New Paradoxes of Risky Decision Making

Michael H. Birnbaum
California State University, Fullerton and Decision Research Center, Fullerton

During the last 25 years, prospect theory and its successor, cumulative prospect theory, replaced expected utility as the dominant descriptive theories of risky decision making. Although these models account for the original Allais paradoxes, 11 new paradoxes show where prospect theories lead to self-contradiction or systematic false predictions. The new findings are consistent with and, in several cases, were predicted in advance by simple "configural weight" models in which probability-consequence branches are weighted by a function that depends on branch probability and ranks of consequences on discrete branches. Although they have some similarities to later models called "rank-dependent utility," configural weight models do not satisfy coalescing, the assumption that branches leading to the same consequence can be combined by adding their probabilities. Nor do they satisfy cancellation, the "independence" assumption that branches common to both alternatives can be removed. The transfer of attention exchange model, with parameters estimated from previous data, correctly predicts results with all 11 new paradoxes. Apparently, people do not frame choices as prospects but, instead, as trees with branches.

Keywords: cumulative prospect theory, decision making, expected utility, rank-dependent utility, paradox

Following a period in which expected utility (EU) theory (Bernoulli, 1738/1954; Savage, 1954; von Neumann & Morgenstern, 1947) dominated the study of risky decision making, original prospect theory (OPT) became the focus of empirical studies of decision making (Kahneman & Tversky, 1979). OPT was later modified (Tversky & Kahneman, 1992) to assimilate rank- and sign-dependent utility (RSDU). The newer form, cumulative prospect theory (CPT), was able to describe the classic Allais paradoxes (Allais, 1953, 1979) that were inconsistent with EU without violating stochastic dominance. CPT simplified and extended OPT to a wider domain.

CPT describes the "four-fold pattern" of risk seeking and risk aversion in the same person. When a person prefers the expected value of a gamble to the gamble itself, that person is exhibiting "risk aversion." For example, most people prefer $50 for sure rather than the risky gamble with a 50% chance to win $100 and otherwise receive nothing. When a person prefers the gamble over its expected value, the person is described as "risk seeking." In the "four-fold pattern," the typical participant shows risk seeking for binary gambles with small probabilities to win large prizes and risk aversion for gambles with medium to high probability to win. For gambles with strictly nonpositive consequences, this pattern is reversed. Such reversal is known as the *reflection* effect. Finally, CPT describes risk aversion in mixed gambles, also known as *loss aversion*, a tendency to prefer sure gains over mixed gambles with the same or higher expected values.

Many important articles contributed to the theoretical and empirical development of these theories (Abdellaoui, 2000, 2002; Battalio, Kagel, & Jiranyakul, 1990; Bleichrodt & Luis, 2000; Camerer, 1989, 1992, 1998; Diecidue & Wakker, 2001; Gonzalez & Wu, 1999; Karni & Safra, 1987; Luce, 2000, 2001; Luce & Fishburn, 1991, 1995; Luce & Narens, 1985; Machina, 1982; Prelec, 1998; Quiggin, 1982, 1985, 1993; Schmeidler, 1989; Starmer & Sugden, 1989; Tversky & Wakker, 1995; von Winterfeldt, 1997; Wakker, 1994, 1996, 2001; Wakker, Erev, & Weber, 1994; Wu & Gonzalez, 1996, 1998, 1999; Yaari, 1987). Because of these successes, CPT has been recommended as the new standard for economic analysis (Camerer, 1998; Starmer, 2000), and it was recognized in the Nobel Prize in Economics in 2002.

However, evidence has been accumulating in recent years that systematically violates both versions of prospect theory. Some authors have criticized CPT (Baltussen, Post, & van Vliet, 2006; Barron & Erev, 2003; Blavatskyy, 2005; Brandstätter, Gigerenzer, & Hertwig, 2006; Gonzalez & Wu, 2003; González-Vallejo, 2002; Hertwig, Barron, Weber, & Erev, 2004; Humphrey, 1995; Levy & Levy, 2002; Lopes & Oden, 1999; Luce, 2000; Marley & Luce, 2005; Neilson & Stowe, 2002; Payne, 2005; Starmer, 1999, 2000; Starmer & Sugden, 1993; Weber & Kirsner, 1997; Wu, 1994; Wu & Gonzalez, 1999; Wu & Markle, 2005; Wu, Zhang, & Abdelloui, 2005). Not all criticisms of CPT have been received without controversy (Baucells & Heukamp, 2006; Fox & Hadar, 2006; Rieger & Wang, 2008; Wakker, 2003), however, and some conclude that CPT is the "best," if imperfect, description of decision making under risk and uncertainty (Camerer, 1998; Harless & Camerer, 1994; Starmer, 2000; Wu, Zhang, & Gonzalez, 2004).

My students and I have been testing prospect theories against an older class of models known as "configural weight" models (Birnbaum, 1974a; Birnbaum & Stegner, 1979). In these models, the weight of a stimulus (branch) depends on relationships between

that stimulus and others in the same set. A generic class of configural weight models includes CPT as a special case, as well as other special cases that are compared against CPT in the present article. In this article, I summarize the case against both versions of prospect theory and show that simple configural weight models provide more accurate descriptions of risky decision making.

In configural weight models, weights of probability-consequence branches depend on the probability or event leading to a consequence and the relationships between that consequence and consequences of other branches in the gamble. These models led me to reexamine old results and to deduce new properties that can be used to test among classes of models (Birnbaum, 1997). The "new paradoxes" are behavioral properties that create systematic self-contradictions in prospect theories. The properties tested are also implied by EU theory; therefore, systematic violations of these properties also contradict EU. I refer to these properties as *paradoxes* because, like the Allais paradoxes (Allais, 1953, 1979), they are stronger than simple violations of the predictions of a model; they are phenomena that lead to self-contradiction when analyzed by widely accepted theory with any functions and any choice of parameters. However, the paradoxes can be resolved by rival theory.

The mass of evidence has now reached the point where I conclude that neither version of prospect theory can be retained as a descriptive model of decision making. The violations of CPT are largely consistent with a model that is a special case of a configural weight model (Birnbaum, 1974a; Birnbaum & Stegner, 1979) known as the special transfer of attention exchange (TAX) model (Birnbaum & Chavez, 1997). Also more accurate than CPT is another type of configural model known as the rank-affected multiplicative weights (RAM) model. The violations of CPT also rule out other related models, such as rank-dependent utility (RDU) of Quiggin (1993) as well as certain other models that share some of its properties.

On the basis of the growing case against CPT, RDU, and RSDU, Luce (2000) and Marley and Luce (2001, 2005) have recently developed a new subclass of configural models, gains decomposition utility (GDU), which they have shown has similar properties to the TAX model but is distinct from it. These three models (TAX, RAM, and GDU) share the following idea: People treat gambles as trees with branches rather than as prospects or probability distributions.

There are two cases made in this article. The easier case to make is the negative one, which is to show that empirical data strongly refute both versions of prospect theory as accurate descriptions. The positive case is necessarily more tentative; namely, that the special TAX model, which correctly predicted some of the violations of CPT in advance of experiments, gives a better description of both old and new data.

Because a model correctly predicted results in a series of new tests, it does not follow that it will succeed in every new test that might be devised. Therefore, the reader may decide to accept the negative case (CPT and its relatives are false) and dismiss the positive case, favoring TAX, as a series of lucky coincidences. In this case, a better theory supported by diagnostic evidence is required.

Some introductory examples help distinguish characteristics of prospect theories from the configural weight models reviewed here. Consider the following choice:

A:	.01 probability to win $100	B:	.01 probability to win $100
	.01 probability to win $100		.02 probability to win $45
	.98 probability to win $0		.97 probability to win $0

Each gamble is represented by an urn containing 100 marbles that are otherwise identical, except for color. The urn for Gamble A contains one red marble and one blue marble, either of which pays $100, and it has 98 white marbles that pay $0 (nothing). Urn B contains one red marble paying $100, two green marbles that pay $45, and it has 97 white marbles that pay $0. A marble will be drawn blindly, at random, from the chosen urn, and the prize will depend on the color of marble drawn. Would you rather draw a ball from A or from B?

In OPT, people are assumed to simplify such choices by editing (Kahneman & Tversky, 1979). Gamble A has two branches with probability .01 to win $100. In prospect theory, these two branches are combined to form a two-branch gamble, A', with one branch of .02 to win $100 and a second branch of .98 to win $0. If a person were to combine the two branches leading to the same consequence, then A and A' would be the same, so the choice between A and B would be the same as that between A' and B, as follows:

A':	.02 probability to win $100	B:	.01 probability to win $100
	.98 probability to win $0		.02 probability to win $45
			.97 probability to win $0

In CPT, the equivalence of these choices is guaranteed by its most general representation, with or without any additional steps of editing (proof in Birnbaum & Navarrete, 1998, pp. 57–58). As is clear in this study, this equivalence, common to both original and cumulative prospect theories, is empirically false.

In OPT, it was also assumed that people cancel common branches. In the example, A and B share a common branch of .01 to win $100, so this branch might be cancelled before a choice is made. If so, then the choice between A and B should be the same as the following:

A":	.01 probability to win $100	B':	.02 probability to win $45
	.99 probability to win $0		.98 probability to win $0

In CPT, the representation does not, however, satisfy cancellation in general. But if people were assumed to use cancellation as an editing rule before evaluating the gambles, then they might satisfy this property as well. As we see in this article, this property can also be rejected when it is tested empirically.

These two principles, combination and cancellation, are violated by branch-weighting theories such as TAX, RAM, and GDU. Thus, the three-branch gamble, A, and the two-branch gamble A', which are equivalent in prospect theory, are different in RAM and TAX, except in special cases. Furthermore, these models do not assume that people "trim trees" by canceling branches common to both alternatives in a choice.

It will be helpful to preview three other issues that distinguish descriptive decision theories: the source of risk aversion, the effects of splitting of branches, and the origins of loss aversion.

Two Theories of Risk Aversion

The term *risk aversion* refers to the empirical finding that people often prefer a sure thing over a gamble with the same or even higher expected value (EV). Consider the following choice:

F: $45 for sure   G: .50 probability to win $0
                       .50 probability to win $100

This represents a choice between F, a "sure thing" to win $45, and a two-branch gamble, G, with equal chances of winning $0 or $100. The lower branch of G is .5 to win $0, and the higher branch is .5 to win $100. Most people prefer $45 for sure rather than gamble G, even though G has a higher expected value of $50; therefore, they are said to exhibit *risk averse* preferences.

Two distinct ways of explaining such risk aversion are illustrated in Figures 1 and 2. In this example, a person prefers any cash value greater than $33.3 to the gamble with equal chances of winning $100 or $0, and this same person prefers the gamble to any cash value less than $33.3. In EU theory, it is assumed that people choose F over G (denoted F > G) if and only if EU(F) > EU(G), where

$$EU(F) = \sum p_i u(x_i), \qquad (1)$$

and $u(x)$ is the utility (subjective value) of the cash prize, x. In Figure 1, there is a nonlinear transformation from objective money to utility (subjective value). If this utility function, $u(x)$, is a concave downward function of money, x, then the expected utility of G (fifty-fifty to win $100 or $0) can be less than that of F($45 for sure). For example, if $u(x) = x^{.63}$, then $u(F) = 11.0$, and $EU(G) = .5u(0) + .5u(100) = 9.1$.

Because $EU(F) > EU(G)$, EU can imply preference for F over G. Figure 1 shows that on the utility continuum, the balance point on the transformed scale (the expectation) corresponds to the utility of $33.3. Thus, this person should be indifferent between a sure gain of $33.3 and gamble G (denoted $33.3 ~ G). The cash
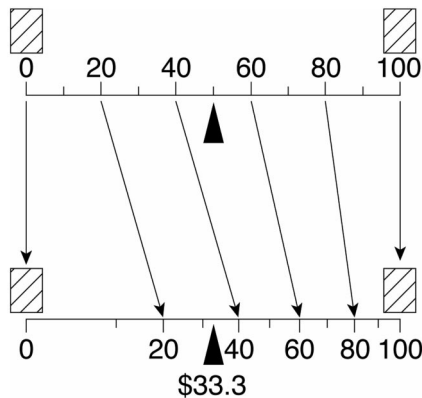

Figure 1. Utility theory of risk aversion. Expected utility is the center of gravity. The gamble, G = ($100, .5; $0), is represented as a probability distribution, with half of its weight at 0 and half at 100. The expected value is 50, the balance point in the upper portion of the figure. In the lower portion of the figure, distance along the scale corresponds to utility, illustrated with the function $u(x) = x^{.63}$. Marginal differences in utility between $20 increments decrease as one goes up the scale. The balance point on the utility axis corresponds to $33.3, which is the certainty equivalent of this gamble.
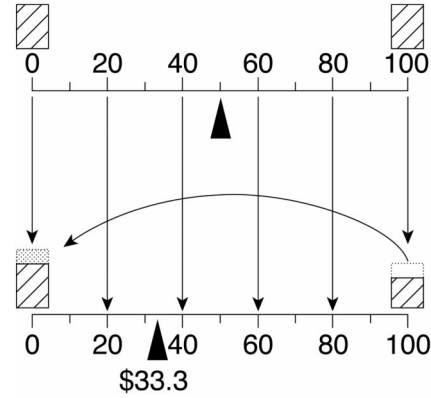

Figure 2. A configural weight theory, illustrated here with $u(x) = x$. Suppose one third of the weight of the higher branch (win $100) is transferred to the lower branch (win $0), then the certainty equivalent (balance point) would be $33.3, same as in Figure 1. Expected utility (see Figure 1) and configural weighting (as illustrated in this figure) can both describe risk aversion, but they make different predictions that can be tested.

value with the same utility as a gamble is known as the gamble's *certainty equivalent*, $CE(G) = u^{-1}[EU(G)]$. In this case, $CE(G) = $33.3. Similarly, EU can accommodate risk seeking by means of a positively accelerated $u(x)$ function and risk neutrality with a linear utility function.

A second way to explain risk aversion is shown in Figure 2. In the TAX model illustrated, one third of the weight of the higher branch is taken from the branch to win $100 and assigned to the lower valued branch to win $0. The weights of the lower and higher branches are thus 2/3 and 1/3, respectively, so the balance point corresponds to a CE of $33.3. In this example, the transformation from money to utility is linear, and it is weighting rather than utility that produces risk aversion. Intuitively, the extra weight applied to the lowest consequence represents a transfer of attention from the highest to lowest consequence of the gamble.

Although the difference between the utility and weight theories (Figure 1 vs. Figure 2) might seem merely two equivalent mathematical tricks to account for the same thing, these two models lead to very different implications that can be tested empirically, as shown in the *Violations of Gain–Loss Separability* section.

Whereas EU (see Figure 1) attributes risk aversion to the utility function, the configural weight TAX model (see Figure 2) attributes risk aversion to a transfer of attention from the higher to the lower valued branch. When people give more attention (place more weight) on the branch leading to lower consequence, they will be risk averse, all other factors being the same.

Why might people place more weight on lower valued consequences? Birnbaum and Stegner (1979) theorized that judges who evaluate risky or uncertain objects have asymmetric costs for over- as opposed to underestimation of value. Participants in the 1979 study were asked to judge the value of used cars on the basis of blue book value and advice from people who examined the cars. Birnbaum and Stegner theorized that the configural weights could be manipulated by instructions to identify with the buyer, seller, or a neutral judge who was asked to evaluate the "fair" price of the cars. Birnbaum and Stegner found that when people are asked to

advise buyers, they place greater weight on the lower estimates of value, and when advising sellers, they place more weight on the higher estimates.

According to the theory, people place more weight on lower estimates because the more costly error for a buyer is paying too much for a car, whereas the more costly error for the seller is to accept too low a value. Birnbaum, Coffey, Mellers, and Weiss (1992) presented a derivation showing that configural weighting models can be deduced from asymmetric cost functions (see also Birnbaum & McIntosh, 1996). Weber (1994) reviewed this intuitive basis for configural weighting and contrasted it with other intuitions that lead to different predictions. Diecidue and Wakker (2001) presented the intuitive ideas behind the rank-dependent models. As shown in The Case Against Prospect Theories in Choice section below, these alternate intuitions that might seem similar when described in words lead to very different predictions that can be tested empirically.

Quiggin's (1993) RDU, Luce and Fishburn's (1991, 1995) RSDU, Tversky and Kahneman's (1992) CPT, Marley and Luce's (2001) idempotent, lower GDU, Birnbaum's (1974a) range model, his (1997) RAM, and TAX models are all members of a generic class of *configural weight* models in that they can account for risk aversion by the assumption that branches with lower consequences receive greater weight. In these models, configural weighting describes risk aversion apart from the nonlinear transformation from money to utility (Lopes, 1990).

These two ways to represent risk aversion are not mutually exclusive; therefore, both mechanisms might combine to produce risk aversion or risk seeking in a given person. The next issue subdivides these models into three groups that can also be compared by experiment with each other.

Two Theories of Branch Splitting

Let $G = (x, p; y, q; z, r)$ represent a three-branch gamble that yields monetary consequences of x with probability p, y with probability q, and z otherwise ($r = 1 - p - q$). A *branch* of a gamble is a probability (event)–consequence pair that is distinct in the display to the decision maker.

Consider the two-branch gamble $G = (\$100, .5; \$0, .5)$. Suppose we split the branch of .5 to win \$100 into two branches of .25 to win \$100. That creates a three-branch gamble, $G' = (\$100, .25; \$100, .25; \$0, .5)$, which is a "split" form of G, which is called the *coalesced* form. It should be clear that there are many ways to split G but only one way to coalesce branches from G' to G. According to *upper coalescing*, $G \sim G'$. Similarly, suppose we split the lower branch of G, $G'' = (\$100, .5; \$0, .25; \$0, .25)$, which is another split version of G with two lower branches of .25 to win \$0. By *lower coalescing*, it is assumed that $G \sim G''$.

We can divide theories in three classes: First, some models satisfy all types of coalescing (including RDU, RSDU, CPT, EU, and OPT with the editing principle of combination). Second, there are models that violate both upper and lower coalescing (including RAM, TAX, and subjectively weighted utility [SWU] also called "stripped" prospect theory, without the editing rules). Third, some models satisfy some but not all forms of coalescing; for example, idempotent, lower GDU satisfies upper coalescing and violates lower coalescing, except in a special case in which it reduces to

RDU. According to the RDU/RSDU/CPT models, a gamble has the same value (utility) whether branches are split or coalesced.

According to models that violate coalescing, however, the sum of the weights of the two splinters need not equal the weight of the coalesced branch. In particular, these models (as fit to data) imply that the total weight of the splinters typically exceeds the weight of the coalesced branch.

Theories that violate coalescing can also be further subdivided into those that satisfy idempotence (e.g., RAM and TAX) and those that do not satisfy this property (e.g., OPT without its editing rules). *Idempotence* is the assumption that $x \sim (x, p_1; x, p_2; \ldots; x, p_n)$ for all n and $\{p_i\}$ such that $\sum_{i=1}^{n} p_i = 1$.

Two Theories of Loss Aversion

The term *loss aversion* has been used to refer both to a behavioral phenomenon and to a class of theories that might account for the phenomenon (Tversky & Kahneman, 1992). This circular terminology creates confusion if one believes that the phenomena can be replicated but that the theory is not correct. As Schmidt and Zank (2005) noted, empirical phenomena of "risk aversion" and "loss aversion" should be distinguished from theories of those phenomena. When the same terms are used for both the phenomenon to be explained and for a particular account of that phenomenon, it can lead to theoretical confusion. I use the term *loss aversion* to refer to the behavioral finding that people show risk aversion for mixed gambles. Although the data for loss aversion are less numerous and less consistent than the evidence for risk aversion (Birnbaum, 2006; Birnbaum & Bahra, 2007; Ert & Erev, in press), researchers have reported that undergraduates will only accept gambles to win x or lose y with equal probability if the amount to win exceeds twice the amount to lose (Tversky & Kahneman, 1992). Two ideas have been proposed to account for this finding.

