on all four tests in that study, $\chi^2(3) = 237.6$. The assumption that choices are independent fits significantly worse than the model that allows individual differences. $\chi^2(1) = 237.6 - 10.7 = 226.9$. This difference was also significant when the frequency of zero violations (which had a low expectation in this model) was excluded from the analysis. In sum, these analyses show that there are “true” individual differences but the significant majority of participants tested violated rather than satisfied stochastic dominance in this recipe.

Coalescing. One way to analyze coalescing is to compare two choices in which *C* and *D* are the split forms of *A* and *B*, respectively. The model of Fig. 8 then would be modified to test the assumption that $A \succ B \iff C \succ D$. Let a = the probability of choosing *A* over *B*, and set $c = 1$ and $d = 1$. In this case, derivations show that the statistical comparison of the frequencies of *AD* and *BC* (the standard test of correlated proportions) is the proper test of coalescing. Under the assumptions, the probabilities of both cells are equal to $(1 - e)e$, so statistical rejection of equality would refute these assumptions.

Table 7 shows the results of this test of coalescing for Choices 5 and 11. In every condition of each study, the number of people showing the *G–GS+* pattern (bold type) significantly exceeds the number showing the opposite reversal of preferences, *G+GS–*. Similar results were found in separate analysis of Choices 7 and 13.

A second way to analyze Choices 5 and 11, illustrated in Table 8, is to compare two classes of models. According to prior TAX or RAM models, people should violate dominance on Choice 5 and satisfy it on Choice 11. CPT (and its class of models) require satisfaction in both choices. In relation to Fig. 8, we again let $A \succ B$ represent satisfaction of stochastic dominance on

Table 7
Coalescing Choices 5 and 11

Condition	Choice pattern			
	<i>G+GS+</i>	<i>G+GS–</i>	<i>G–GS+</i>	<i>G–GS–</i>
Text	56	8	93	14
Reversed	67	3	89	10
Pies	97	22	210	21
Pies*	93	12	172	26
Tickets	95	3	224	16
List	137	8	203	26
Small	85	8	143	28
Semi-split	267	29	321	63
Marbles FU	19	3	46	6
Marbles UF	15	5	50	11
Decumulative	71	23	292	55
Control-rev	60	6	81	11
Birnbaum 99b	458	57	632	71
Birnbaum 99b lab	54	12	155	24
Total	1574	199	2711	382

Table 8
Theoretical analysis of Choices 5 and 11

Choice 5	Choice 11	
	<i>C(GS+)</i>	<i>D(GS–)</i>
<i>A(G+)</i>	$a(1 - e)(1 - e) + (1 - a)e(1 - e)$	$a(1 - e)e + (1 - a)ee$
<i>B(G–)</i>	$ae(1 - e) + (1 - a)(1 - e)(1 - e)$	$ae + (1 - a)(1 - e)e$

Choice 5 (i.e., $G+ \succ G–$), and let $C \succ D$ represent satisfaction of dominance on the split version in Choice 11. Let a = the probability of satisfying stochastic dominance on both Choices 5 and 11 (i.e., $c = 1$). The rival hypothesis (e.g., TAX) implies that people will violate dominance on Choice 5 and satisfy it on Choice 11 (i.e., $d = 0$), so the two latent choice patterns are *AC* and *BC*. To test if $a > 1/2$, here we compare the frequency in cell *BC* against that in *AC*. If $a = 1$, cells *BC* and *AD* should be equal, so the binomial test of equality of these cells is a test of CPT (and other models like it). If $a = 0$, cells *AC* and *BD* should be equal, which can also be tested by binomial applied to the columns in Table 7.

Fitting the decumulative probability condition of Study 4, 292 of the 441 (66%) who completed both items violated stochastic dominance on Choice 5 and satisfied it on Choice 11 (*BC*); there were 23 who had the opposite pattern (*AD*); 71 people satisfied dominance twice (*AC*) and 55 violated it twice (*BD*). According to the model, $a = .056$, $e = .179$, and the minimal $\chi^2 = 3.78$ indicating that if we use this model to estimate the proportion who conform to CPT or TAX, we find that less than 6% fit CPT, allowing for “errors.” Analysis of the other conditions and to Choices 7 and 13 also yielded estimates of a to be significantly less than 1/2.

Upper and lower cumulative independence. Both of these properties are of the form $A \succ B \implies C \succ D$. For upper cumulative independence, let $A \succ B$ represent $S' \prec R'$ and let $C \succ D$ represent $S''' \prec R'''$. This rule implies that only three latent preference patterns in Fig. 8 are possible: *AC*, *BC*, and *BD*. This model implies that the probability of *BC* should be greater than that of *AD*. Therefore, the test of correlated proportions (which compares *BC* against *AD*) is the appropriate test of this null hypothesis. Table 9 presents the numbers of participants in each condition who showed each of these choice patterns in Choices 10 and 9. In all 14 tests, the data clearly violate this implication; the frequencies show significant deviations in the opposite direction from that required by any CPT model.

By the same reasoning, one can show that the same test is also appropriate for lower cumulative independence, once $A \succ B$ is identified with $S \succ R$ and $C \succ D$ is identified with $S'' \succ R''$. The relevant frequencies are shown in Table 10 for Choices 6 and 8. In the table, in all cases except one (List), the frequency of violations

Table 9
Tests of upper cumulative independence

Condition	Choice pattern			
	$S'S'''$	$S'R'''$	$R'S'''$	$R'R'''$
Text	23	14	*69	66
Reversed	41	10	*51	67
Pies	80	31	*135	103
Pies*	64	28	*112	98
Tickets	71	16	*156	96
List	92	27	*154	100
Small	56	20	*126	57
Semi-split	155	59	*254	216
Marbles FU	23	5	*28	18
Marbles UF	17	5	*39	22
Decumulative	137	33	*173	95
Control-rev	30	19	*60	47
Birnbaum 99b	235	84	*408	486
Birnbaum 99b lab	57	9	*106	73
Total	1081	360	*1871	1544

Choices 10 and 9. $S' = (\$110, .8; \$44, .1; \$40, .1)$, $R' = (\$110; .8; \$98, .1; \$10, .1)$, $S''' = (\$98, .8; \$40, .2)$, $R''' = (\$98, .9; \$10, .1)$. Asterisks indicate cases where preference reversals that violate the property are significantly more frequent than reversals consistent with the property.

Table 10
Tests of lower cumulative independence

Condition	Choice pattern			
	SS''	RS''	SR''	RR''
Text	29	21	35	84
Reversed	22	15	*36	95
Pies	46	29	*74	196
Pies*	33	39	*73	155
Tickets	52	51	*74	163
List	92	*99	45	134
Small	43	48	53	120
Semi-split	135	80	*127	343
Marbles FU	10	11	17	35
Marbles UF	18	11	21	33
Decumulative	87	62	*120	174
Control-rev	26	28	29	71
Birnbaum 99b	156	147	*220	692
Birnbaum 99b lab	46	29	*57	114
Total	795	670	*981	2409

Choices 6 and 8. $S = (\$44, .1; \$40, .1; \$2, .8)$, $R = (\$98, .1; \$10, .1; \$2, .8)$, $S'' = (\$44, .2; \$10, .8)$, $R'' = (\$98; .1; \$10, .9)$. Asterisks indicate cases where preference reversals that violate the property are significantly more frequent than reversals consistent with the property.

exceeds the frequency of consistent switches. In eight of nine cases with significant differences, eight are significant violations.

Branch independence. Restricted branch independence is the property that $S \succ R \iff S' \succ R'$. It implies that cells *BC* and *AD* should be empty, except for error, and therefore, they should be equal. The relevant data for Choices 6 and 10 are shown in Table 11, where the frequency of SR' is found to be more frequent than the opposite switch of preferences in 13 of 14 conditions,

Table 11
Tests of branch independence

Condition	Choice pattern			
	SS'	RS'	SR'	RR'
Text	24	13	*41	93
Reversed	26	25	33	85
Pies	62	47	59	178
Pies*	56	36	50	159
Tickets	50	37	*77	177
List	74	44	63	191
Small	44	33	*52	134
Semi-split	115	100	*147	326
Marbles FU	14	14	14	32
Marbles UF	13	9	*26	35
Decumulative	100	71	*107	163
Control-rev	23	28	31	75
Birnbaum 99b	147	173	*229	665
Birnbaum 99b lab	46	19	*55	124
Total	794	649	*984	2437

Comment. The participants were less risk-averse than expected from prior parameters. Note that RR' is the modal pattern in every case.