In CPT, loss aversion is represented by the utility function, as defined on gains and losses. The utility account of loss aversion assumes that $|u(-x)| > u(x)$ for $x > 0$; that is, that "losses loom larger than gains." Tversky and Kahneman (1992) theorized that the utility for losses is a multiple of the utility for equal gains:

$$u(-x) = -\lambda u(x),$$

$$x \geq 0, \tag{2}$$

where λ is the loss-aversion coefficient. Suppose $\lambda = 2$. If so, then it follows from this model of CPT that a fifty-fifty gamble to either win or lose \$100 will have a negative utility, assuming the weightings of .5 to win and .5 to lose are equal. A number of other expressions have been suggested to describe how utilities of gains relate to losses of equal absolute value; these are reviewed and compared in Abdellaoui, Bleichrodt, and Paraschiv (2007). These variations all attribute loss aversion to the utility function.

A second way to represent loss aversion, in contrast, is the idea that loss aversion is another consequence of configural weighting. As in Figure 2, suppose the weights of two equally likely branches to lose \$100 or win \$100 are 2/3 and 1/3, respectively. Even with $u(x) = x$ for both positive and negative values of x (e.g., $u(-100) = -u(100) = -100$), this model implies that this mixed gamble has a value of –\$33, so people would avoid such mixed gambles.

These two theories of loss aversion—utility versus weight—make different predictions for a testable property known as *gain–loss separability*, as is shown in the *Violations of Gain–Loss Separability* section. These theories also give different accounts of buying and selling prices (sometimes called the "endowment" effect) that can be tested. The configural weighting theory of buying and selling prices (Birnbaum & Stegner, 1979) assumes that buyers place more weight on lower estimates, and sellers place relatively more weight on higher estimates of value. According to the model of Tversky and Kahneman (1991), however, the difference between willingness to pay (buyer's prices) and willingness to accept (seller's) prices is due to a kink in the utility function, as in Equation 2 (Birnbaum & Zimmermann, 1998). As is the case with risk aversion, these two factors are not mutually exclusive; both factors could contribute to the phenomenon.

Notation and Terminology

There are a few other terms that should be defined here. Preferences are said to be *transitive* if for all A, B, and C, $A > B$ and $B > C$ implies $A > C$. All of the models considered in this article are transitive; intransitive models are considered in detail elsewhere (Birnbaum, 2008a, 2008b; Birnbaum & Gutierrez, 2007; Birnbaum & LaCroix, 2008).

Consequence monotonicity is the assumption that if one consequence in a gamble is improved, holding everything else constant, then the gamble with the better consequence should be preferred. For example, if a person prefers \$100 to \$50, then the gamble $G = (\$100, 0.5; \$0)$ should be preferred to $F = (\$50, 0.5; \$0)$.

Branch independence is weaker than Savage's (1954) "sure thing" axiom. It holds that if two gambles have an identical probability-consequence branch, then the value of the consequence on that branch can be changed without altering the order of the gambles induced by the other branches. For example, for three-branch gambles, branch independence requires

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

if and only if

$$S' = (x, p; y, q; z', r) > R' = (x', p'; y', q'; z', r), \quad (3)$$

where the consequences (x, y, z, x', y', z') are all distinct, all probabilities are greater than zero and sum to 1 in each gamble, so $p + q = p' + q'$. The branch (z, r) is known as the *common branch* in the first choice, and (z', r) is the common branch in the second choice. This principle is weaker than Savage's independence axiom because it holds for gambles with equal numbers of branches of known probability and also because it does not presume coalescing.

When the distributions and number of consequences in the gambles are restricted to be the same in all four gambles ($p = p'$; $q = q'$), then this property is termed *restricted branch independence* (RBI). The special case of restricted branch independence, in which corresponding consequences retain the same ranks, is termed *comonotonic restricted branch independence*, also known as comonotonic independence (Wakker et al., 1994). Cases of (unrestricted) branch independence in which only the probability distribution is systematically varied are termed *distribution independence*.

Stochastic dominance (first-order stochastic dominance) is the relation between nonidentical gambles, F and G, such that for all values of x, the probability of winning x or more in gamble G is greater than or equal to the probability of winning x or more in gamble F. If so, then G is said to *stochastically dominate* gamble F. The statement that preferences satisfy stochastic dominance means that if G dominates F, then F will not be preferred to G.

Subjectively Weighted Utility and Prospect Theories

Subjectively Weighted Utility Theory

Let $G = (x_1, p_1; x_2, p_2; \ldots; x_i, p_i; \ldots; x_n, p_n)$ represent a gamble to win x_i with probability p_i, where the n outcomes are mutually exclusive and exhaustive. Edwards (1954, 1962) considered subjectively weighted utility (SWU) models of the form,

$$SWU(G) = \sum_{i=1}^{n} w(p_i)u(x_i). \quad (4)$$

Edwards (1954) discussed both S-shaped and inverse-S-shaped functions as candidates for the weighting function, $w(p)$. When $w(p) = p$, this model reduces to EU. Karmarkar (1978, 1979) also worked with this type of model as a description of risky decision making. Edwards (1962) theorized that weighting functions might differ for different configurations of consequences. He made reference to a "book of weights" with different pages for different cases, each of which would describe a weighting function, $w_C(p)$.

Edwards (1962) wrote:

> The data now available suggest the speculation that there may be exactly five pages in that book, each page defined by a class of possible payoff arrangements. In Class 1, all possible outcomes have utilities greater than zero. In Class 2, the worst possible outcome (or outcomes, if there are several possible outcomes all with equal utility), has a utility of zero. In Class 3, at least one possible outcome has a positive utility and at least one possible outcome has a negative utility. In Class 4, the best possible outcome or outcomes has a utility of zero. And in Class 5, all possible outcomes have negative utilities. (p. 128)

Prospect Theory

Kahneman and Tversky's (1979) prospect theory is similar to that of Edwards (1962), except it was restricted to gambles with no more than two nonzero consequences, and it reduced the number of pages in the book of weights to two—prospects (gambles) with and without the consequence of zero. This means that the 1979 theory is silent on many of the tests and properties that are described in this article unless assumptions are made about how to extend it to gambles with more than two nonzero consequences. One way to extend OPT is to treat the theory as the special case of the Edwards (1962) model.

Another way to generalize it is to use Equation 4 as "stripped" prospect theory (Starmer & Sugden, 1993). The term *stripped* was used to indicate that the editing principles and the use of differential weighting for the consequence of \$0 have been removed, though different weighting functions would be allowed for positive and negative consequences. A third way to extend it is to use CPT (Tversky & Kahneman, 1992), defined in the RDU and CPT section. Each of these approaches leads to a different theory, so it

is best to consider "prospect theory" as a large family of different, contradictory theories.

Besides the restriction in OPT to prospects with no more than two nonzero consequences, the other new feature of OPT was the idea that an editing phase precedes the evaluation phase. Kahneman (2003) described the development of the editing rules in an article reviewing his collaboration with Amos Tversky that led to his winning a share of the 2002 Nobel Prize. He described how, when they were completing their 1979 article, they worried that their model could be easily falsified. For example, Equation 4 predicts that people can violate stochastic dominance when every consequence of one gamble is less than every consequence of the other gamble (Birnbaum, 1999b; Fishburn, 1978). To rescue the model from such implications they thought were false, they added editing rules that provided preemptive excuses for potential refuting evidence. The editing rules also provided ways to generalize OPT to gambles with more than two nonzero branches.

Editing Rules in Prospect Theory

The six principles of editing are as follows:

A. Combination: Probabilities associated with identical outcomes are combined. This principle implies coalescing.

B. Segregation: A riskless component is segregated from risky components. "The prospect (300, .8; 200, .2) is naturally decomposed into a sure gain of 200 and the risky prospect (100, .8)" (Kahneman & Tversky, 1979, p. 274).

C. Cancellation: Components shared by both alternatives are discarded from the choice. "For example, the choice between (200, .2; 100, .5; –50, .3) and (200, .2; 150, .5; –100, .3) can be reduced by cancellation to a choice between (100, .5; –50, .3) and (150, .5; –100, .3)." (Kahneman & Tversky, 1979, pp. 274–275). If subjects cancel common components, then they will satisfy branch independence, as described in the *Decomposition of the Allais Common Consequence Paradox* section. Note that this example shows how to handle three-branch gambles that are otherwise excluded from the model and that it implies RBI.

D. Dominance: Transparently dominated alternatives are recognized and eliminated. Without this principle, Equation 4 violates dominance in cases in which people do not.

E. Simplification: Probabilities and consequences are rounded off. This principle, combined with cancellation, could produce violations of transitivity.

F. Priority of editing: Editing precedes and takes priority over evaluation. Kahneman and Tversky (1979) remarked, "Because the editing operations facilitate the task of decision, it is assumed that they are performed whenever possible" (p. 275). The editing rules are unfortunately imprecise, contradictory, and conflict with the equations of prospect theory. This means that OPT often has the misfortune (or luxury) of predicting opposite results, depending on what principles are invoked or the order in which they are applied (Stevenson, Busemeyer, & Naylor, 1991). This makes the theory easy to use post hoc but makes it difficult to use as a predictive scientific model.

In Kahneman and Tversky's (1979) article, there were 14 choice problems, none of which involved gambles with more than two nonzero consequences. To fit these 14 problems, there were two functions (value and probability), a status quo point, and six editing principles. In addition, special biases were postulated, such

as consequence "framing," which contradicts the editing principle of segregation. For example, when a participant is asked to choose between $45 for sure and a 50-50 gamble to win $100 or $0, most people take the sure $45. But if the person is given $100 contingent on their taking one of two losing gambles, then he or she prefers a 50-50 gamble to lose nothing or lose the $100 rather than take a sure loss of $55. But if people can integrate the $100 endowment into the losing gambles, then they would see that the two choices are objectively the same. Similarly, if they could segregate the "sure thing" in both cases, then they would also see that both choices are objectively the same. Because of these complexities and self-contradictions, it might seem that the editing rules make prospect theory untestable. Nevertheless, if each editing rule is taken as a separate scientific theory, it can be isolated and tested one by one. So far, none of the editing rules has been found to be consistent with results of such direct tests.

RDU and CPT

RDU theory was proposed as a way to explain the Allais paradoxes without violating transparent dominance (Quiggin, 1982, 1985, 1993). CPT (Tversky & Kahneman, 1992) was considered an advance over OPT because it applied to gambles with more than two nonzero consequences, and because it removed the need for the editing rules of combination and dominance detection, which are automatically guaranteed by the representation. CPT uses the same representation as RSDU (Luce & Fishburn, 1991, 1995), though the two theories were derived from different assumptions (Luce, 2000; Wakker & Tversky, 1993). CPT is also more general than OPT in that it allows different weighting functions for positive and negative consequences.

For gambles with strictly nonnegative consequences, RDU, RSDU, and CPT all reduce to the same representation. With $x_1 \geq x_2 \geq \ldots \geq x_i \geq \ldots \geq x_n \geq 0$, the representation is:

$$RDU(G) = \sum_{i=1}^{n} \left[W^+ \left(\sum_{j=1}^{i} p_j \right) - W^+ \left(\sum_{j=1}^{i-1} p_j \right) \right] u(x_i), \quad (5)$$

where $RDU(G)$ is the rank-dependent expected utility of gamble G; $W^+(P)$ is the weighting function of decumulative probability, $P_i = \sum_{j=1}^{i} p_j$, which monotonically transforms decumulative probability to decumulative weight and assigns $W^+(0) = 0$ and $W^+(1) = 1$.

For gambles of strictly negative consequences, a similar expression to that in Expression 5 is used, except $W^-(P)$ replaces $W^+(P)$, where $W^-(P)$ is a function that assigns cumulative weight to cumulative probability of negative consequences. For gambles with mixed gains and losses, with consequences ranked such that: $x_1 \geq x_2 \geq \ldots \geq x_n \geq 0 > x_m \geq \ldots x_{n+2} \geq x_{n+1}$, CPT utility is the sum of two terms, as follows:

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

$$+ \sum_{i=n+1}^{m} \left[W^- \left(\sum_{j=n+1}^{i} p_j \right) - W^- \left(\sum_{j=n+1}^{i-1} p_j \right) \right] u(x_i). \quad (6)$$

This is the same as the representation in RSDU. Note that the overall utility of a mixed gamble is just the sum of the evaluations

of the "good" and "bad" parts of the gamble. This additive representation implies gain–loss separability, as is shown in the *Violations of Gain–Loss Separability* section.

The parameterized model of CPT further specifies the functions as follows: The utility ("value") function is represented by $u(x) = x^\beta$, for $x \geq 0$. Tversky and Kahneman (1992) reported a best-fit value of $\hat{\beta} = 0.88$. They fit negative consequences with the assumption that $u(-x) = -\lambda u(x)$, where λ, which represents "loss aversion," was estimated to be 2.25. The cumulative weighting function for losses, $W^-(P)$, and the decumulative weighting function for gains, $W^+(P)$ were estimated to be nearly identical and were both fit as inverse-S functions. Tversky and Kahneman (1992) fit $W^+(P) = P^\gamma/[P^\gamma + (1 - P)^\gamma]^{1/\gamma}$, where $\hat{\gamma} = 0.61$. This function was also fit by the equation, $W^+(P) = cP^\gamma/[cP^\gamma + (1 - P)^\gamma]$, where $\hat{c} = 0.724$ and $\hat{\gamma} = 0.61$ (Tversky & Fox, 1995; Tversky & Wakker, 1995). Throughout the remainder of the present article, predictions are calculated using these parameter values. For a review of functional forms and best-fit parameter estimates in CPT, see Stott (2006).

RAM, TAX, and GDU Models

Six preliminary comments are in order. First, the approach described here, like that of OPT or CPT, is purely descriptive. The models are intended to describe and predict humans' decisions and judgments rather than to prescribe how people should decide. Second, as in OPT and CPT, utility (or value) functions are defined on changes from the status quo rather than total wealth states. See also Edwards (1954, 1962) and Markowitz (1952).

Third, the approach is psychological; how stimuli are described, presented, or framed is part of the theory (Edwards, 1954, 1962). Indeed, these models were originally developed as models of social judgment, perceptual psychophysics, and buying and selling prices (Birnbaum, 1974a; Birnbaum, Parducci, & Gifford, 1971; Birnbaum & Stegner, 1979). Two situations that are objectively equivalent but described differently must be kept distinct in theory if people respond differently to them. For example, the objective probability of "heads" in a coin toss equals the probability of drawing a black card from a standard deck, but people might have different subjective probabilities for these events.

Fourth, RAM, TAX, and GDU models, like SWU, OPT, and CPT, emphasize the role of weighting of probability; it is in the details of this weighting that these theories differ. Whereas CPT works with decumulative probabilities, RAM, TAX, and GDU work with branch probabilities.

Fifth, although I use the terms *utility* and *value* interchangeably and use the notation $u(x)$ to represent utility (value), rather than $v(x)$ as in OPT and CPT, the $u(x)$ functions of configural weight models are treated as psychophysical functions and should not be confused with "utility," as defined by EU, nor should these functions be imbued with other excess meaning. The estimated utility functions in different models of risky decision making will in fact be quite different.

The psychophysical functions that convert objective probabilities to decision weights will reflect principles of psychophysics and judgment, including contextual effects (Birnbaum, 1974c, 1982, 1992a, 1992b; Mellers & Birnbaum, 19982; Rose & Birnbaum, 1975). Subjective scales of utility of money and probability will also show individual differences. In addition, configural weighting will be affected by individual differences and by ma-

nipulations that affect the judge's point of view, such as instructions to identify with a buyer or seller (Birnbaum et al., 1992; Birnbaum & Stegner, 1979).

Despite these sources of variation, models preserving scale convergence (the assumption that two ways to measure utility for the same person in the same context should be the same) are preferred over those that require a new scale of utility in each new situation (Birnbaum, 1974a, 1982; Birnbaum & Sutton, 1992; Mellers, Ordóñez, & Birnbaum, 1992).

Sixth, in my description of configural weight models, I pointed out that "the weight of an item depends in part on its rank within the set" (Birnbaum, 1974a, p. 559). Thus, the configural weight models have some similarities to the models that later became known as "rank-dependent" weighting models. However, in the RAM or TAX models, the weight of each branch depends on the ranks of discrete branch consequences, whereas in the RDU models, decumulative weight is a monotonic function of decumulative probability. In other words, the definition of *rank* differs in the two approaches. For example, there are exactly two ranks (lower and higher) in two-branch gambles, and there are exactly three ranks (lowest, middle, highest) in three-branch gambles, no matter what their probabilities are.

The RAM Model

In RAM, the weight of each branch of a gamble is the product of a function of the branch's probability multiplied by a constant that depends on the rank and augmented sign of the branch's consequence (Birnbaum, 1997). Augmented sign takes on three levels, $-$, $+$, and 0. Rank refers to the rank of the branch's consequence relative to other branches in the gamble, ranked such that $x_1 \geq x_2 \geq \ldots \geq x_n$.

$$RAMU(G) = \frac{\sum_{i=1}^{n} a(i, n, s_i) t(p_i) u(x_i)}{\sum_{i=1}^{n} a(i, n, s_i) t(p_i)}, \qquad (7)$$

where RAMU(G) is the utility of gamble G in the RAM model, $t(p)$ is a strictly monotonic function of probability, i and s_i are the rank and augmented sign of the branch's consequence, and $a(i, n, s_i)$ are the rank and augmented sign-affected branch weights. Rank takes on levels of $i = 1$ or 2 in two-branch gambles; $i = 1, 2,$ or 3 in three-branch gambles, and so on. The function $t(p)$ describes how a branch's weight depends on its probability, apart from these configural effects.

For choices between two-, three-, and four-branch gambles on strictly positive consequences, it has been found that rank weights are approximately equal to their branch ranks. In other words, the rank weights in a three-branch gamble are 3 for the lowest branch, 2 for the middle branch, and 1 for the branch with the highest consequence. In two-branch gambles, lower and upper branches have weights of 2 and 1, respectively. In practice, $t(p)$ is approximated by a power function, $t(p) = p^\gamma$ (typically, $0 < \gamma < 1$), and $u(x)$ is also approximated by a power function, $u(x) = x^\beta$, where typically $0 < \beta < 1$. For gambles with small stakes (pocket money), $u(x) = x$ for \$1 $< x <$ \$150 provides a good enough approximation to illustrate the model. These parameters roughly

approximate the data of Tversky and Kahneman (1992). The model with these parameters are called the "prior" RAM model. The term *prior* is used to indicate that parameters were taken to approximate previous data and used to predict new phenomena of choice that had not been previously tested.

For two-branch gambles on positive consequences, this model implies that *cash equivalents* (CEs) are an inverse-S function of probability to win the higher consequence. For example, with $u(x) = x$, CEs of gambles, $G = (\$100, p; \$0)$, are given by $CE(G) = \dfrac{1 \cdot p^{\gamma} \cdot 100}{1 \cdot p^{\gamma} + 2 \cdot (1 - p)^{\gamma}}$, which gives a good approximation of empirical data of Tversky and Kahneman (1992), as shown in Figure 9 of Birnbaum (1997). It is important to keep in mind that the probability weighting function, $t(p) = p^{\gamma}$, is a negatively accelerated function, not an inverse S, even though CEs are an inverse-S function of probability in binary gambles.

Equation 7 allows for different weighting of positive, negative, and zero branches with the same rank (Birnbaum, 1997). However, one might simplify further by assuming that in mixed gambles (with positive and negative consequences), consequences are weighted in the same fashion as for strictly positive gambles. This means that negative consequences get greater weight than equally probable positive consequences in mixed gambles. This approach contrasts with CPT, in which the utilities of negative consequences are changed by a "loss aversion" parameter, but the decumulative and cumulative weighting functions for positive and negative consequences are approximately the same.

Another simplification can be added that will imply reflection. To describe gambles with consequences that are all less than or equal to zero, one can use the same equations with absolute values of consequences and then multiply the result by –1. This is probably an oversimplification, because reflection is not perfect (Birnbaum, 2006). This simplification implies that the order for strictly nonnegative gambles $\langle A, B, \succ \rangle$ and for strictly nonpositive gambles $\langle -A, -B, \succ \rangle$ will show "reflection"; that is, $A \succ B \Leftrightarrow -A \prec -B$, where –A is the same as A, except that each positive consequence has been converted from a gain to a loss.

When the judge's point of view is manipulated (e.g., buyer's vs. seller's prices), it is assumed that the only parameters affected are the rank weights (Birnbaum, 1997; Birnbaum & Beeghley, 1997; Birnbaum et al., 1992; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992; Birnbaum & Zimmermann, 1998). The rank weights for the buyer's point of view are similar to those that reproduce choices but are typically more risk averse. The rank weights for the seller's point of view show greater weight of branches leading to the better consequences.

Intuitively, the RAM model says that people evaluate each gamble as a weighted average in which the weight of a probability-consequence branch is a function of probability and rank of the branch. If branch weights are independent of rank, and if $\gamma = 1$, then this model reduces to EU. However, when $\gamma \neq 1$ and branch weights are unequal, this model violates coalescing.

The TAX Model

Like the RAM model, the TAX model represents the utility of a gamble as a weighted average of the utilities of the consequences. Weights also depend on probability and rank of the branches; however, in TAX, the branch weights result from trans-fers of attention from branch to branch. Intuitively, a decision maker deliberates by attending to the possible consequences of an action. Those branches that are more probable deserve more attention, but branches leading to lower valued consequences also deserve greater attention if a person is risk averse. In the TAX model, these shifts in attention are represented by weights transferred from branch to branch. If there were no configural effects, then each branch would have weights purely as a function of probability, $t(p)$. However, depending on the participant's point of view ("risk attitude"), weight is transferred from branch to branch. Let $\omega(p_i, p_k, n)$ represents the weight transferred from branch k to branch i ($k \leq i$, so $x_k \geq x_i$, hence transferred from branches with higher consequences to branches with lower consequences). The TAX model can then be written as:

$$TAXU(G) = \frac{\sum_{i=1}^{n} t(p_i)u(x_i) + \sum_{i=1}^{n} \sum_{k=1}^{i} [u(x_i) - u(x_k)]\omega(p_i, p_k, n)}{\sum_{i=1}^{n} t(p_i)}.$$

(8)

This model is fairly general. Note that if the weight transfers are all zero, then this model reduces to a subjectively weighted average utility model (Birnbaum, 1999a); unlike SWU and OPT, this model satisfies idempotence. Several other interesting special cases of Equation 8 have been studied by Marley and Luce (2001, 2005).

The "special" TAX model (Birnbaum, 1999b; Birnbaum & Stegner, 1979) assumes that all weight transfers are the same fixed proportion of the weight of the branch giving up weight, as follows:

$$\omega(p_i, p_k, n) = \begin{vmatrix} \dfrac{\delta \cdot t(p_k)}{n + 1}, & \delta > 0 \\ \dfrac{\delta \cdot t(p_i)}{n + 1}, & \delta \leq 0 \end{vmatrix}.$$

(8a)

In this TAX model, the amount of weight transferred between any two branches is a fixed proportion of the (transformed) probability of the branch losing weight. If lower ranked branches have more importance (as they would do in a "risk-averse" person), then it is theorized that weight is transferred from branches with higher consequences to those with lower valued consequences, that is, $\delta > 0$. Intuitively, this model assumes that the total amount of attention available is fixed, and when attention is shifted from one branch to another, what is taken from one branch is given to the others.[1]

In a binary gamble with two equiprobable, positive branches (assuming $\delta > 0$ and $\gamma < 1$), splitting the upper branch tends to make a gamble better, and splitting the lower branch tends to make

---

[1] Three calculators are freely available from http://psych.fullerton.edu/mbirnbaum/calculators/

These calculators also allow calculations of RAM, TAX, the CPT model of Tversky and Kahneman (1992), as well as the revised model in Tversky and Wakker (1995). A CPT calculator by Veronika Köbberling is also linked from the same URL.

it worse. Splitting both upper and lower branches tends to make such a gamble worse. Variations of this model in which the denominator in Expression 8a is n or $n - 1$ instead of $n + 1$ have also been considered to represent cases in which splitting both upper and lower branches will have no effect or improve the gamble, respectively (Birnbaum, 2007b).

This model is equivalent to that of Birnbaum and Chavez (1997), who based it on the "revised" configural weight model in Birnbaum and Stegner (1979). This model was also illustrated in Birnbaum and McIntosh (1996, Figure 3). However, the notational conventions have been changed from earlier presentations, so that $\delta > 0$ in the present article corresponds to a weight transfer from higher ranked to lower ranked consequences, which was represented by $\delta < 0$ in previous articles.

When lower valued branches receive greater weight ($\delta > 0$), this special TAX model can be written for three-branch gambles, $G = (x_1, p_1; x_2, p_2; x_3, p_3)$, where $x_1 \geq x_2 \geq x_3 > 0$ as follows:

$$TAXU(G) = \frac{Au(x_1) + Bu(x_2) + Cu(x_3)}{A + B + C}. \quad (9)$$

where

$$A = t(p_1) - 2\delta t(p_1)/4$$

$$B = t(p_2) - \delta t(p_2)/4 + \delta t(p_1)/4$$

$$C = t(p_3) + \delta t(p_1)/4 + \delta t(p_2)/4$$

In three-branch gambles with positive consequences, when $\delta = 1$, one fourth of the probability weight of any higher valued branch is transferred to each lower valued branch.

In two-branch gambles, with $\delta = 1$, one third of the weight of the higher branch is given to the lower branch (as in Figure 2).

The special TAX model has three parameters, γ represents the psychophysical function for probability, $t(p) = p^{\gamma}$, β represents the utility (value) function of monetary consequences, $u(x) = x^{\beta}$, and δ represents the configural transfer of weight (which affects risk aversion or risk seeking).

One can roughly approximate the data of Tversky and Kahneman (1992) with $\beta = 1$, $\gamma = 0.7$, and $\delta = 1$ for nonnegative consequences less than \$150. I use the term *prior* TAX model in reference to these parameter values. Although best-fit estimates of $\hat{\beta}$ are typically less than 1 (Birnbaum & Chavez, 1997; Birnbaum & Navarrete, 1998), the linear approximation is used here to simplify the presentation and because it suffices to reproduce most of the phenomena reviewed here. This simplification also shows that most of the major phenomena can be explained in terms of weights rather than by nonlinear utility. However, it should also be noted that 71% of individuals tested by Birnbaum and Navarrete (1998) had best-fit estimates of $\hat{\beta}$ less than 1.

For gambles composed of strictly positive, strictly negative, or mixed consequences, different values of δ are allowed in TAX. But a simpler model appears to give a decent first approximation; namely, suppose the same δ can be used for gambles with nonnegative and mixed consequences. For gambles with strictly nonpositive consequences, the same value of δ can be used with absolute values of the consequences, except using reflection to generate the predictions. With these simplifying assumptions (one value of δ instead of three), the model implies the fourfold pattern

of risk seeking and risk aversion, the Allais paradoxes, loss aversion, reflection, and it also describes all 11 new paradoxes reviewed here. Similarly, the model does not require a kink in the utility function for positive and negative consequences to account for these phenomena. Results with mixed gambles (with positive and negative consequences) show that the simplifying assumption, $u(x) = x$, for both positive and negative consequences is not completely accurate (Birnbaum, 2006, 2007b). Nevertheless, these assumptions do give a reasonable first approximation to the results of Birnbaum and Bahra (2007), which refute both versions of prospect theory.

For binary gambles of the form, $G = (x, p; 0, 1 - p)$, with the prior parameters, certainty equivalents in the TAX model are also an inverse-S function of probability to win the larger prize:

$$CE(x, p; 0) = u^{-1}\left[\frac{(1 - \delta/3)p^{\gamma}}{p^{\gamma} + (1 - p)^{\gamma}} u(x) \right]$$

Figure 3 shows predicted certainty equivalents for binary gambles of the form $G = (\$100, p; \$0, 1 - p)$ as a function of the parameters γ and β, with $\delta = 0$. When $\gamma < 1$, the curves have an inverse-S shape, and when $\gamma > 1$, the curves have an S shape. The value of β shifts the curves up or down, influencing risk aversion. These predictions are quite similar to those of CPT, so it should be clear that researchers cannot test very well between TAX and CPT by studying binary gambles.

It is also important to keep in mind that this figure does not represent the probability weighting function in the TAX model, which is $t(p) = p^{\gamma}$. As is shown in several sections below, this distinction is crucial to understanding how TAX predicts choices between three-branch gambles. I show that the interpretation in CPT that this curve represents the decumulative weighting function leads it to wrong predictions. In CPT, this curve is interpreted to mean that people give less weight to consequences near the
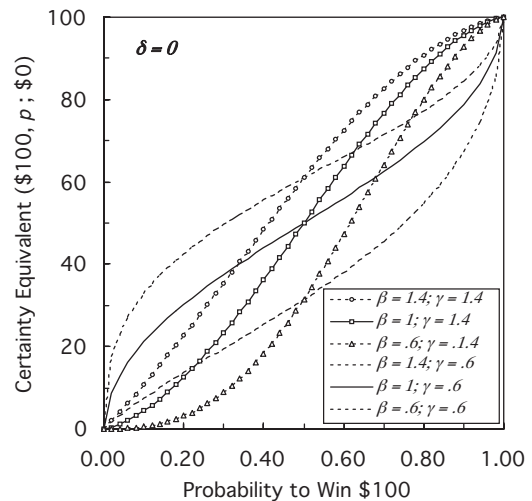

Figure 3. Predicted certainty equivalents in the transfer of attention exchange model for gambles of the form, $G = (\$100, p; \$0)$, as a function of probability to win \$100, with separate curves for different values of the utility parameter, β, and the probability weighting parameter, γ. The configural weight transfer parameter, δ, is fixed to 0. As in cumulative prospect theory, these curves have an inverse-S shape when $\gamma < 1$.

median of the probability distribution than they do to consequences ranked at either extreme of the probability distribution. As is shown in several sections below, when this implication of CPT is tested directly, it is shown to be empirically false.

Figure 4 shows the effects of the configural weight transfer parameter, δ (with the other parameters fixed: $\gamma = 1$ and $\beta = 1$). When $\delta > 0$, weight is transferred from the higher consequence to the lower one, and there is an upper limit on the value of the certainty equivalent as $p \rightarrow 1$, creating a discontinuity. When $\delta < 0$, weight is transferred from the branch with the lower consequence to the branch with the higher consequence, and there is a lower limit on the certainty equivalent as $0 \leftarrow p$. Such a discontinuity was postulated by Birnbaum and Stegner (1979) to account for buying and selling prices as well as earlier results with evaluative and moral judgments (Birnbaum, 1973b; Riskey & Birnbaum, 1974).

For example, a person who has done one very bad deed is rated low in morality, despite doing a large number of good deeds. It appears that the worst deed a person has done sets an upper limit on the highest moral evaluation that person can achieve (Birnbaum, 1973b; Riskey & Birnbaum, 1974). Similarly, for buying prices, any chance that the merchandise is defective sets an upper limit on the price a person is willing to pay. For sellers, any chance that the merchandise is valuable tends to set a lower limit on the amount they will accept. For gambles, it means that as long as it is possible the gamble pays $0, there is an upper limit on the evaluation of that gamble that falls below the value of the highest consequence. A similar discontinuity was assumed in the weighting function of Kahneman and Tversky (1979) to account for the "certainty effect" of Allais.

In practice, the parameters of TAX for the typical undergraduate produce an inverse-S curve with a gap at the upper end. The median estimates reported by Birnbaum and Navarrete (1998) were $\hat{\beta} = 0.41$, $\hat{\gamma} = 0.79$, and $\hat{\delta} = 0.95$. (Because the authors used
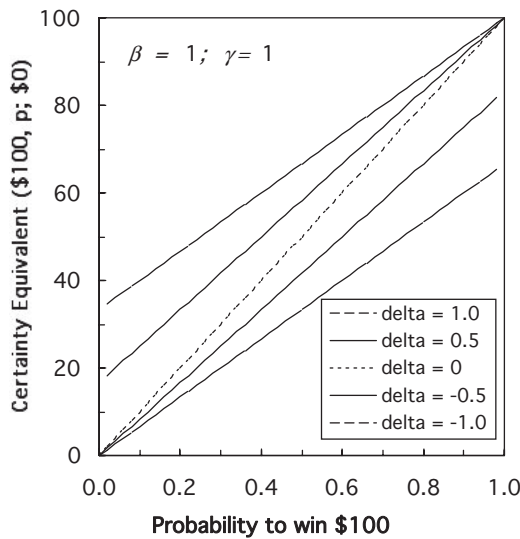
the previous notation convention in their article, the value of $\hat{\delta}$ in that article was reported as -0.95).

Another way to understand the intuitions of the special TAX model is to see that it can be interpreted as a model in which people respond to both a gamble's expectation and also to its spread. For example in binary, gambles, $G = (x, p; y, 1 - p)$, $x \geq y > 0$, $\delta \geq 0$, special TAX simplifies to the following:

$$TAX(G) = \frac{t(p)}{t(p) + t(1-p)} u(x) + \frac{t(1-p)}{t(p) + t(1-p)} u(y)$$
$$- \frac{(\delta/3)t(p)}{t(p) + t(1-p)} |u(x) - u(y)|.$$

Note that this is the sum of a weighted average of utilities (first two terms) and a third term that depends on the spread of the utilities. When $\delta > 0$, people are averse to the spread of consequences (analogous to risk or variance) in the gambles. When $p = 1/2$, this equation further reduces to:

$$TAX(G) = \frac{u(x) + u(y)}{2} - \frac{\delta}{6} |u(x) - u(y)|.$$

Connections between rank-affected configural weighting and this *range* form of the model are further discussed in Birnbaum (1974a, 1982); Birnbaum and Stegner (1979); Birnbaum et al. (1992); and Birnbaum et al. (1971). It can be seen that when $\delta = 0$, the special TAX model reduces to a subjectively weighted average utility model. (The subjectively weighted average utility model satisfies idempotence and therefore does not violate stochastic dominance in the same way as does stripped prospect theory; Birnbaum, 1999b.)

Viscusi's (1989) prospective reference theory (PRT) is also a special case of the TAX model in which $\delta = 0$ and $t(p) = \kappa p + (1 - \kappa)/n$, $0 \leq \kappa \leq 1$. EU theory is also a special case of TAX in which $\delta = 0$ and $t(p) = p$. As is shown in the *Violations of RBI* section, the TAX model produces violations of RBI only when $\delta \neq 0$, but in CPT, these violations are produced when $\gamma \neq 1$. PRT implies no violations of RBI.

The GDU Model

Luce (2000, p. 200) proposed a "less restrictive theory" that satisfies a property known as (lower) gains decomposition but that does not necessarily satisfy coalescing. Marley and Luce (2001) presented a representation theorem for GDU, and Marley and Luce (2005) showed that this model is similar to TAX with respect to many of the new paradoxes. The key idea is that a multibranch gamble can be decomposed into a series of two-branch gambles. The decomposition can be viewed as a tree in which a three-branch gamble is resolved in two stages: first, the chance to win the lowest consequence, and otherwise to win a binary gamble to win one of the two higher prizes. Binary gambles are represented by RDU, as follows:

$$GDU(x, p; y) = W(p)u(x) + [1 - W(p)]u(y), \qquad (10)$$

where $GDU(G)$ is the utility of G according to this model.

For a three-branch gamble, $G = (x, p; y, q; z, 1 - p - q)$, where $x > y > z \geq 0$, the lower gains decomposition rule (Luce, 2000, pp. 200–202) can be written as follows:

Figure 4. Predicted certainty equivalents in the transfer of attention exchange model for gambles of the form, $G = (\$100, p; \$0)$, as a function of probability to win $100, with a separate curve for each value of δ (labeled as "delta"), with $\beta = \gamma = 1$.

$$GDU(G) = W(p + q)*GDU(x, p/(p + q); y)$$
$$+ [1 - W(p + q)]u(z). \quad (11)$$

Note that this utility is decomposed into a gamble either to win the worst outcome, z, or to win a binary gamble between x and y otherwise. Marley and Luce (2001) have shown that RDU is a special case of GDU in which the weights take on a special form.

To illustrate this model, let $u(x) = x^\beta$, and let the weighting function be approximated by the expression developed by Prelec (1998) and by Luce (2000):

$$W(p) = \exp[-\delta(-\ln p)^\gamma]. \quad (12)$$

With these assumptions, this model has a total of three parameters for nonnegative gambles, two for the weighting function and one for the utility function. Luce (2000, pp. 200–202) showed how this model could account for certain phenomena that refute CPT.

Marley and Luce (2001, 2005) showed that lower GDU is similar to TAX in that it violates coalescing and properties derived from coalescing, but it is distinct from TAX. Birnbaum (2005b, 2007b) noted that this model satisfies upper coalescing. Luce, Ng, Marley, and Aczél (in press) have recently developed models that violate both coalescing and idempotence. Such models satisfy gains decomposition but do not necessarily satisfy binary RDU. Such models allow a utility for gambling, apart from winning.

The Case Against Prospect Theories in Choice

Violations of Coalescing: Splitting Effects

Because RDU, RSDU, and CPT models satisfy coalescing and transitivity, they cannot explain "event-splitting" effects. Starmer and Sugden (1993) and Humphrey (1995) found that preferences depend on how branches are split or coalesced. Luce (2000) expressed reservations concerning these tests because they were conducted between groups of participants. Results that are observed between subjects do not always replicate within subjects (Birnbaum, 1999a). In this case, however, subsequent research has determined that event-splitting effects (violations of coalescing combined with transitivity) are robust and can be demonstrated within subjects (Birnbaum, 1999c, 2004a, 2007b; Humphrey, 1998, 2000, 2001a, 2001b).

Consider the Choices 1.1 and 1.2 of Table 1. Each row of Table 1 represents a different choice, which has been embedded among many other choices and presented to a number of participants.

In each choice, a marble will be drawn from an urn, and the color of marble drawn blindly and randomly will determine your prize. You can choose the urn from which the marble will be drawn. Which urn would you choose in Problem 1.1, shown below?

A:	85 red marbles to win \$100	B:	85 black marbles to win \$100
	10 white marbles to win \$50		10 yellow marbles to win \$100
	05 blue marbles to win \$50		05 purple marbles to win \$7

On a different trial, mixed among other choices, participants are asked, "Which urn would you choose in Problem 1.2?"

A':	85 black marbles to win \$100	B':	95 red marbles to win \$100
	15 yellow marbles to win \$50		05 white marbles to win \$7

Note that A' is the same as A, except for coalescing, and B' is also the same prospect as B. So if a person obeys coalescing, then that person should make the same choice between A and B as between A' and B', apart from random "error." Birnbaum (2004a) presented these two choices included among other choices to 200 participants, finding that 63% (significantly more than half) chose B over A, and 80% (significantly more than half) chose A' over B'. Of 200 participants, 96 switched from B to A', but only 13 switched from A to B' ($z = 7.95$).

According to OPT, if people used the editing rule of combination, then they would edit Problem 1.1 into Problem 1.2 before making a choice, so both forms of this choice should yield the same decisions. According to CPT, no one should switch, except by chance, because even without the editing rule of combination, the representation satisfies coalescing.