Choices 6 and 10. $S = (\$44, .1; \$40, .1; \$2, .8)$, $R = (\$98, .1; \$10, .1; \$2, .8)$, $S' = (\$110, .8; \$44, .1; \$40, .1)$, $R' = (\$110; .8; \$98, .1; \$10, .1)$. Asterisks indicate cases where preference reversals that violate the inverse-S weighting function of CPT are significantly more frequent than reversals consistent with the property.

including seven significant cases. This approach might underestimate the incidence of violations if different people had different types of violations, since *BC* and *AD* might both be large but equal in frequency.

A more sensitive test of branch independence, one that allows a distinction between “true” violations and those produced by “error” can be conducted in studies having at least two repetitions of each choice. Data from the lab sample of Birnbaum (1999b) are presented in Table 12 for two repetitions each of Choices 6 and 10, and Choices 17 and 12. Each set can be fit to the model of Fig. 8. When the model is fit with all parameters free, a , c , d , and e can be estimated from the data. The parameter, a , reflects whether the experimental design is “well-tuned,” because if it is far from 1/2, one cannot expect to see violations of branch independence. If restricted branch independence holds, $c = d = 1$. If both types of violations of branch independence are “real,” we expect to see people who have SR' and RS' patterns in both repetitions; in other words, we expect to see both $SR'SR'$ and $RS'RS'$. Instead, the RS' pattern is rare and is rarely repeated. With all parameters free, the estimates are $a = 0.348$, $c = 0.568$, $d = 1$, and $e = 0.172$ for Choices 6 and 10 and $a = 0.659$, $c = 0.752$, $d = 0.998$, and $e = 0.186$, with $\chi^2(11) = 15.0$ and 12.6, respectively (both ns). When d is fixed to 1, fit is unaffected, but when c and d are both fixed to 1 (assuming branch independence), $\chi^2(13) = 27.9^*$ and 29.6*, respectively. These results indicate that branch independence is sig-

Table 12

Response patterns for two tests of restricted branch independence with two repetitions of each choice

Choice pattern	Observed frequencies	
	Choices 6 and 10	Choices 17 and 12
SS'SS'	11	32
SS'SR'	4	12
SS'RS'	2	3
SS'RR'	2	2
SR'SS'	8	3
SR'SR'	8	11
SR'RS'	1	4
SR'RR'	14	9
RS'SS'	3	4
RS'SR'	2	3
RS'RS'	1	3
RS'RR'	5	4
RR'SS'	5	5
RR'SR'	9	5
RR'RS'	3	4
RR'RR'	43	20

The first two letters indicate the response pattern on the first repetition, the second indicate response pattern on the second. Data from Birnbaum (1999b).

nificantly violated and the finding that $d = 1$ indicates that only the SR' pattern appears to be “real.”

Appendix D. Evaluation or choice?

There are two types of explanations of paradoxes of choice. They may arise in the evaluation of gambles or in the comparison of alternatives. Transitive theories (including EU, RDU, RSDU, CPT, TAX, and RAM among others) attempt to predict choices as the result of a comparison of the subjective values (utilities) of the gambles. The decision maker is assumed to choose the gamble with the higher evaluation. These models imply transitivity.

In contrast, the choice approach attempts to account for results in terms of models of comparison that may or may not satisfy transitivity. Leland (1994) reviews this approach.

There has been considerable debate whether or not there are situations in which people systematically depart from transitivity (Birnbaum et al., 1999; Fishburn, 1992; Iverson & Falmagne, 1985; Loomes et al., 1991; Luce, 2000, pp. 38–39, 197–198; Starmer, 1999; Sopher & Gigliotti, 1993; Tversky, 1969). See also Appendix C.

Although choice processes and variables unique to choice (e.g., correlation among consequences) may provide methods that affect the rate of violation of stochastic dominance, it seems unlikely that choice processes alone will account for the violations reported here and in previous research. Violations of stochastic dominance, branch independence, and cumulative in-

dependence are observed in judgment studies, where each gamble is judged on its own, and separated by many trials from other gambles.

Models of choice require two or more gambles be presented in each trial for comparison, so that differences and distributions of differences become explanatory variables. However, when participants are asked to judge individual gambles, presented one at a time, it is unclear how to apply choice models.