However, both RAM and TAX, with parameters estimated from previous data, predicted this reversal correctly. In those models, splitting the branch to win \$100 makes B better than B', and splitting the lower branch (to win \$50) makes A worse than A'. Table 1 shows calculated CEs for both TAX and CPT; both models predict the choice correctly in Choice Problem 1.2, but there is no version of CPT (no functions and parameters) that can predict the reversal in Problem 1.1.

The certainty equivalents (TAX) for A' and B' are \$75.7 and \$62.0, respectively; however, the certainty equivalents of A and B are \$68.4 and \$69.7, respectively. So prior TAX predicted this reversal correctly and others like it in Birnbaum (2004a).

According to prior CPT, $A > B$ and $A' > B'$; however, any CPT model implies $A > B \Leftrightarrow A' > B'$. These findings and other data showing violations of coalescing (Birnbaum, 2004a, 2007b) refute CPT and all theories that satisfy coalescing, including those of Lopes and Oden (1999); Becker and Sarin (1987); Chew (1983); and Chew, Epstein, and Segal (1991), among others (see Luce, 1998, 2000).

Violations of First-Order Stochastic Dominance

Birnbaum (1997) deduced that RAM and TAX would violate stochastic dominance when choices were constructed from a special recipe, illustrated in Choice Problem 2 of Table 1. The calculated certainty equivalents of the gambles according to TAX are shown in the right portion of Table 1. According to TAX, the dominant gamble (I, shown on the left) has a certainty equivalent of only \$45.8, and the dominated gamble (J) has a value of \$63.1, violating dominance. Any RDU, RSDU, or CPT theory (with any functions and parameters) must satisfy stochastic dominance.

After this prediction had been set in print in Birnbaum (1997, pp. 93–94), Birnbaum and Navarrete (1998) tested it empirically, finding 73% violations in Problem 2; about 70% of 100 undergraduates violated first-order stochastic dominance in four variations of choices like Problem 2. Birnbaum, Patton, and Lott (1999) found 73% violations with five new variations of this recipe and a new sample of 110 undergraduates. In these studies, significantly more than half the participants violated stochastic dominance.

The development of this example is illustrated in Figure 5. Start with a root gamble, $G_0 = (\$96, .9; \$12, .1)$. Split the lower branch (.1 to win \$12) into two splinters, one of which has a slightly better consequence (.05 to win \$14, and .05 to win \$12), yielding $G^+ = (\$96, .9; \$14, .05; \$12, .05)$. G^+ dominates G_0. However, accord-

Table 1

Choice Problems, Percentages Choosing Gamble on the Right, TAX and CPT Cash Equivalents in Tests of Coalescing (Problems 1.1 and 1.2), Stochastic Dominance (Problems 2, 3.1, 3.2, 3.3, 4), and Upper Tail Independence (Problems 5.1, 5.2, 6.1, 6.2)

Choice problem	Choice		% Choosing second gamble	Prior TAX		Prior CPT	
	First gamble	Second gamble		First	Second	First	Second
1.1	A: 85 to win $100 10 to win $50 05 to win $50	B: 85 to win $100 10 to win $100 05 to win $7	62	68.4	69.7	**82.2**	79.0
1.2	A′: 85 to win $100 15 to win $50	B′: 95 to win $100 05 to win $7	26	75.7	62.0	82.2	79.0
2, 3.1	I: 90 to win $96 05 to win $14 05 to win $12	J: 85 to win $96 05 to win $90 10 to win $12	73	45.8	63.1	**70.3**	69.7
3.2	M: 85 to win $96 05 to win $96 05 to win $14 05 to win $12	N: 85 to win $96 05 to win $90 05 to win $12 05 to win $12	06	53.1	51.4	70.3	69.7
3.3	I′: 90 to win $97 05 to win $15 05 to win $13	J′: 85 to win $90 05 to win $80 10 to win $10	57	46.8	57.6	**73.3**	66.6
4	K: 90 to win $96 05 to win $14 05 to win $12	L: 25 to win $96 05 to win $90 70 to win $12	35	45.8	35.8	72.3	35.0
5.1	s: 43 to win $92 07 to win $68 50 to win $0	t: 48 to win $92 52 to win $0	34	32.3	29.8	33.4	33.2
5.2	u: 43 to win $97 07 to win $68 50 to win $0	v: 43 to win $97 05 to win $92 52 to win $0	62	33.4	36.7	**35.1**	34.9
6.1	w: 80 to win $110 10 to win $44 10 to win $40	x: 80 to win $110 10 to win $96 10 to win $10	67	65.0	69.0	**83.5**	79.9
6.2	y: 80 to win $96 10 to win $44 10 to win $40	z: 90 to win $96 10 to win $10	33	60.3	57.2	75.0	71.4

Note. All choice percentages differ significantly from 50%. Entries in bold show significant errors. TAX = transfer of attention exchange; CPT = cumulative prospect theory.

ing to both TAX and RAM models, G^+ should seem worse than G_O because the increase in total weight of the lower branches outweighs the increase in the .05 sliver's consequence from $12 to $14.

Starting again with G_O, split the higher valued branch of G_O, constructing $G^- = (\$96, .85; \$90, .05; \$12, .10)$, which is dominated by G_O. According to the configural weight models, this split increases the total weight of the higher branches, which improves the gamble despite the decrease in the .05 sliver's consequence from $96 to $90. Both RAM and TAX models, with their prior parameters, predict that people will prefer G^- over G^+, in violation of stochastic dominance.

Transitivity, coalescing, and consequence monotonicity imply satisfaction of stochastic dominance in this recipe: $G_O = (\$96, .9; \$12, .1) \sim (\$96, .9; \$12, .05; \$12, .05)$, by coalescing. By consequence monotonicity, $G^+ = (\$96, .9; \$14, .05; \$12, .05).> (\$96, .9; \$12, .05; \$12, .05)$; $G_O = (\$96, .9; \$12, .1) \sim (\$96, .85; \$96, .05; \$12, .10)$, by coalescing; and $(\$96, .85; \$96, .05; \$12, .10) >$ $(\$96, .85; \$90, .05; \$12, .10) = G^-$, by consequence monotonicity. By transitivity, $G^+ = (\$96, .90; \$14, .05; \$12, .05) > G_O > (\$96, .85; \$90, .05; \$12, .10) = G^-$, so $G^+ > G^-$. This derivation shows that if these three principles hold, then people would not show this violation, except by chance. Systematic violations imply that these assumptions are not descriptive. In general, the recipe is as follows: $G^+ = (x, p; y^+, q; y, 1 - p - q)$ versus $G^- = (x, p - s; x^-, s; y, 1 - p)$, where $x > x^- > y^+ > y \geq 0$, and all of the probabilities are positive. In the case of Choice Problem 2 in Table 1, $p = 0.9$ and $r = s = 0.05$.

Violations of stochastic dominance in judged buying and selling prices. Perhaps violations are due to some comparative process, such as cancellation, that depends on contrasts between features of the gambles rather than on the evaluation of each separate gamble (cf. González-Vallejo, 2002; Leland, 1994). For example, suppose people cancel the branches to win $96 and $12, which have nearly equal probabilities, and choose on the basis of the remainder. To test this notion, Birnbaum and Yeary (2001) obtained judgments of
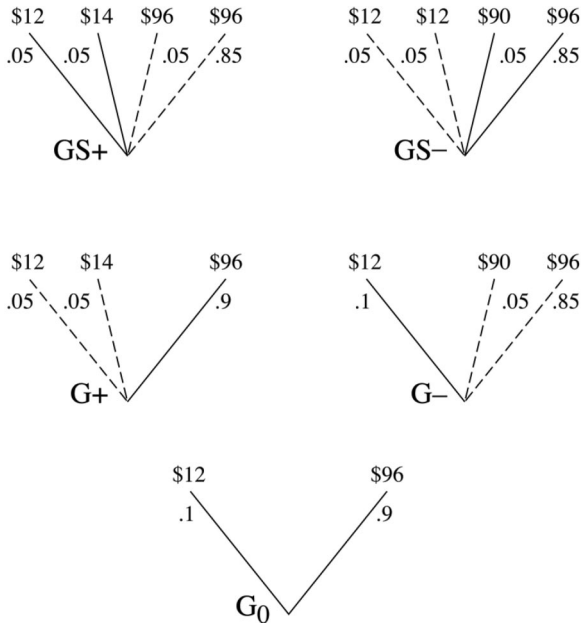
Figure 5. Cultivating and weeding out violations of stochastic dominance. Starting at the root, $G_0 = (\$96, 0.9; \$12, 0.1)$, split the upper branch to create $G^- = (\$96, .85; \$90, .05; \$12, .10)$, which is dominated by G_0. Split the lower branch of G_0, creates $G^+ = (\$96, 0.9; \$14, .05; \$12, .05)$, which dominates G_0. According to configural weight models, G^- is preferred to G^+ because splitting increases the relative weight of the higher or lower branches, respectively. A second round of splitting weeds out violations to low levels in the choice between GS^+ and GS^-.

the "highest buying price" for each of 166 gambles and judgments of the "lowest selling prices" of the same gambles, including eight used by Birnbaum and Navarrete (1998), presented separately. Mixed in among 166 trials in each task were eight gambles in four tests of stochastic dominance in this recipe. Both buying prices and selling prices were significantly higher for dominated gambles like J than for dominant ones like I. For buying prices, people offered an average of $53.52 to buy $G^- = (\$96, .85; \$90, .05; \$12, .10)$, but they offered an average of only $34.52 for the dominant gamble, $G^+ = (\$96, .9; \$14, .05; \$12, .05)$. Similarly, from the seller's viewpoint, people asked for $71.29 to sell the dominated gamble and for $65.49 for the dominant gamble. Violations of stochastic dominance are found in judged values of single gambles, in which a person has no opportunity to cancel nearly equal branches. Therefore, it seems most likely that violations of stochastic dominance are produced in the evaluations of individual gambles, rather than produced by some comparative processes like cancellation that is unique to choice.

Are the violations due to random "error?" In choice studies with replications, it is possible to estimate "random error" rates for each choice. Models of error allow one to estimate the percentage that "truly" violates stochastic dominance and the percentage that does so by "error" alone. In Birnbaum (2004b, Study 3), for example, each of 156 participants completed four choices that were variations of Choice Problem 2 in Table 1, testing stochastic dominance recipe. These four choices were intermixed among a number of other choices. Of the 156 participants, there were 79

who had four violations, 31 with three violations, 21 with two, 13 with one, and 12 with zero violations.

Suppose there are two types of participants: those who truly violate stochastic dominance in this recipe and those who truly satisfy it, apart from error. Then the probability that a person satisfies stochastic dominance on the first three presentations of the choice and violates it on the fourth (*SSSV*) is given by the following:

$$P(SSSV) = a(1 - e)^3 e + (1 - a)e^3(1 - e),$$

where a is the probability that a person "truly" satisfies stochastic dominance, and e is the probability of making an "error" in reporting one's "true" preference. In this case, the person who truly satisfies stochastic dominance has correctly reported her or his preference three times and made one error $((1 - e)^3 e)$, whereas the person who truly violates it has made three "errors" and one correct report. Similar expressions can be written for each of the 15 other response patterns.

When this "true and error" model was fit to observed frequencies of the 16 response combinations in Birnbaum (2004b), it indicated that 83% of participants "truly" violated stochastic dominance (on all four choices, because $a = .17$), except that people made "errors" on 15% of their choices. Thus, these results cannot be attributed to random error. Analyses with other error models (Carbone & Hey, 2000) lead to the same conclusions.

Choices formatted with decumulative probability. (Birnbaum, 2004b, Study 4) presented choices using decumulative probabilities, a procedure that should help people "see" dominance. Examine the following variation of Problem 2:

I: .90 to win $96 or more   *J*: .85 to win $96 or more
  .95 to win $14 or more   .90 to win $90 or more
  1.00 to win $12 or more   1.00 to win $12 or more

Because the definition of stochastic dominance can be given in decumulative form, it was thought that this format might reduce violations. It is easy to see that the probability of getting $96 or more is higher in *I* than in *J*; the probability of getting $90 or more is the same; the probability of getting $14 or more is higher in *I* than in *J*; and the probability of getting $12 or more is the same.

Despite the theory that this display format should make it easier to see dominance, with this decumulative probability format, the true-and-error model estimated that 92% of 445 participants "truly" violated stochastic dominance and that the error rate was 12%. The finding of a higher rate of true violation is surprising because this condition was thought to be one in which detection of stochastic dominance and use of RDU or CPT might be facilitated.

Stochastic dominance in dependent gambles. The displays above represent choices between independent gambles. It seems reasonable that the ticket or marble drawn from one urn would be independent of what would be randomly drawn from a different urn. Perhaps the rate of violation would be different in dependent gambles, in which the state space is the same for both gambles, as in the arrangement used by Savage (1954) and illustrated in Figure 6. Consider a single urn containing 100 tickets numbered from 1 to 100. A ticket drawn from this urn will determine the prize according to a schedule that depends on one's decision, as displayed in Figure 6. Birnbaum (2006) explored this situation with three variations of such dependent gambles, in which the ticket number drawn from a single urn determined the prizes for both gambles. Despite the fact that people should find it easy to perceive domi-

Figure 6. Aligned matrix format showing test of stochastic dominance in coalesced form. About 70% of participants violated stochastic dominance on this problem presented in this format (Birnbaum, 2006).

nance in such displays, 72% violations of stochastic dominance were observed in Choice Problem 3.1 in this format. Rates of violation of stochastic dominance with two other formats for display of such dependent gambles were the same or even higher.

Summary. By 2006, my students and I had completed 41 studies with a total of 11,405 participants testing first-order stochastic dominance in choice using 15 different formats for displaying gambles and choices (Birnbaum, 2004b, 2006). Violation of stochastic dominance with this recipe has been a very robust finding.

These studies confirm that violations of stochastic dominance are observed with or without branch juxtaposition, and when branches are listed in increasing or decreasing order of their consequences. They are observed when probabilities are presented as decimals, as pie charts, as percentages, as natural frequencies, or as lists of equally likely consequences (Birnbaum, 2004b). They are observed with or without the event framing used by Tversky and Kahneman (1986).

Systematic violations of stochastic dominance have been observed among men, women, undergraduates, college graduates, and holders of doctoral degrees (Birnbaum, 1999c). Although the rate of violation declines with increased education, the rate of violation among undergraduates is about 70%, among college graduates it is about 60%, and among doctorates who have studied decision making, the rate is still quite high: about 50%. Majority violations have been observed when undergraduates are tested in class, in the lab, or via the World Wide Web, with consequences that are purely hypothetical or when real chances for prizes are possible (Birnbaum & Martin, 2003).

The same type of violation has been found in three-branch gambles and in five-branch gambles (Birnbaum, 2005a). It has been found with hypothetical prizes in the millions and with chances for real prizes less than $100 (Birnbaum, 2005b, 2007b). It has been found with gambles on gains, loses, and with mixed consequences (Birnbaum, 2006). I conclude that any theory that proposes to be descriptive must reproduce these violations and should account for manipulations that increase or decrease their incidence. Besides RDU, RSDU, and CPT, other descriptive theories that assume or imply stochastic dominance are also refuted by such evidence (e.g., Becker & Sarin, 1987; Lopes & Oden, 1999).

It is important to distinguish first-order stochastic dominance, which must be satisfied by RDU/RSDU/CPT, from other types of "stochastic dominance," such as that discussed by Levy and Levy (2002), Wakker (2003), and Baucells and Heukamp (2006). Although CPT can use its nonlinear weighting function to account for results of Levy and Levy (2002), CPT cannot handle violations of first-order stochastic dominance.

It is also important to distinguish the kinds of violations of stochastic dominance predicted by TAX from those predicted by OPT without its editing principle of dominance detection and other restrictions. The "stripped" version of OPT predicts violations of stochastic dominance of a type that is not predicted by TAX and that has not been observed empirically. For example, with plausible parameters, this version of OPT predicts that people should prefer $H = (\$103, 0.01; \$102, 0.01; \$100, 0.98)$ over $G = (\$110, 0.5; \$105, 0.5)$, even though the worst consequence of G is better than the best consequence of H (Birnbaum, 1999b). Unlike OPT, TAX does not violate stochastic dominance in this way. Because the TAX utility is a weighted average of the utilities of the consequences, the cash equivalent value of a gamble in TAX must fall in the interval between the lowest consequence and highest consequence of the gamble. For the same reason, TAX satisfies idempotence, whereas OPT does not; that is, TAX implies that $x \sim G = (x, p_1; x, p_2; x, p_3; \ldots; x, p_n)$, where $\sum_{i=1}^{n} p_i = 1$, for any splitting of the same consequence. In stripped OPT, however, $x > (x, 0.01; x, 0.99) > (x, 0.5; x, 0.5)$ for $x > 0$. The purpose of the editing rules in OPT was to avoid such implausible predictions (Kahneman, 2003).

Event Splitting and Stochastic Dominance

RAM and TAX models imply that splitting can be used not only to cultivate violations of stochastic dominance but also to weed them out (Birnbaum, 1999c). As shown in the upper portion of Figure 5, one can split the lowest branch of G^-, which makes the split version, GS^-, seem worse, and split the highest branch of G^+, which makes its split version, $GS+$, seem better, so in these split forms, GS^+ seems better than GS^-. As predicted by RAM and TAX models, most people prefer GS^+ over GS^- and G^- over G^+ (Birnbaum, 1999c, 2000, 2001; Birnbaum & Martin, 2003). The choice between GS^+ and GS^- is the same (objectively) as the choice between G^+ and G^-, except the choice is presented in canonical split form. By *canonical split form* of a choice, I mean that both gambles of a choice are split so that the probabilities on

corresponding ranked branches are equal in the two gambles, and the number of branches is minimal.

Consider Problems 3.1 and 3.2 in Table 1. In Birnbaum (2004b), 342 participants were asked to choose between *I* and *J* and between *M* and *N*, which were intermixed among other choices. There were 71% violations of stochastic dominance in the coalesced form (Problem 3.1) and only 5.6% violations in the canonical split form of the same choice (Problem 3.2). It was found that 224 participants (65.5%, significantly more than half) preferred *J* to *I* and *M* to *N*, violating stochastic dominance in Problem 3.1 and satisfying it in Problem 3.2 (choice between GS^- and GS^+). Only 3 participants had the opposite reversal of preferences ($z = 14.3$). Similar reversals refuting any model implying coalescing (including RSDU/RDU/CPT) have been obtained in 25 studies with 7,809 participants (Birnbaum, 1999c, 2000, 2001, 2004a, 2004b, 2006, 2007b; Birnbaum & Martin, 2003).

The variable of *form* of a choice (coalesced as in Problem 3.1 or canonical split as in Problem 3.2) is therefore an extremely powerful device to markedly reduce and nearly eliminate violations of stochastic dominance. It is interesting that coalescing appears different, perceptually, in different types of formats for displaying choices, but it has (in every case tested) had the same effect. For example, in the matrix display of dependent gambles, the split version of Figure 6 differs only in having vertical lines placed to identify consequences for the four branches in each alternative of the split form. Despite the different ways that this variable appears, this variable (form) has had the same type of effect in all of the display formats tested so far, reversing majority violations to small minorities.

Predicting choice proportions and satisfactions of stochastic dominance. A descriptive model should be able to predict when stochastic dominance will or will not be violated. As noted above, branch splitting can be used to reduce 70% violations to 6% violations. This result is consistent with five models: RAM, TAX, GDU, PRT, and a branch-counting heuristic. These models make different predictions for other manipulations. Birnbaum (2005a) conducted a series of five studies manipulating features of the recipe illustrated in Figure 5 to compare the accuracy of these models in predicting satisfaction and violation. These studies also tested heuristic models that (a) people just average the values of the consequences or (b) they count the number of branches with consequences favoring one gamble or the other.

Birnbaum (2005a) tested the counting heuristic, in which people choose the gamble with the greater number of branches with higher consequences. According to this heuristic, violations would be minimal in Choice Problem 3.3 in Table 1.

I':	90 black to win $97	*J'*:	85 red to win $90
	05 yellow to win $15		05 blue to win $80
	05 purple to win $13		10 white to win $10

Here, the dominant gamble (*I'*) has higher consequences on two of three branches, yet 57% of 394 participants (significantly more than half) still choose the dominated gamble, *J'*, instead of the dominant gamble *I'*, as predicted by prior TAX, which has CEs of $57.6 and $46.8, respectively. For CPT, the CEs of *I'* and *J'* satisfy dominance, for any parameters.

Both consequence averaging and contrast counting imply that the majority should violate stochastic dominance in Choice Prob-

lem 4 of Table 1. If TAX and its parameters are correct, however, then the majority should satisfy stochastic dominance in Choice 4.

In order to predict the new choice percentage for Problem 4, the following probabilistic model was used:

$$P(G^-, G^+) = \frac{1}{1 + \exp\{\alpha[TAX(G^+) - TAX(G^-)]\}}, \quad (13)$$

where $P(G^-, G^+)$ is the predicted choice probability for violating stochastic dominance in this choice, and α is the logistic spread parameter that maps the difference in utility into a predicted choice probability. This choice model is similar to models of Thurstone (1927) and Luce (1959, 1994). In order to estimate α, Birnbaum (2005a) simply took the value that makes the predicted choice probability for Problem 2 (same as Problem 3.1) in the "prior" TAX model to be 0.70 to match the approximate 70% violations observed in previous research; this value is $\hat{\alpha} = 0.049$. According to the prior TAX model and Equation 13, Problem 4 should produce 38% violations. Birnbaum (2005a, Study 2) tested a new group of 232 undergraduates in the same context of filler choices to check this new choice and other new predictions.

Of 232 participants, 72% (significantly more than half) violated stochastic dominance in the comparison of *I* and *J* (replicating Problem 3.1), but only 35% of the same people (significantly less than half) violated stochastic dominance in Problem 4. There were 103 who violated stochastic dominance in the choice between *I* and *J and satisfied* it in the choice between *K* and *L*, compared with only 18 who had the opposite reversal ($z = 7.7$), refuting both heuristic models. These obtained rates of violation (72% and 35%) are not far from the predictions made by the TAX model combined with Equation 13 (70% and 38%).

Thus, the majority does not always violate stochastic dominance. People largely satisfy it when the canonical split form is used, as in Problem 3.2, and 65% satisfy it when probability of the highest consequence is reduced sufficiently (Problem 4). The hypothesis that people ignore probability can therefore be rejected. Intermediate probabilities yielded intermediate results, suggesting that people are attending to probability and responding to it as predicted by a quantitative model. Five studies found that TAX gave the best predictions to new data, followed by Viscusi's (1989) prospective reference theory, which is a special case of TAX, followed by RAM. RAM implies a violation of probability monotonicity. Transferring probability from the highest branch to the middle branch in gamble *J* should have made that gamble better according to RAM, a prediction that was not confirmed.

Analysis in TAX. Figure 7 shows an analysis of Choice Problems 3.1(same as Problem 2) and 4 with respect to two parameters of the TAX model, δ and γ. (The exponent of the utility function, β, has very little influence on these predictions; $\beta = 1$ in this analysis.) The region above and to the left of the upper curve in Figure 7 shows where stochastic dominance will be satisfied in both Choice Problems 3.1 and 4, according to TAX. The open circle at the intersection of $\delta = 0$ and $\gamma = 1$ represents EU theory, which is a special case of TAX that satisfies stochastic dominance. The region below the upper curve and above the lower dashed curve shows combinations of parameters for which people should violate stochastic dominance in Choice Problem 3.1 and satisfy it in Choice Problem 4. The solid circle at the intersection of $\delta = 1$ and $\gamma = 0.7$ shows the prediction of the "prior" parameters of
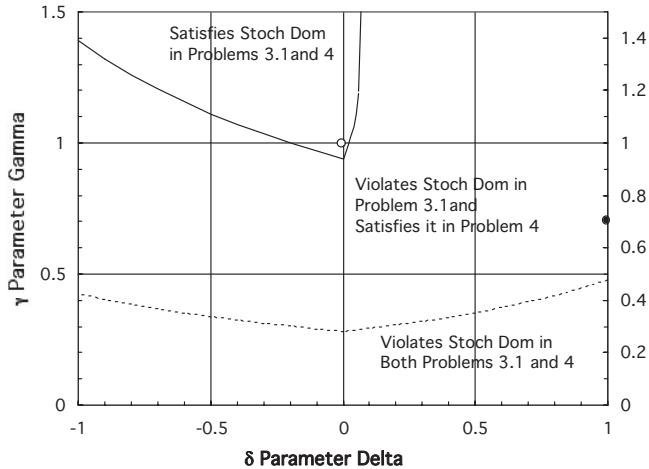
Figure 7. Analysis of violations of stochastic dominance (Stoch Dom) in Choice Problems 3.1 (same as Problem 2) and 4 according to the transfer of attention exchange (TAX) model. Expected utility is a special case of TAX with $\delta = 0$ and $\gamma = 1$ (open circle); it satisfies stochastic dominance. The solid circle shows parameters of TAX used in previous studies ($\delta = 1$ and $\gamma = 0.7$); with these parameters, TAX violates stochastic dominance in Choice Problem 3.1 but satisfies it in Choice Problem 4.

TAX. One interpretation of the 35% who violated stochastic dominance in Choice Problem 4 is that these violations are produced by individuals whose TAX parameters fall below the dashed curve in Figure 7.

For this manipulation (Problem 4) and others studied in Birnbaum (2005a), heuristics and CPT were failures because the heuristics continued to predict violations, and CPT predicts no violations.

Priority heuristic fails. The priority heuristic of Brandstätter et al. (2006) predicts that the majority should satisfy stochastic dominance in Choice Problems 2, 3.1, and 3.3, in which most people violate it, and it fails to predict satisfaction of stochastic dominance in Choice Problem 3.2, in which most people satisfy it. By adding an editing rule of dominance detection, the priority heuristic could be modified to satisfy stochastic dominance in Problem 3.2, but that would still not account for violations of stochastic dominance. Birnbaum (2008a, 2008b) devised a new variation, F = ($89, 0.7; $88, 0.1; $11, 0.2$) versus G = ($90, 0.8; $13, 0.1; $12, 0.1$). This example was constructed so that G is better than F on all four of the variables used by the priority heuristic. Despite the fact that G stochastically dominates F and despite the fact that G is predicted by the priority heuristic to be chosen (because of the smaller probability of the lowest consequence), 71% of 408 undergraduates chose F over G. Even with its use of EV ratio as the first step, the priority heuristic with any order of considering the four dimensions fails to predict systematic violations of this property in this new variation.

Domain of violation of stochastic dominance in TAX. Figure 7 shows that the prediction of violation of stochastic dominance in Problem 3.1 is "robust" in that many combinations of plausible parameters in TAX imply the violations in this special recipe. However, Figure 7 does not imply that people will often violate stochastic dominance. To understand how "often" the TAX model

implies violations of stochastic dominance, Birnbaum (2004a) simulated choices between three-branch gambles. Three "random" numbers, uniformly distributed between 0 and 1, were sampled by computer and divided by their sum to produce three probabilities summing to 1. Next, three consequences were independently drawn from a uniform distribution between $0 and $100. Pairs of such gambles were drawn independently to form choices. In 1,000,000 choices thus simulated, TAX and CPT with their prior parameters made the same predictions in 94% of these cases. Among these million cases, one third of the choices had a stochastic dominance relation, but only 1.8 per 10,000 were predicted violations of stochastic dominance by TAX. This means that TAX rarely violates stochastic dominance in such an environment. An experimenter sampling choices by such a random algorithm would be unlikely to find a choice containing this predicted violation.

One might ask: Is TAX just so flexible that it can account for anything? The answer is no. Given the following two properties, both present in many sets of data (e.g, Gonzalez & Wu, 1999; Tversky & Kahneman, 1992), TAX is forced to violate stochastic dominance in Birnbaum's (1997) original recipe: Assuming $u(x) = x$, if people are (a) risk averse for 50-50 gambles, then $\delta > 0$, and if people are simultaneously (b) risk seeking for positive consequences with small p, then $\gamma < 1$. With $\delta > 0$ and $\gamma < 1$, TAX violates stochastic dominance in Problem 3.1 (same as in Problem 2), as shown in Figure 7. So, TAX had to violate stochastic dominance in this special recipe in order to simultaneously account for typical results with two-branch gambles. As is shown in the *Violations of RBI* section, violations of RBI imply that $\delta \neq 0$ and the Allais paradoxes rule out the assumptions that $\gamma = 1$ and $\delta = 0$. So, findings of Allais paradoxes and violations of RBI also compel TAX to violate stochastic dominance in this recipe.

Equation 13 was fairly accurate in predicting variations of the recipe that produces stochastic dominance in three-branch gambles (Problems like 3.1, 3.3, and 4). TAX and RAM combined with Equation 13 correctly predict that the majority should satisfy dominance in Problem 3.2 with four-branch gambles in canonical split form. However, the difference in utility is small relative to the extremity of the choice proportion. To account for the 6% rate of violations in canonical split form, Equation 13 would require a different value of α to fit this choice proportion. Diederich and Busemeyer (1999) used an analogous interpretation to explain why the rate of violation of stochastic dominance in choices between dependent gambles varies inversely with the correlation between consequences.

Upper Tail Independence

Wu (1994) reported systematic violations of "ordinal" independence. These were not full tests of ordinal independence, as defined in Green and Jullien (1988); however, the property tested by Wu is implied by RDU/RSDU/CPT. It has been called "upper tail independence" because it tests whether the upper tail of a distribution can be manipulated to reverse preferences induced by other aspects of the choice. The property follows from transitivity, upper coalescing, and comonotonic RBI, so it should be satisfied by CPT. Wu found systematic violations, which he reasoned, might be due either to violations of CPT or to editing rules that contradict CPT.

Birnbaum (2001) constructed Problems 5.1 and 5.2 in Table 1 from a similar problem in Wu (1994). Any RDU, RSDU, or CPT model implies that people should prefer t = ($92, 0.48; $0, 0.52) > s = ($92, 0.43; $68, 0.07; $0, 0.5) in Problem 5.1 if and only if they prefer v = ($97, 0.43; $92, .05; $0, 0.52) > u = ($97, 0.43; $68, 0.07; $0, 0.5) in Problem 5.2. There were n = 1,438 people tested via the Web who had chances to win cash prizes. Significantly more than half (66%) preferred s to t, and significantly more than half (62%) preferred v to u, contradicting this property. These results contradict RDU, RSDU, or CPT with any parameters and are predicted by TAX.

Because this implication rests on upper coalescing, violations refute lower GDU as well as CPT. There was a computational rounding error in Luce (2000, pp. 201–202) that made it appear that the lower GDU model could account for violations of upper tail independence.

A new test of upper tail independence (Birnbaum, 2005b, Experiment 2) is displayed in Choice Problems 6.1 and 6.2. Unlike earlier examples, the lowest consequences are not zero in all four gambles. The second choice is created from the first by reducing the consequence on the common branch of 80 marbles from $110 to $96 on both sides, and then coalescing it with the branch of 10 marbles to win $96 on the right side. There were 503 participants. As predicted by TAX, the majority was significantly reversed (67% chose the risky gamble in Problem 6.1, but only 33% chose it in Problem 6.2), contradicting both CPT and GDU.

Upper Cumulative Independence

Violations of RBI found by Birnbaum and McIntosh (1996) contradict the inverse-S-weighting function of CPT needed to account for CEs of binary gambles and for Allais paradoxes (Gonzalez & Wu, 1999; Wu & Gonzalez, 1998). Birnbaum (1997) restated this apparent contradiction between experiments in different labs more precisely in the form of two new, within-person paradoxes that are to the class of RDU/RSDU/CPT models as the Allais paradoxes are to EU. Just as the Allais paradoxes show that there is no utility scale in EU that can reproduce the paradoxical choices, there are no functions and parameters in RDU/RSDU/CPT that allow these models to reproduce violations of upper or of lower cumulative independence.

Lower and upper cumulative independence can be deduced from transitivity, consequence monotonicity, coalescing, and comonotonic RBI. Thus, violations of these properties are paradoxical to any theory that implies these properties, including EU, RDU, RSDU, and CPT, among others. These predictions were set in print (Birnbaum, 1997) before experiments were done. Proofs for the general class of models satisfying coalescing are given in Birnbaum (1997).

Upper cumulative independence can be written as ($z' > x' > x > y > y' > z > 0$):

$$S' = (z', 1 - p - q; x, p; y, q)$$

$$< R' = (z', 1 - p - q; x', p; y', q)$$

$$\Rightarrow S''' = (x', 1 - p - q; y, p + q) < R''' = (x', 1 - q; y', q)$$

This property is implied by CPT/RSDU/RDU. Consider Choice Problems 7.1 and 7.2 in Table 2. In the pies* condition of Birn-

baum (2004b), probability was displayed using pie charts in which slices of the pie had areas proportional to probabilities. Significantly more than half (70% of 305 participants) chose R' = ($110, 0.8; $98, 0.1; $10, 0.1) > S' = ($110, 0.8; $44, 0.1; $40, 0.1) and significantly more than half (58% of the same 305 participants) chose S''' = ($98, 0.8; $40, 0.2) > R''' = ($98, 0.9; $10, 0.1), contrary to RDU/RSDU/CPT: 112 changed preferences in the direction violating the property, and only 28 switched in a manner consistent with the property (z = 7.10). (In Table 2, these results correspond to $N > M$ and $O > P$.)

Such violations of upper cumulative independence can be interpreted as a self-contradiction in the CPT weighting function for these models (Birnbaum et al., 1999, Appendix). For example, for the RDU/RSDU/CPT models, $R' > S' \Leftrightarrow$

$$W(.8)u(110) + [W(.9) - W(.8)]u(98) + [1 - W(.9)]u(10) >$$

$$W(.8)u(110) + [W(.9) - W(.8)]u(44) + [1 - W(.9)]u(40)$$

$$\Leftrightarrow [W(.9) - W(.8)][u(98) - u(44)] >$$

$$[1 - W(.9)][u(40) - u(10)]$$

$$\Leftrightarrow \frac{u(98) - u(44)}{u(40) - u(10)} > \frac{1 - W(.9)}{W(.9) - W(.8)}.$$

Similarly, in this class of models, $R''' < S''' \Leftrightarrow$

$$W(.8)u(98) + [W(.9) - W(.8)]u(98) + [1 - W(.9)]u(10) <$$

$$W(.8)u(98) + [W(.9) - W(.8)]u(40) + [1 - W(.9)]u(10)$$

$$\Leftrightarrow \frac{u(98) - u(44)}{u(40) - u(10)} < \frac{u(98) - u(40)}{u(40) - u(10)} < \frac{1 - W(.9)}{W(.9) - W(.8)}.$$

Thus, this family of CPT models leads to self-contradiction when it attempts to analyze this result because the same ratio of weights cannot be both smaller and larger than the same ratio of differences in utility.

Birnbaum and Navarrete (1998) investigated 27 different tests of this property with 100 participants. Summing across these tests within each person, there were 67 people who showed more reversals of preference in violation of upper cumulative independence than in agreement with it. Only 22 people had more reversals consistent with this property; the remaining 11 people either showed the same number of each type or showed no reversals. Birnbaum et al. (1999) tested six different variations with 110 new participants. Birnbaum (1999c) compared results from 1,224 people recruited from the Web (including highly educated ones) against 124 undergraduates tested in the lab. Birnbaum (2004b) investigated this property with 3,440 participants using a dozen different procedures for displaying gambles. Birnbaum (2006) tested another 663 participants with three different presentation formats involving dependent gambles, with strictly positive, strictly negative, and with mixed consequences. These results were also replicated among "filler trials" in five studies of stochastic dominance (Birnbaum, 2005a) with 1,467 participants. As of this writing, there have been 26 studies with 33 variations of the test and 7,186 participants. Results consistently violate upper cumulative independence: More people reverse preferences in the direction contradicting the property (but consistent with the predictions of prior TAX) than make reversals consistent with the property.

Table 2

Choice Problems Testing Upper (Problems 7.1 and 7.2) and Lower Cumulative Independence (Problems 8.1 and 8.2) and Dissection of Allais Paradox (Problems 10.1–10.5)

Choice problem	Choice		% choosing second gamble	Prior TAX		Prior CPT	
	First gamble	Second gamble		First	Second	First	Second
7.1	M: 80 to win $110 10 to win $44 10 to win $40	N: 80 to win $110 10 to win $98 10 to win $10	70	65.0	69.6	**83.5**	80.1
7.2	O: 80 to win $98 20 to win $40	P: 90 to win $98 10 to win $10	42	68.0	58.3	75.7	72.8
8.1	Q: 05 to win $96 05 to win $12 90 to win $3	R: 05 to win $52 05 to win $48 90 to win $3	62	8.8	10.3	**11.6**	9.5
8.2	S: 05 to win $96 95 to win $12	T: 10 to win $52 90 to win $12	26	18.3	16.7	19.9	17.9
9.1	A: $1M for sure	B: 10 to win $2M 89 to win $1M 01 to win $2	42	1,000 K *1,000 K*	810 K *742 K*	1,000 K	**1065 K**
9.2	C: 11 to win $1M 89 to win $2	D: 10 to win $2M 90 to win $2	76	125 K *75 K*	236 K *138 K*	132 K	248 K
9.3	X: 10 to win $1M 01 to win $1M 89 to win $2	Y: 10 to win $2M 01 to win $2 89 to win $2	37	154 K *97 K*	**172 K** *93 K*	132 K	**248 K**
10.1	E: 10 to win $98 90 to win $2	F: 20 to win $40 80 to win $2	38	13.3	9.0	16.9	10.7
10.2	G: 10 to win $98 10 to win $2 80 to win $2	H: 10 to win $40 10 to win $40 80 to win $2	64	9.6	11.1	**16.9**	10.7
10.3	I: 10 to win $98 80 to win $40 10 to win $2	J: 10 to win $40 80 to win $40 10 to win $40	54	30.6	40.0	38.0	40.0
10.4	K: 80 to win $98 10 to win $98 10 to win $2	L: 80 to win $98 10 to win $40 10 to win $40	43	62.6	59.8	67.6	**74.5**
10.5	M: 90 to win $98 10 to win $2	N: 80 to win $98 20 to win $40	78	54.7	68.0	67.6	74.5

Note. Entries in bold show significant errors. TAX = transfer of attention exchange; CPT = cumulative prospect theory; M = million; K = thousand; italics indicate predictions for previous weighting parameters, $u(x) = x^{0.8}$.

Lower Cumulative Independence

Lower cumulative independence, also deduced by Birnbaum (1997) from coalescing, consequence monotonicity, comonotonic RBI, and transitivity (implications of the class of RSDU/RDU/ CPT models), is the following:

$$S = (x, p; y, q; z, 1 - p - q) > R = (x', p; y', q; z, 1 - p - q)$$

$$\Rightarrow S'' = (x, p + q; y', 1 - p - q) > R'' = (x', p; y', 1 - p).$$

Choice Problems 8.1 and 8.2 in Table 2 show a test of this property. In Birnbaum (2004b, pies*), it was found that 62% (significantly more than half) chose S = ($96, 0.05; $12, 0.05; $3, 0.9) over R = ($52, 0.05; $48, 0.05; $3, 0.9), but only 26% (significantly less than half) chose S″ = ($96, 0.05; $12, 0.95) over R″ = ($52, 0.1; $12, 0.9); in addition, significantly more participants switched in the direction violating the property than in the direction consistent with it. As in the case of upper cumulative

independence, violations of lower cumulative independence create self-contradiction if one assumes any RSDU, RDU, or CPT model. There have been 26 studies with 33 problems and 7,186 participants in a variety of display formats, creating a very strong case against this property and models that imply it. Summing the 27 tests of this property in Birnbaum and Navarrete (1998) for each person, there were 64 people (significantly more than half) who had more reversals of the type refuting lower cumulative independence than consistent with it. Only 7 showed no reversals or had an equal split.