Birnbaum and Yeary (submitted) presented 166 gambles, and asked each participant to judge the highest price a buyer should pay to buy each gamble. In a separate block of 166 trials, participants judged the lowest price a seller should accept to sell each gamble. Among these trials, separated by many intervening trials, were eight gambles that permitted four tests of stochastic dominance. Results were similar to those found in choice. For example, the median judged buying price of gamble $I = (\$96, .9; \$14, .05; \$12, .05)$ was \$30 whereas the median buying price for the inferior gamble $J = (\$96; .85; \$90, .05; \$12, .05)$ was \$60! Similarly, median judged lowest selling prices were \$73.5 and \$81.5 for the same gambles, respectively, also significantly violating stochastic dominance. Thus, to find violations of stochastic dominance, it is not necessary to present choices; people violate stochastic dominance in judgment, where gambles are presented one at a time and data are forced to be transitive.

References

- 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.
- Becker, J., & Sarin, R. (1987). Lottery dependent utility. *Management Science*, 33, 1367–1382.
- 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). Pabst, Germany: Lengerich.
- Birnbaum, M. H. (2004). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology*, 48(2), 87–106.
- 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., & Jou, J. W. (1990). A theory of comparative response times and “difference” judgments. *Cognitive Psychology*, 22, 184–210.
- 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., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45, 792–804.
- Birnbaum, M. H., & Navarrete, J. 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). Violations of stochastic dominance, branch independence, upper and lower cumulative independence in judgments of buying and selling prices of gambles. Submitted for publication.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review*, 100, 432–459.
- Camerer, C. F. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 207–251). Boston: Kluwer Academic Publishers.
- 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.
- Carbone, E. (1997). Investigation of stochastic preference theory using experimental data. *Economics Letters*, 57, 305–311.
- Carbone, E., & Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty*, 20(2), 161–176.
- 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.
- Fishburn, P. C. (1978). On Handa’s “New theory of cardinal utility” and the maximization of expected return. *Journal of Political Economy*, 86(2), 321–324.
- Fishburn, P. (1992). Nontransitive preferences and normative decision theory. In J. Geweke (Ed.), *Decision making under risk and uncertainty: New models and empirical findings* (pp. 3–10). Dordrecht: Kluwer Academic Publishers.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency format. *Psychological Review*, 102, 684–704.
- Harless, D. W. (1992). Actions vs. prospects: The effect of problem representation on regret. *American Economic Review*, 82, 634–649.
- Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62, 1251–1290.
- Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty*, 11, 263–274.
- Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences*, 10, 131–153.
- Iverson, G., & Luce, R. D. (1998). The representational approach to psychophysical and judgmental problems. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision-making* (pp. 1–79). San Diego: Academic Press.
- Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, 58(9), 723–730.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Keller, L. R. (1985). The effects of problem representation on the surething and substitution principles. *Management Science*, 31, 738–751.
- Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, 9, 151–172.
- Levy, M., & Levy, H. (2002). Prospect Theory: Much ado about nothing. *Management Science*, 48, 1334–1349.
- Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Erlbaum.
- Loomes, G., Starmer, C., & Sugden, R. (1992). Are preferences monotonic: Testing some implications of regret theory. *Economica*, 59, 17–33.
- Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Journal of Economic Theory*, 59, 425–440.
- Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39, 641–648.
- 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: Erlbaum.
- 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.
- Machina, M. J. (1982). Expected utility analysis without the independence axiom. *Econometrica*, 50, 277–323.
- Marley, A. A. J., & Luce, R. D. (2001). Rank-weighted utilities and qualitative convolution. *Journal of Risk and Uncertainty*, 23(2), 135–163.

- 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.
- Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behaviour. *Theory and Decision*, 46, 141–158.
- 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. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty*, 6, 235–254.
- Sopher, B., & Gigliotti, G. (1993). Intransitive cycles: Rational choice or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision*, 35, 311–336.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286 (Reprinted 1994, 101, 266–270).
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- 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. (2003). The Data of Levy and Levy (2002), “Prospect theory: Much ado about nothing?” Support Prospect Theory. *Management Science*, 49, 979–981.
- Wakker, P. P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, 7, 147–176.
- 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.