The "Classic" Allais Paradoxes

The chief arguments against EU in Kahneman and Tversky (1979) were variations of "classic" Allais paradoxes (Allais & Hagen, 1979). The *constant ratio paradox* can be illustrated by the choice between F = $3000 for sure and G = ($4000, 0.8; $0, 0.2) and the choice between F′ = ($3000, 0.25; $0, 0.75) and G′ =

($4000, 0.2; $0, 0.8). According to EU, $F > G \Leftrightarrow F' > G'$; however, Kahneman and Tversky (1979) reported that 80% of 95 people chose the sure $F = \$3000$ over G, and only 35% chose $F' > G'$. Both CPT and TAX reproduce this phenomenon; for CPT, the CEs are $3,000, $2,357, $678, and $779 for F, G, F', G', respectively; whereas, for TAX, they are $3,000, $1,934, $633, and $733, respectively. In this case, and for other properties defined on binary gambles and sure things, TAX and CPT make virtually identical predictions.

The common consequence paradox of Allais (1953, 1979) is illustrated with Choices Problems 9.1 and 9.2 in Table 2. From EU, one can deduce that $A > B \Leftrightarrow C > D$. However, many people choose A over B and prefer D over C. This pattern of empirical choices violates EU. However, SWU, PRT, RDU, CPT, RAM, TAX, and GDU can all account for this finding, as can many other models.

A great deal of research on risky decision making has been based on the study of choices that can be represented by means of line segments connecting points inside a triangle, as in Figure 8. The upper left, lower right, and lower left corners of the triangle represent "sure things" to win x, y, or z, respectively. Each point inside the triangle corresponds to a three-branch gamble. Points on the line segments connecting corners represent two-branch gambles. For example, in Choice Problems 9.1 and 9.2, $(x, y, z) =$ ($2M, $1M, $2).

The panels in the left and right of Figure 8 show iso-utility contours according to CPT and TAX with their prior parameters. The similarity of these figures should be obvious. Clearly, research confined to this paradigm is not the way to compare TAX and CPT. Although it might be useful for testing EU theory (Machina, 1987), this restricted paradigm is not useful for testing among nonexpected utility theories such as TAX and CPT.

To understand the effects of the triangle restriction, consider a random experimenter who studies choices between three-branch gambles. This experimenter also restricts attention to gambles
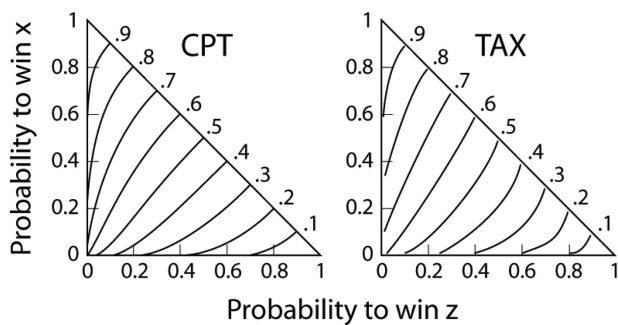

Figure 8. Analysis of studies that limit the number of distinct consequences within a choice to three. For $x > y > z \geq 0$, the ordinate represents the probability to win x, the abscissa represents the probability to win z; otherwise, y is won. Experiments in this representation allow choices between up to three-branch gambles but restrict the number of distinct consequences from six to only three. Curves show iso-utility contours for cumulative prospect theory (CPT; left) and transfer of attention exchange (TAX; right), given their parameters estimated from previous data. Although studies designed in this paradigm test expected utility, they do not allow diagnostic tests among nonexpected utility models such as CPT and TAX.

yielding consequences uniformly distributed between $0 and $100, with probabilities constructed by drawing three random numbers uniformly distributed between 0 and 1 and dividing each by their sum. Birnbaum (2004a) simulated results for random choices in which this experimenter chose three-branch gambles in which all six consequences were free to vary. In that case, TAX and CPT (using their previously estimated parameters) agreed in 94% of these choices. However, when three-branch gambles are constrained to use only three distinct consequences in a choice, CPT and TAX agree in 99% of such restricted choices. Unless the experiment is quite large, this experimenter is unlikely to find a single test in which the models make different predictions. The amount of agreement also depends on the spacing of the consequences; for example, with (x, y, z) fixed to ($100, $20, $0), prior TAX and CPT agree in 99.5% of such "random" choices.

Therefore, studies of classic paradoxes "trapped inside the triangle" (e.g., as analyzed by Machina, 1982, 1987; Harless & Camerer, 1994; Hey, 2005; Hey & Orme, 1994; Wakker, 2001; Wu & Gonzalez, 1998) cannot be used to strongly test among nonexpected utility theories like TAX and CPT. Such tests of the "sure thing" principle confound their tests of independence with the property of coalescing.

Instead, researchers need to think outside the triangle and tease apart variables that are typically confounded in this research (Humphrey, 2001a). Birnbaum's (1997) recipe for violations of stochastic dominance in three-branch gambles cannot be found inside the triangle, for example, because it requires at least four distinct levels of the consequences. Nor can the general statement of RBI be found inside the triangle because it is defined on six distinct consequences. Nor does the triangle provide a good way to represent coalescing and splitting; instead, this needlessly restricted space invites experimenters to study manipulations of probability that do not distinguish the theories.

Figure 8 may be a useful device for conceptualizing violations of EU, but one needs to break out of this paradigm to compare theories like CPT and TAX. To unconfound the Allais paradox, researchers can test RBI with six distinct consequences (as in Expression 3) and can separate those tests of coalescing from the tests of independence.

Decomposition of the Allais Common Consequence Paradox

In order to test among the models, dissect the Allais paradox into transitivity, coalescing, and RBI, as illustrated below:

$A = (\$1M, 1) > B = (\$2M, 0.1; \$1M, 0.89; \$2, 0.01)$

$\Leftrightarrow$ (coalescing & transitivity)

$A' = (\$1M, 0.1; \$1M, 0.89; \$1M, 0.01) >$

$B = (\$2M, 0.1; \$1M, 0.89; \$2, 0.01)$

$\Leftrightarrow$ (restricted branch independence)

$A'' = (\$1M, 0.1; \$2, 0.89; \$1M, 0.01) >$

$B' = (\$2M, 0.1; \$2, 0.89; \$2, 0.01)$

$\Leftrightarrow$ (coalescing & transitivity)

$$C = (\$1M, 0.11; \$2, 0.89) > D = (\$2M, 0.1; \$2, 0.90)$$

The first step converts A to its split form, A'; $A \sim A'$ by coalescing; by transitivity, $A' > B$. In the third step, the consequence on the common branch (.89 to win \$1,000,000) has been changed to \$0 on both sides, so by RBI, $A'' > B'$. By coalescing branches with identical consequences, we see that $C > D$. So if people obeyed these three principles (coalescing, transitivity, and RBI), then there would be no paradox, except by chance. Instead, in Choice Problem 9.1, most people choose $A > B$, in Choice Problem 9.2, most people choose $D > C$. Such violations show that at least one of these three principles is not descriptive.

Different theories attribute Allais paradoxes to different causes (Birnbaum, 1999b), as listed in Table 3. SWU and "stripped" OPT, RAM, TAX, and GDU attribute them to violations of coalescing. In contrast, the class of RDU, RSDU, and CPT attribute the paradox to violations of RBI. OPT uses an additive representation that violates coalescing and implies branch independence, and it uses the editing principle of cancellation, which also implies branch independence; however, OPT is difficult to place in the table because it also has the editing principle of combination, which implies coalescing. But if OPT assumes both editing principles, then it cannot show the Allais paradox, whereas Kahneman and Tversky (1979) presented the theory as a theory of Allais paradoxes. To handle these contradictory multiple predictions, OPT (with or without combination) is permitted to occupy two of the four cells in Table 3. Similarly, the equations of CPT imply coalescing and violations of RBI, but the editing principle of cancellation would imply RBI, so CPT (with or without cancellation) is also allowed to occupy two of the four cells in Table 3.

Birnbaum (2004a) tested among the theories in Table 3 by using tests like those in Choice Problems 10.1–10.5 in Table 2, which teases out branch independence from coalescing. According to EU, people should make the same choice in all five choices because all of the choices are the same, except for a common branch (80 marbles to win \$2 in Choice Problems 10.1 and 10.2, 80 marbles to win \$40 in Choice Problem 10.3, or 80 marbles to win \$98 in Choice Problems 10.4 and 10.5). These five choices were presented, intermixed among other choices, to 349 participants whose data are summarized in Table

Table 3
Comparison of Decision Theories by Two Properties

	Branch independence	
Coalescing	Satisfied	Violated
Satisfied	EU (CPT*/OPT*)	RDU/RSDU/CPT*/SPA
Violated	SWU/PRT/OPT*	RAM/TAX/GDU

Note. Prospect theories make different predictions with and without their editing rules. The editing rule of combination implies coalescing, and cancellation implies branch independence. With or without the editing rule of combination, cumulative prospect theory (CPT) satisfies coalescing. The rank-affected multiplicative (RAM) and transfer of attention exchange (TAX) models violate both branch independence and coalescing. EU = expected utility theory; OPT = original prospect theory; PRT = prospective reference theory; RDU = rank-dependent utility; RSDU = rank- and sign-dependent utility theory. SPA = security potential, aspiration level; SWU = subjectively weighted utility; GDU = gains decomposition utility.

2. Each percentage, except for that in Choice Problem 10.3, is significantly different from 50%. By tests of correlated proportions, each successive contrast between rows is also significant, as is the difference between Choice Problem 10.2 and 10.4.

CPT, RSDU, SWU, GDU, TAX, and RAM models, as fit to previous data, predict the empirical modal choices in Choice Problems 10.1, 10.3, and 10.5 of Table 2 correctly. That is, all models (except EU and prospect theories with both cancellation and combination) predict the classic versions of Allais paradoxes in these three problems. However, RDU, RSDU, or CPT imply that Choice Problems 10.1 and 10.2 should be the same, except for error, and that Choice Problems 10.4 and 10.5 should agree as well because these differ only by coalescing. SWU (including stripped OPT and PRT) implies that the choices can differ in Choice Problem 10.1 and 10.2, but people should make the same choices in Choice Problem 10.2 and 10.4 because these differ only by RBI. RAM, TAX, and GDU models predict reversals from Choice Problem 10.1 to 10.2, from Choice Problem 10.2 to 10.4, and from Choice Problem 10.4 to 10.5. Indeed, results show that all reversals predicted by these models are significant.

Suppose people used the editing rule of combination. If so, then they would make the same decision in both Choice Problem 10.1 and 10.2 and the same decision in Choice Problem 10.4 and Choice 10.5. So, the editing principle of combination can be refuted.

Suppose people used the editing rule of cancellation. If so, then they would make the same decisions in Choice Problems 10.2, 10.3, and 10.4. But these differ, so the editing rule of cancellation and the assumption of RBI can be refuted.

Now suppose that people only cancelled on some proportion of the trials, with CPT governing the rest; if so, then there would be (weaker) violations of branch independence in the same direction between Choice Problems 10.2 and 10.4 as between Choice Problems 10.1 and 10.5. Instead, observed violations are substantial and *opposite* the direction needed by CPT to account for Allais paradoxes. So, these results (and others in Birnbaum, 2004a, 2007b) refute both OPT and CPT, with or without their editing rules of cancellation and combination. Put another way, the results fall in the lower right cell in Table 3, the one cell in which neither version of prospect theory can lay a claim.

TAX and RAM correctly predicted (before the experiments were begun) all of the modal choices in which the choice proportions significantly differed from 0.5 in Birnbaum (2004a). According to TAX or RAM, the Allais constant consequence paradoxes are due to violation of coalescing, and violations of branch independence actually reduce their magnitude. Note that all splitting and coalescing operations make the "risky" gamble (left side) worse and the "safe" gamble (on the right side) better as we proceed from Choice Problem 10.1 to 10.5 in Table 2. Splitting the lower branch on the left side from Choice Problem 10.1 to 10.2 makes the risky gamble worse, and coalescing the upper branches on the left side from Choice Problem 10.4 to 10.5 makes the risky gamble worse again. Similarly, splitting the higher branch in the safe gamble (right side) from Choice Problem 10.1 to 10.2 improves this safe gamble as does coalescing the lower branches on the right side from Choice Problem 10.4 to 10.5.

Birnbaum (2007b) further refined the experiment to investigate upper and lower coalescing in safe and risky gambles separately. There were significant violations of upper coalescing, which vio-

late the idempotent, lower GDU model. Violations of lower coalescing in nonnegative gambles were smaller but statistically significant. Birnbaum (2007b) found violations of both upper and lower coalescing in mixed gambles.

Analysis of Coalescing in Allais Paradox

To fit group data with large consequences (Birnbaum, 2007b, including Choice Problems 9.2 and 9.3), TAX required a nonlinear utility function, $u(x) = x^{0.8}$, along with its previous weighting parameters (predictions shown in italics in Table 2).

Table 4 shows the number of participants (out of 200) who showed each preference pattern on two replications of Allais paradox (Choice Problems 9.2 and 9.3), which differ only by coalescing (Birnbaum, 2007b). In Table 4, S indicates preference for the "safe" gamble (11 marbles to win $1,000,000 and 89 to win $2) rather than the "risky" (R) gamble (10 marbles to win $2,000,000 and 89 to win $2). Here $SRSS$, for example, refers to the pattern choosing S on the first replicate of Choice Problem 9.2, in which branches were coalesced, R on Choice Problem 9.3, in which branches were split, and S on both of these choices in the second replicate.

A five-parameter "true and error" model was fit to these 16 response frequencies (as in Birnbaum, 2004b). The parameters are the "true" probabilities that a person has the preference patterns RR', RS', SR', and SS' in Choice Problems 9.1 and 9.2, and the probabilities of making a random "error" on Choice Problems 9.2 and 9.3, respectively. For these problems (Series A; see Table 4),

Table 4

Numbers of Participants (out of 200) Showing Each Choice Combination in Tests of Coalescing in Problems 9.2 and 9.3 of Table 2

Choice pattern	Series A		Series B	
	Observed	Fitted	Observed	Fitted
RR' RR'	31	29.0	19	19.4
RR' RS'	17	20.7	11	13.6
RR' SR'	4	5.1	5	4.2
RR' SS'	7	3.9	6	3.5
RS' RR'	25	20.7	18	13.6
RS' RS'	56	58.4	62	63.7
RS' SR'	3	3.9	7	3.5
RS' SS'	9	11.2	17	17.4
SR' RR'	3	5.1	3	4.2
SR' RS'	3	3.9	2	3.5
SR' SR'	4	3.1	1	1.4
SR' SS'	1	3.6	1	4.0
SS' RR'	5	3.9	3	3.5
SS' RS'	14	11.2	15	17.4
SS' SR'	3	3.6	2	4.0
SS' SS'	14	11.9	28	23.1

Note. R = chose "risky" gamble in Problem 9.2; S = chose "safe" gamble in Problem 9.2; R' = chose "risky" gamble in Problem 9.3; S' = chose "safe" gamble in Problem 9.3. The pattern RR' RS', for example, represents choice of the "risky" gamble (listed as second gamble in Table 2) on the first replicate of 9.2 and 9.3, and choice of "risky" gamble in the second replicate of 9.2, but choice of the "safe" gamble (first in Table 2) on the second replicate of Problem 9.3. Series B used different gambles (Birnbaum, 2007b). The true-and-error model has been fit to the observed frequencies, the "fitted" values are predictions of that model.

estimates are 0.27, 0.60, 0.02, and 0.11, respectively; $\hat{e}_{9.1} = 0.14$, and $\hat{e}_{9.2} = 0.20. = 0.20$. This model provides an acceptable fit to the observed data, $\chi^2(10, N = 200) = 10.2$, shown by the similarity of the observed and fitted frequencies in Table 4.

Put another way, this model indicates that 60% of the participants truly switched from R in Choice Problem 9.2 to S' in Choice Problem 9.3 and that 2% switched in the other direction. Table 4 shows that 56 people (28%) switched in this fashion on both replicates, compared with 4 who made the opposite switch both times. A three-parameter model assuming no reversals fit much worse, $\chi^2(2, N = 200) = 141.1$, so these data refute coalescing. These results (and others, as in Series B, which showed that 62% had the RS' pattern) contradict any theory that assumes coalescing, including CPT/RSDU/RDU. These results agree with RAM, TAX, and GDU.

In summary, although many models can account for the basic Allais paradoxes, neither OPT nor CPT with or without their editing principles of cancellation and combination can account for the dissection of the Allais paradoxes. Similarly, the next test also shows evidence against both versions of prospect theory.

Violations of Gain–Loss Separability

According to CPT/RSDU or OPT, the overall utility of a mixed gamble is the sum of two terms, one for the gain part and another for the loss part of the gamble. These models therefore imply *gain–loss separability*, which can be expressed for three-branch gambles as follows:

If $B^+ = (y, p; x, q; 0, 1 - p - q) >$
$$A^+ = (y', p'; x', q'; 0, 1 - p' - q')$$
and if $B^- = (0, 1 - r - s; z, r; v, s) >$
$$A^- = (0, 1 - r' - s'; z', r'; v', s'),$$
then $B = (y, p; x, q; z, r; v, s) >$
$$A = (y', p'; x', q'; z', r'; v', s'),$$

where x, y, x', y' are gains and z, v, z', v' are losses; $r + s = 1 - p - q$ and $r' + s' = 1 - p' - q'$. Intuitively, if you prefer the good part of B to the good part of A, and if you prefer the bad part of B to the bad part of A, then you should prefer B to A.

Wu and Markle (2005) devised a test from a choice in Levy and Levy (2002, Experiment 2), shown in Choice Problems 12.1, 12.2, and 12.3 in Table 5.

Wu and Markle (2005) found that the majority prefers $B^+ > A^+$ and $B^- > A^-$; however, the majority does not prefer $B > A$, contrary to CPT/RSDU and any model that satisfies gain–loss separability, including OPT. This example is consistent with the TAX model, with the simplifications listed in the note of Table 5, including the use of just one δ and $u(x) = x$. This point deserves emphasis: This model describes "loss aversion" by greater weight assigned to branches with negative consequences, and it can fit these violations of gain–loss separability without assuming that "loss aversion" has any effect on utility.

In the model of Birnbaum (1997), branch weights depend not only on probability and rank but also on the augmented sign of the consequences. That is, a branch's weight in that model depends on whether the branch leads to a positive, zero, or negative consequence, and it also depends on the rank. That model uses more parameters than the simple version of TAX used here to fit the Wu

Table 5
Tests of Gain–Loss Separability

Choice problem	Choice		% Choosing second gamble	Prior TAX		Prior CPT	
	First gamble	Second gamble		First	Second	First	Second
12.1	A^+: 25 to win $2,000 25 to win $800 50 to win $0	B^+: 25 to win $1,600 25 to win $1,200 50 to win $0	72	497	552	**601**	551
12.2	A^-: 50 to win $0 25 to lose $800 25 to lose $1,000	B^-: 50 to lose $0 25 to lose $200 25 to lose $1,600	72	−358	−276	**−379**	−437
12.3	A: 25 to win $2,000 25 to win $800 25 to lose $800 25 to lose $1,000	B: 25 to win $1,600 25 to win $1,200 25 to lose $200 25 to lose $1,600	38	−280	−300	−107	−179
12.4	F^+: 25 to win $100 25 to win $0 50 to win $0	G^+: 25 to win $50 25 to win $50 50 to win $0	71	14	21	**25**	19
12.5	F^-: 50 to lose $0 25 to lose $50 25 to lose $50	G^-: 50 to lose $0 25 to lose $0 25 to lose $100	65	−21	−14	**−20**	−25
12.6	F: 25 to win $100 25 to win $0 25 to lose $50 25 to lose $50	G: 25 to win $50 25 to win $50 25 to lose $0 25 to lose $100	52	−25	−25	−9	−15
12.7	F': 25 to win $100 25 to win $0 50 to lose $50	G': 50 to win $50 25 to lose $0 25 to lose $100	24	−16	−34	−9	−15
12.8	H: 25 to win $100 25 to win $0 25 to lose $0 25 to lose $100	I: 25 to win $50 25 to win $50 25 to lose $50 25 to lose $50	57	−30	−20	−13	−11

Note. Prior transfer of attention exchange (TAX) here assumes $u(x) = x$, $-\$100 \leq x \leq \$100$, $\gamma^+ = \gamma^- = \gamma^\pm = 0.7$; $\delta^+ = \delta^- = \delta^\pm = 1$. This model is undoubtedly oversimplified, but it correctly predicts all eight modal choices, including the case of indifference in Problem 12.6. CPT = cumulative prospect theory. Bold entries show incorrect predictions of the modal choice.

and Markle (2005) result; however, the more complicated model was not needed for this result.

A new series of tests of gain–loss separability by Birnbaum and Bahra (2007) found data compatible with this (probably oversimplified) TAX model in a direct test, shown in Choice Problems 12.4–12.8. According to the TAX model, with simplifying assumptions in Table 5, people should be indifferent in Choice Problem 12.6. The observed choice proportion was 0.52, not significantly different from the predicted value of 0.5.

Wu and Markle (2005) also reported violations of gain–loss separability in binary gambles, which (if they can be replicated) refute OPT as well as CPT. They noted that the case of mixed gambles has been investigated in only a few empirical studies, yet the topic of "loss aversion" has been treated in many theoretical articles on the basis of the assumption of gain–loss separability.

In both versions of prospect theory, the utility function shows a kink at zero, consistent with the idea that "losses loom larger than gains." This is represented in CPT by the equation, $u(-x) = -\lambda u(x)$, $x \geq 0$, where λ, estimated to be about 2.25, is the "loss looming" factor. This utility function depends on gain–loss separability. If these separable theories are false, then utility curves derived from those theories become meaningless.

To account for the data, Birnbaum and Bahra (2007) did not need to postulate a kinked utility function. Do their findings

disprove the kinked utility curve of prospect theory? No, because that conclusion would require a proof of the null hypothesis. What Wu and Markle (2005) and Birnbaum and Bahra showed is that gain–loss separability can be rejected, which shows that the argument given for the kinked utility function is false. Birnbaum and Bahra showed that configural weighting can account for the behavioral phenomena of loss aversion and violation of gain–loss separability without assuming a kinked utility function. However, it should also be kept in mind that TAX does not rule out nonlinear or kinked utility functions.

Violations of RBI

Restricted branch independence should be satisfied by EU, SWU, PRT, OPT, and by any other theory that includes the editing principle of cancellation. However, this property should be violated according to RDU/RSDU/CPT (aside from the editing rule of cancellation) and by RAM, TAX, and GDU. The direction of violations, however, is opposite in the two groups of models. Wakker et al. (1994) failed to find systematic violations predicted by CPT, but their study was not designed to test RAM or TAX. As shown by Birnbaum and McIntosh (1996), violations of this property are not easy to find unless researchers know the parameters in

advance or use special experimental designs that allow for individual differences.

A number of experiments using the Birnbaum and McIntosh (1996) design subsequently reported systematic violations of a special case of RBI (Birnbaum & Chavez, 1997; Birnbaum & Navarrete, 1998; Birnbaum & Veira, 1998). This special case can be written as follows: $(0 < z < y' < y < x < x' < z')$:

$$S = (x, p; y, p; z, 1 - 2p) > R = (x', p; y', p; z, 1 - 2p)$$

$$\Leftrightarrow$$

$$S' = (z', 1 - 2p; x, p; y, p) > R' = (z', 1 - 2p; x', p; y', p).$$

In this special case, two branches have equal probabilities, and only the value and rank of the common branch, $(z, 1 - 2p)$ or $(z', 1 - 2p)$, changes (from z smallest in the first choice to z' highest in the second).

In this paradigm, RSDU, RDU, CPT, RAM, TAX, and GDU models reduce to what Birnbaum and McIntosh (1996) called the "generic rank-dependent configural weight" (p. 92) model, also known as the rank-weighted utility model (Luce, 2000; Marley & Luce, 2001, 2005). This generic model can be written for the choice between S and R as follows:

$$S > R$$

$$\Leftrightarrow$$

$$w_1 u(x) + w_2 u(y) + w_3 u(z) > w_1 u(x') + w_2 u(y') + w_3 u(z),$$

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

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

There will be an SR' violation of RBI (i.e., $S > R$ and $R' > S'$) if and only if the following:

$$\frac{w_2}{w_1} > \frac{u(x') - u(x)}{u(y) - u(y')} > \frac{w'_3}{w'_2},$$

where w_1 and w_2 are the weights of the highest and middle branches, respectively (when both have probability p in S and R), and w'_2 and w'_3 are the weights of the middle and lowest branches (with probability p) in S' and R', respectively. RAM, TAX, and GDU models imply this type of violation, that is, SR'. With prior parameters, the weight ratios are 2/1 > 3/2 in both RAM and TAX.

CPT also systematically violates RBI; however, CPT with its inverse-S-weighting function violates it in the opposite way from that of RAM and TAX. The $W(P)$ function estimated by Tversky and Kahneman (1992) is shown as the solid curve in Figure 9. Note that this function has an inverse-S shape (steeper near 0 and 1 than in the middle) and that it crosses the identity line. Define a *weakly inverse-S function* as any strictly increasing monotonic function from zero to one satisfying the following, for all $p < p^*$: $W(2p) - W(p) < W(p) - W(0)$ and $W(1 - p) - W(1 - 2p) < W(1) - W(1 - p)$. A *strongly inverse-S function* is one that is weakly inverse
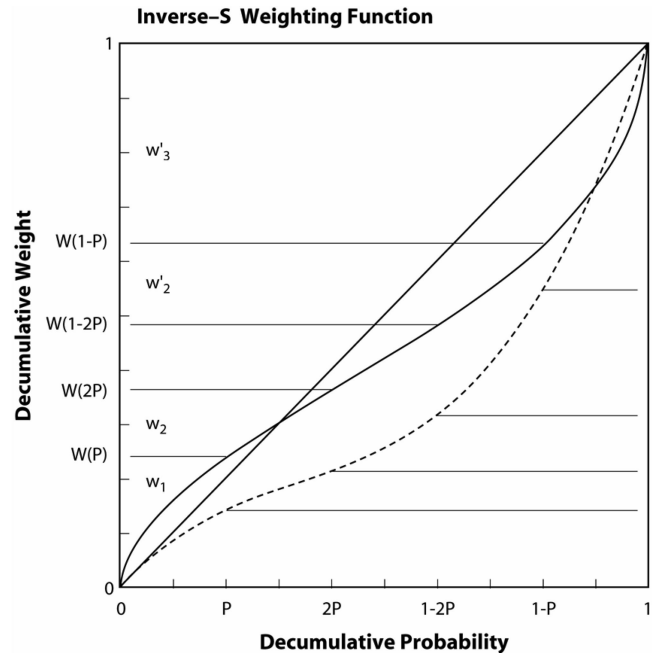

Figure 9. The solid curve is strongly inverse S, and the dashed curve is weakly inverse S. In cumulative prospect theory, the weight of the highest branch in $(x, p; y, p; z, 1 - 2p)$ is greater than the weight of the middle branch; $w_1 > w_2$; that is, $1 > w_2/w_1$. Similarly, the weight of the middle branch in $(z', 1 - 2p; x, p; y, p)$ is less than that of the lowest branch; therefore, $w'_2 < w'_3 \Rightarrow 1 < w'_3/w'_2$. Together, these two conditions imply that violations of branch independence can only be of the form $S < R$ and $S' > R'$; that is, the RS' pattern of violations follows from either type of inverse-S function.

S and also crosses the identity line (i.e., for all $p < p''$, $W(p) > p$, and for all $p > p''$, $W(p) < p$). If the weakly inverse S is rejected, then the stronger version is also rejected. Such functions are illustrated by the dashed and solid curves in Figure 9, respectively. In any inverse-S function, therefore, $w_2 < w_1$ and $w'_2 < w'_3$. It follows that:

$$\frac{w_2}{w_1} < 1 < \frac{w'_3}{w'_2}.$$

Therefore, CPT with such a function (weakly or strongly inverse S) implies that violations of RBI, if they are observed, should have the opposite ordering from that predicted by RAM and TAX; that is, CPT implies $R > S$ and $R' < S'$, called the RS' pattern.

Consider Choice Problems 13.1 and 13.2 in Table 6 (from Birnbaum & Chavez, 1997). Intuitively, CPT predicts $R > S$ in Choice Problem 13.1 because the lowest branches are the same and the highest branch gets more weight than the middle branch. In Choice Problem 13.2, however, CPT implies $R' < S'$ because the upper branches are the same and the lowest branch gets more weight than the middle branch. Contrary to this prediction, however, significantly more than half the participants chose $S > R$ in Choice Problem 13.1, and significantly more than half chose $R' > S'$ in Choice Problem 13.2. These violations are significant and opposite of the predictions of CPT with its inverse-S-weighting function, which is required by CPT to handle the standard Allais

Table 6
Tests of Restricted Branch Independence (Choice Problems 13.1 and 13.2) and 4-Distribution Independence (Problems 14.1 and 14.2)

Choice problem	Choice		% Choosing risky gamble	Prior TAX		Prior CPT	
	Safe	Risky		Safe	Risky	Safe	Risky
13.1	S: 25 to win $44 25 to win $40 50 to win $5	R: 25 to win $98 25 to win $10 50 to win $5	40	20.0	19.2	19.7	**28.1**
13.2	S': 50 to win $111 25 to win $44 25 to win $40	R': 50 to win $111 25 to win $98 25 to win $10	62	57.2	60.7	**69.5**	64.3
14.1	S: 01 to win $110 20 to win $49 20 to win $45 59 to win $4	R: 01 to win $110 20 to win $97 20 to win $11 59 to win $4	34	21.7	20.6	20.9	**25.1**
14.2	S': 59 to win $110 20 to win $49 20 to win $45 01 to win $4	R': 59 to win $110 20 to win $97 20 to win $11 01 to win $4	51	49.8	50.0	**71.9**	67.2

Note. Data from Birnbaum and Chavez (1997, $n = 100$). Entries in bold designate cases in which a model fails to predict the modal choice. TAX = transfer of attention exchange; CPT = cumulative prospect theory. Entries in bold show incorrect predictions of modal choices.

paradoxes (Wu & Gonzalez, 1998). In all 12 tests in Birnbaum and Chavez (1997), more people showed the SR' pattern of reversals than showed the RS' pattern.

Figure 10 shows an analysis of Choice Problems 13.2 and 13.3 according to the parameterized CPT model. CPT is flexible enough to handle any combination of preferences, including either pattern of violation, depending on the value of γ. When $\gamma < 1$, the decumulative weighting function has the inverse-S shape, and CPT implies the RS' pattern, and when $\gamma > 1$, it has an S shape, in which case CPT implies the SR' pattern. When $\gamma = 1$, CPT implies no violations of RBI: Depending on the utility function,

people should either prefer the "safe" gamble in both cases (see SS' in Figure 10) or prefer the risky gamble in both cases (RR').

Figure 11 shows an analysis of Choice Problems 13.1 and 13.2 in the special TAX model. With the configural weight transfer parameter, $\delta = 0$, there can be no violations of RBI. However, when δ is either positive or negative, special TAX implies the SR' pattern of violation of RBI. This pattern is opposite that predicted by CPT with the inverse-S-weighting function.

Empirically, there are significantly more reversals of the type SR' than of the opposite (Birnbaum & McIntosh, 1996; Birnbaum & Navarrete, 1998; Weber & Kirsner, 1997). This same pattern of
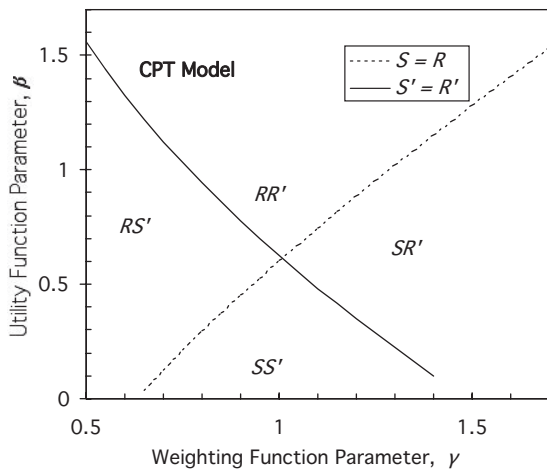

Figure 10. Analysis of Choice Problems 13.1 and 13.2 (from Table 6) according to parameterized cumulative prospect theory (CPT) model. When $\gamma < 1$, the decumulative weighting function is inverse S, and the only type of violation possible is RS'; that is, $R > S$ and $S' > R'$. Instead, data show that the opposite pattern, SR', is significantly more frequent, which contradicts the inverse-S-weighting function required by CPT to account for Allais paradoxes.
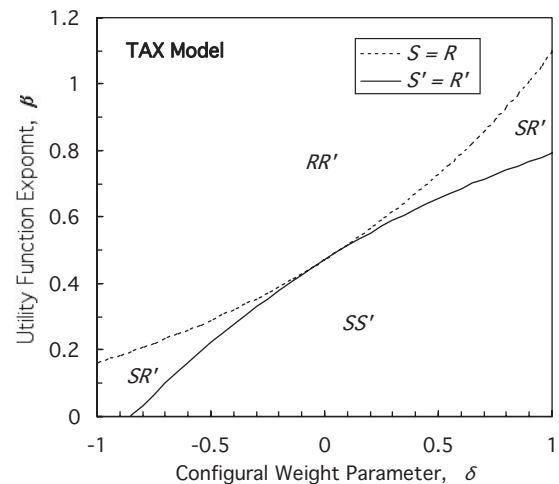

Figure 11. Analysis of Choice Problems 13.1 and 13.2 according to the special transfer of attention exchange (TAX) model. This model allows only one type of violation of RBI, SR', which can occur only when the configural weight parameter, δ, is not zero. The "prior" parameters ($\beta = 1$ and $\delta = 1$) predict this pattern for Problems 13.1 and 13.2, as do many other combinations of parameters (e.g., $\beta = 0.6$ and $\delta = 0.3$).

violations has been prevalent when $p = 1/3$ (Birnbaum & McIntosh, 1996), $p = .25$ (Birnbaum & Chavez, 1997; Birnbaum & Navarrete, 1998), $p = .2$ (Birnbaum et al., 1999), $p = .1$ (Birnbaum, 1999c, 2004b, 2005a, 2006, 2007b; Birnbaum & Navarrete, 1998), and $p = .05$ (Birnbaum, 1999c, 2004b).

If people used the editing principle of cancellation, then they would not violate RBI. In research on this property, it has been assumed that if many choices in a study would allow such a cancellation, then participants might learn to use this editing principle in order to simplify the decisions. For that reason, research on this topic used many other filler trials in which probabilities and consequences differed. The crucial tests were then interspersed among these other trials (Birnbaum & Chavez, 1997; Birnbaum & McIntosh, 1996). Is it possible, however, that some people use cancellation some of the time? If so, then the observed rate of violation is smaller than it would be without the partial use of this editing rule.

There have been 36 studies with 10,240 participants testing RBI in choice. These show that violations of RBI are significantly more frequent in opposition to predictions of CPT (or any RDU or RSDU model with an inverse-S-weighting function) than in agreement with them. Birnbaum and Navarrete (1998) included 27 tests of this property for each participant, with different probability distributions and different levels of the consequences. Of 100 participants, 65 (significantly more than half) had more violations of the SR' type than of the opposite. Evidence has shown similar results despite a variety of procedural manipulations (Birnbaum, 2004b, 2005a, 2006, 2007b). Birnbaum (2004b) fit the "true and error" model to replicated choices; this analysis indicated that the SR' violations cannot be attributed to random "error," but violations of the opposite pattern (RS', predicted by CPT) can be set to zero with no reduction in fit. In addition, the same pattern has been found to be frequent and substantial in judged buying and selling prices of gambles and investments (Birnbaum & Beeghley, 1997; Birnbaum & Veira, 1998; Birnbaum & Zimmermann, 1998). Because the same pattern of results occurs in both judgment and choice, it seems likely that they are produced by the evaluation of the gambles rather than by the choice between gambles.

Wakker (2001) summarized evidence from studies in which the "sure thing" principle is tested as a confounded combination of both RBI and coalescing. Although he acknowledged contrary evidence, Wakker concluded that evidence supporting the inverse-S-weighting function is "overwhelming." But this conclusion depends on the assumption of a family of models, including CPT, that assume coalescing, and experiments were not designed to test that model. Pure tests of RBI provide stronger evidence against the inverse-S-weighting function than the evidence summarized by Wakker, so it can be said that evidence against Wakker's conclusion is more than overwhelming. How can these apparently contradictory results be resolved? According to the rank-dependent models, there is a serious contradiction; however, TAX and RAM models reconcile both sets of results. The TAX model fits both the confounded tests cited by Wakker as well as the pure tests of RBI. It also fits studies of distribution independence, which also test the inverse-S-weighting function; these are taken up in the next four sections.

The priority heuristic of Brandstätter et al. (2006) also predicts violations of RBI; however, like CPT, it predicts the opposite pattern from what has been observed. According to this model,

people should choose $R > S$ because the lowest branches are the same but the highest consequence on the risky gamble is better; and they should choose $S' > R'$ because the highest consequences are the same, and S' has the better lowest consequence.

From the calculations for CEs of binary gambles as a function of probability (e.g., as in Figure 3), it might be thought that the configural weight, TAX and RAM models use the "same" weighting function as is used in CPT. However, that inference would be wrong because in RAM and TAX, that figure does not represent the probability weighting function, which is approximated with a power function, $t(p) = p^\gamma$. Results for tests of RBI and distribution independence, described in sections on RBI and distribution independence, show that no inverse-S decumulative weighting function (whether strongly or weakly inverse S) is compatible with empirical choices. Thus, despite their agreement with CPT for binary gambles, TAX and RAM do not use the same weighting function as in CPT. Instead, their weighting functions can reproduce the violations of RBI, which the RDU, RSDU, and CPT models with the inverse S cannot do. Nor should the weighting function in TAX be confused with that of OPT because OPT implies no violations of RBI.

4-Distribution Independence

4-distribution independence (4-DI) is an interesting property because RDU/RSDU/CPT models imply systematic violations, but the property must be satisfied, according to RAM.

$$S = (z', r; x, p; y, p; z, 1 - r - 2p) >$$

$$R = (z', r; x', p; y', p; z, 1 - r - 2p)$$

$$\Leftrightarrow$$

$$S' = (z', r'; x, p; y, p; z, 1 - r' - 2p) >$$

$$R' = (z', r'; x', p; y', p; z, 1 - r' - 2p),$$

where $z' > x' > x > y > y' > z \geq 0$. According to TAX, this property can be violated, but in the opposite way from that predicted by any inverse-S-weighting function in RSDU/RDU/CPT.

Choice Problems 14.1 and 14.2 in Table 6 illustrate a test of 4-DI. Note that two branches of equal probability ($p = 0.2$ in this example) are nested within a probability distribution in which they are either near the low end of decumulative probability or near the upper end of decumulative probability. According to RDU/RSDU/CPT, this change in the distribution should change the relative weights of these common branches, producing violations of 4-DI. If the weights of the four ranked branches are w_1, w_2, w_3, and w_4, w_2/w_3 should be greater than 1 at the low end of decumulative probability, and w_2'/w_3' should be less than 1 when these two branches are at the upper end of decumulative probability.

According to CPT/RDU/RSDU model with any inverse-S-weighting function, $R > S$ and $S' > R'$, that is, the RS' pattern of violations. RAM allows no systematic violations of this property (proof in Birnbaum & Chavez, 1997, pp. 176–177). According to SWU, OPT, or any theory with prospect theory's cancellation principle, there should be no violations. According to TAX model with its previous parameters, however, people should show the SR' pattern of violations. In a study with 100 participants and 12 tests of 4-DI interspersed among 130 choices, Birnbaum and Chavez

(1997) found that the pattern predicted by TAX was more frequent than the pattern consistent with any inverse-S decumulative weighting function in all 12 tests. In Choice Problems 14.1 and 14.2, for example, there were 23 participants who had the SR' pattern against only 6 who showed the opposite reversal of preferences. These results also favor TAX over models that imply no violations, which includes RAM, SWU, PRT, OPT, and any other theory that assumes cancellation.

3-Lower Distribution Independence

Define 3-lower distribution independence (3-LDI) as follows:

$$S = (x, p; y, p; z, 1 - 2p) > R = (x', p; y', p; z, 1 - 2p)$$

$$\Leftrightarrow S2 = (x, p'; y, p'; z, 1 - 2p') >$$

$$R2 = (x', p'; y', p'; z, 1 - 2p'),$$

where $x' > x > y > y' > z \geq 0$. The name "lower" distribution independence is used to indicate that the lower branch, $(z, 1 - 2p)$, is common to both choices, and it is the probability of this branch (rather than its value) that is changed. Note that the first comparison involves a distribution with probability p to win x or y, whereas the second choice involves a different probability, p'.

According to 3-LDI, people should choose S over R in Choice Problem 15.1 of Table 7 if and only if they choose $S2$ over $R2$ in Choice Problem 15.2, except for error. The CPT model as fit to previous data predicts that people should violate this property by choosing R over S and $S2$ over $R2$, whereas special TAX and RAM predict that people should choose both S and $S2$.

We can again apply the generic model to the choice between S and R as follows:

$S > R \Leftrightarrow$

$$w_1 u(x) + w_2 u(y) + w_3 u(z) > w_1 u(x') + w_2 u(y') + w_3 u(z),$$

where w_1, w_2 and w_3 are the weights of the highest, middle, and lowest ranked branches, respectively, which depend on the value of p (differently in different models). Subtracting the term $w_3 u(z)$ from both sides, the following can be derived:

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

Suppose there is a violation of 3-LDI, in which $R2$ $S2$. By a similar derivation,

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

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

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

A reversal from $S < R$ to $S2 > R2$ can occur with the opposite ordering. But if the ratio of weights is independent of p (e.g., as in EU), then there can be no violations of this property.

According to RAM, this ratio of weights is as follows:

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

Table 7

Tests of 3-Lower Distribution Independence (Choice Problems 15.1 and 15.2), 3-2 Lower Distribution Independence (Choice Problems 16.1 and 16.2), and 3-Upper Distribution Independence (Choice Problems 17.1 and 17.2)

Choice problem	Choice		% choosing risky gamble	Prior TAX		Prior CPT	
	Safe	Risky		Safe	Risky	Safe	Risky
15.1	S: 20 to win $58 20 to win $56 60 to win $2	R: 20 to win $96 20 to win $4 60 to win $2	24	21.7	13.8	19.9	**21.3**
15.2	S2: 45 to win $58 45 to win $56 10 to win $2	R2: 45 to win $96 45 to win $4 10 to win $2	19	36.9	22.9	41.0	35.8
16.1	S0: 50 to win $44 50 to win $40	R0: 50 to win $96 50 to win $4	31	41.3	34.7	41.7	39.3
16.2	S: 48 to win $44 48 to win $40 04 to win $2	R: 48 to win $96 48 to win $4 04 to win $2	34	29.1	24.5	34.7	**37.7**
17.1	S': 80 to win $100 10 to win $44 10 to win $40	R': 80 to win $100 10 to win $96 10 to win $4	56	61.6	63.4	**77.4**	71.7
17.2	S2': 10 to win $100 45 to win $44 45 to win $40	R2': 10 to win $100 45 to win $96 45 to win $4	33	45.9	43.9	50.3	42.6

Note. Entries in bold show incorrect predictions of modal choice. All entries are significantly different from 50%. TAX = transfer of attention exchange; CPT = cumulative prospect theory.

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

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

$$\frac{w_2}{w_1} = \frac{t(p) + (\delta t(p)/4)) - (\delta t(p)/4))}{t(p) - (^2\!/\!_4)\delta t(p)} = \frac{t(p)}{t(p)[1 - \delta/2]} = \frac{w_2'}{w_1'}.$$

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

According to CPT or any RDU/RSDU model, however, there should be violations of 3-LDI if the weighting function is nonlinear. The relation among the weights will be as follows:

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

The inverse-S-weighting function of CPT implies that these ratios differ by a factor of almost 2 in Table 7. Thus, CPT predicts violations of 3-LDI, and both RAM and TAX predict no violations.

In tests like Choice Problems 15 (and 16) in Table 7, CPT goes on offense by predicting violations, leaving TAX and RAM to defend the null hypothesis. Birnbaum (2005b) conducted three studies with 1,578 participants testing these predictions. Results showed that TAX and RAM were more accurate than CPT with its inverse-S-weighting function, which predicted reversals that failed to materialize. Of course, failure to find a predicted effect does not disprove a model, but merely adds to the list of failed predictions by CPT with its prior parameters.

3-2 Lower Distribution Independence

The property of 3-2 lower distribution independence (3-2 LDI) requires:

$$S0 = (x, ^1\!/\!_2; y, ^1\!/\!_2) > R0 = (x', ^1\!/\!_2; y', ^1\!/\!_2)$$

$$\Leftrightarrow S = (x, p; y, p; z, 1 - 2p) > R = (x', p; y', p; z, 1 - 2p).$$

By a similar derivation (to that in 3-LDI), RAM predicts the same choices between $S0$ and $R0$ as between S and R, when $a(2, 2)/a(1, 2) = a(2, 3)/a(1, 3)$, as in the prior RAM model, in which both ratios are 2/1.

Prior TAX also implies that people should make the same choices in $S0$ and $R0$ as between S and R, because the weight ratio in a 50-50 two-branch gamble is $\dfrac{w_2}{w_1} = \dfrac{t(^1\!/\!_2) + (^8\!/\!_3)t(^1\!/\!_2)}{t(^1\!/\!_2) - (^8\!/\!_3)t(^1\!/\!_2)} = \dfrac{1 + (\delta/3)}{1 - (\delta/3)}$, which also equals 2/1 when $\delta = 1$, so this TAX model also satisfies 3-2 LDI.

In contrast, CPT, with its prior parameters, predicts that people should prefer $S0 > R0$ and choose $R > S$ in Choice Problems 16.1 and 16.2 in Table 7, violating 3-2 LDI. Results shown are from Birnbaum (2005b, Experiment 1), with $n = 1,075$. Predicted violations by CPT failed again to materialize.

3-Upper Distribution Independence

The property 3-upper distribution independence (3-UDI) is defined as follows:

$$S' = (z', 1 - 2p; x, p; y, p) > R' = (z', 1 - 2p; x', p; y', p)$$

$$\Leftrightarrow S_2' = (z', 1 - 2p'; x, p'; y, p') >$$

$$R_2' = (z', 1 - 2p'; x', p'; y', p'),$$

where $z' > x' > x > y > y' \geq 0$. This property will be violated, with $S' > R'$ and $S2' < R2'$ if and only if:

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

The opposite reversal of preference would occur when the order above is reversed.

The RAM model implies that:

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

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

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

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

which shows that this weight ratio is not independent of p. For prior TAX, this weight ratio increases from 1.27 to 1.60, which straddles the ratio of differences $(96-44)/(40-4) = 1.44$; therefore, TAX predicts that $S' < R'$ and $S2' > R2'$, the $R'S2'$ pattern of violation in Choice Problems 17.1 and 17.2.

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

$$\frac{w_3}{w_2} = \frac{1 - W(1 - p)}{W(1 - p) - W(1 - 2p)}$$

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

In particular, for the inverse-S-weighting function of Tversky and Kahneman (1992), this ratio of weights increases and then decreases, with a net decrease from 1.99 to 1.71. Therefore, CPT implies violations of the opposite pattern from that predicted by TAX, but CPT predicts no violation in Choice Problem 17, given its prior parameters.

Birnbaum (2005b) found that significantly more than half the 1,075 participants chose $R' > S'$, and significantly more than half chose $S2' > R2'$. This result violates RAM, was not predicted by CPT, and was predicted correctly by TAX. Other cases in which prior CPT predicted violations of 3-LDI and 3-UDI were tested by Birnbaum (2005b), who reported that the predicted violations of CPT failed to materialize.

Properties Involving Judgments

One can ask whether the same models that are used to represent the values of the alternatives in choice would also apply in judgment, where people are asked to judge each alternative separately. The next sections concern cases in which the rank order of judgments can be changed by manipulating the judge's point of view

and in which the rank order of judgments is compared with the order inferred from choice (preference reversals).

Buying and Selling Prices: Viewpoint or Endowment?

Birnbaum, Wong, and Wong (1976) reported an interaction between estimates from different sources in judged buying prices of used cars. This interaction was consistent with the idea that people assign greater configural weight to lower than to higher estimates of value. Birnbaum and Stegner (1979, Experiment 5) reasoned that such configural weighting was the result of asymmetric payoffs to the judge. That is, a buyer makes a worse error by overestimating value than by underestimating value. They conjectured that the configural weight parameter could be manipulated by changing the judge's point of view from buyer to that of seller. Experiments confirmed this hypothesis: The rank order of judgments changed systematically between buying and selling prices.

Birnbaum and Stegner (1979) concluded that whereas buyers assign greater weight to lower estimates of value, sellers assign greater weight to higher estimates of value. Birnbaum and Sutton (1992), Birnbaum and Beeghley (1997), and Birnbaum and Veira (1998) showed that this same pattern of changing configural weight is found in judgments of buying and selling prices of gambles as well as with so-called sure things like used cars. Birnbaum and Zimmermann (1998) showed the same pattern of changing configural weighting for buying and selling prices of investments.

Following Birnbaum and Stegner's (1979) work, the difference between buying and selling prices was later renamed the "endowment" effect (Thaler, 1980). Kahneman, Knetsch, and Thaler (1990) and Tversky and Kahneman (1991) tried to account for the difference between willingness to pay (buying price) and willingness to accept (selling price) by means of a utility function with different slopes for gains and losses of the same absolute size, what they call "loss aversion." However, the theory of Tversky and Kahneman (1991) implies that the ratio of selling to buying prices should be a constant, and therefore, buying prices should be monotonically related to selling prices of the same objects (Birnbaum & Zimmermann, 1998). In order to avoid the wrong prediction that the buying and selling prices of a $5 bill should differ by a factor of two or more, a special exception was made for money or goods held for trade. Even with this special exemption, the "loss aversion" theory had already been refuted by the experiments of Birnbaum and Stegner (1979), who showed that buying and selling prices are not monotonically related to each other (see also Birnbaum, 1982). Nor did the configural weight model of Birnbaum and Stegner (1979) predict that buying and selling prices of a $5 bill should differ.

Thus, the "loss aversion" account of the "endowment" effect is not consistent with judgments of buying and selling prices of either "sure things" of uncertain value (like used cars or stocks) or standard risky gambles. The term *endowment effect* has also created some confusion because it is not necessary to actually endow anyone with anything to observe the effect; one need only instruct the participants to identify with or to advise a buyer, seller, or a neutral judge to observe the effect.

Birnbaum and Zimmermann (1998) combined the theory of Tversky and Kahneman (1991) with CPT of Tversky and Kahne-

man (1992) by assuming that people integrate prices with the prizes of a gamble and that they apply CPT with its assumption of "loss aversion" to the integrated mixed gambles. These assumptions lead to the implication of a property called *complementary symmetry*. Let $B(x, p; y, 1 - p)$ represent the judged "highest buying price" for gamble $G = (x, p; y, 1 - p)$, where $x > y \geq 0$. Let $S(x, 1 - p; y, p)$ represent the "lowest selling price" of the complementary gamble. Complementary symmetry is the assumption that $B(x, p; y, 1 - p) + S(x, 1 - p; y, p)$ is a constant, independent of $|x - y|$ and of p. This property was systematically violated by the data of Birnbaum and Sutton (1992) and refuted in a wider test by Birnbaum and Yeary (2001). Thus, the combination of CPT and "loss aversion" fails to provide a consistent account of buying and selling prices of gambles (Birnbaum & Zimmermann, 1998).

Luce's (2000) still more general approach to integrating prices with prizes by his theory of joint receipts in RSDU was fit to judgments of binary gambles by Birnbaum, Yeary, Luce, and Zhou (2001). Birnbaum et al. also concluded that RAM and TAX models are more accurate descriptions of judged buying and selling prices than any of the models stemming from this representation. In summary, theories in which the relation between buying and selling prices is attributed to "loss aversion" in the utility function have not yet been successful in describing the difference between buying and selling prices. Instead, judgments of buying and selling prices of both sure things and gambles can be better described by the theory that configural weights change in the different points of view.

Birnbaum and Beeghley (1997) collected judgments of the highest buying price and lowest selling prices of three-branch gambles, for which each of three consequences were equally likely. That study included many tests of RBI. The authors found that there were strong violations of RBI in both buying and selling prices. They also found that these violations were different in different viewpoints, which again contradicts the loss aversion theory of buying and selling prices. Violations of RBI were of the SR' type in both viewpoints, which shows that the inverse-S decumulative weighting function of RDU or CPT cannot account for judgments of value. Birnbaum and Beeghley were able to use configural weight models to reproduce these "preference reversals" between buying and selling prices, and also preference reversals between price judgments and choices between gambles. The model assumes that the utility function for money is the same for buying prices, selling prices, and choices, and only the configural weights differ in these three situations. Their estimated weights are displayed in Table 8. The fact that violations of RBI are of the same type in judgment and choice (i.e., of the SR' type) suggests that it is in the evaluation stage rather than in the choice stage that these violations are generated. A similar pattern of violations was found in four-branch gambles by Birnbaum and Veira (1998).

Preference Reversals in Ratings and Price Judgments

Tversky, Sattath, and Slovic (1988) proposed a contingent weight model to account for preference reversals between attractiveness ratings and judged prices of binary gambles of the form $(x, p; 0)$, where $x > 0$. According to this contingent weight model (which should not be confused with Birnbaum's, 1974a, configural weighting model), the relative weights of probability and prize values depend on the task; presumably, probability has greater

Table 8

Estimated Weights of Equally Likely Consequences as a Function of Rank in Three-Branch Gambles (From Birnbaum & Beeghley, 1997)

Experiment	Lowest	Middle	Highest
Buyer's prices	.56	.36	.08
Seller's prices	.27	.52	.21
Choices	.51	.33	.16

Note. Relative weights are normalized to sum to one by dividing by the sum of weights in each case. Values for choices are from Birnbaum and McIntosh (1996). All three experiments are fit with the same utility function, $u(x) = x$, for $0 < x \leq \$148$.

weight in the attractiveness rating task, and prizes have greater weight in the price judgment task. Mellers et al. (1992) found that the empirical changes in rank ordering were not consistent with the contingent weight model. Instead, the data were consistent with the theory that people combined probability and prizes by an additive model in the ratings task and by a multiplicative one in the price judgment task. This theory preserved scale convergence; that is, the utility function was the same in both the additive and multi-plicative models. In addition, the authors found a contextual effect; this change of process occurs only when the gambles are of the form $(x, p; 0)$, where all $x > 0$ or all $x < 0$ and few values were at or near zero.

Choice Response Times and Configural Weighting

Birnbaum and Jou (1990) presented a model of choice response times to describe three major phenomena of choice response times: the distance effect, the end effect, and the semantic congruity effect. The *distance effect* refers to the fact that it takes less time to choose between two stimuli that are farther apart in subjective value. The *end effect* refers to the fact that choices are faster for stimuli at or near the ends of the series of stimuli used in a study. The *semantic congruity effect* refers to the facts that one can decide faster which of two good stimuli is "better" than to decide which of two bad stimuli is better, and it is faster to decide which of two bad stimuli is "worse" than to decide which of two good stimuli is worse. The Birnbaum and Jou (1990) model can be viewed as an extension of random walk models of choice, such as studied by Link (1975, 1992, Busemeyer and Townsend (1993), and others.

There were two new ideas in Birnbaum and Jou (1990): The first was a model of the bias parameters for the semantic congruity effect that enabled a common scale to represent "difference" judgments as well as "more" and "less" choice response times. This model thus accounted not only for the three main phenomena of choice response time but also for direct ratings of subjective "differences." The second new idea was to implement this model of choice response times within Birnbaum's (1974a, Experiment 4) "scale free" test of additive or constant weight averaging models of impression formation. Violations of the additive model in the scale-free tests are taken as evidence of configural weighting. The results showed converging evidence from both choice response times and direct judgments against the additive models in favor of an interaction that was represented by greater configural weighting for stimuli of lower evaluation.

Whereas Busemeyer and Townsend (1993) had used a model without configural weights, Johnson and Busemeyer (2005) presented a random walk model to account for the data of Birnbaum and Beeghley (1997), in which there is configural weighting in evaluation of gambles, but bias parameters for buying and selling prices differ, and configural weighting is the same in both view-points. Johnson and Busemeyer concluded that this model fit the data of Birnbaum and Beeghley (1997) about as well as the purely configural weight model with the same number of parameters. Both models correctly imply changes of rank order between buy-ing and selling prices. As noted by Leth-Steensen and Marley (2000) in reference to Birnbaum and Jou (1990), it can be difficult to distinguish effects of bias parameter from those attributed to the evaluation process. A critical test between these interpretations has not yet been conducted.

Discussion

The list of paradoxes and properties are summarized in Tables 9 and 10. Table 9 contains six properties that must be satisfied by any version of RDU/RSDU/CPT and whose violations were pre-dicted correctly by prior RAM and TAX. Violations of coalescing, stochastic dominance, upper tail independence, and lower and upper cumulative independence create contradictions within any of these theories. Because there are no functions and parameters within those theories that can explain these violations, they can be called "new paradoxes" of choice. Violations of the properties in Table 9 also add to the list of paradoxes of EU theory because EU is a special case of RDU/RSDU/CPT and must also satisfy these properties.

Violations of coalescing are probably at the root of the first five new paradoxes in Table 9. Violations of coalescing disprove the editing principle of combination that was included as an editing principle in OPT. Researchers cannot reject the idea that some participants might use such an editing rule on some occasions, but researchers can reject this editing rule if it is treated as a scientific theory. Any theory that represents the gambles as prospects or probability distributions is refuted by systematic violation of co-alescing. Violations of stochastic dominance also refute RDU, RSDU, CPT, and other theories that have assumed stochastic dominance, including security potential aspiration level theory (Lopes & Oden, 1999) and lottery-dependent utility (Becker & Sarin, 1987), among others.

Gain–loss separability is satisfied by OPT, CPT, and RSDU, so violations of this property (Birnbaum & Bahra, 2007; Wu & Markle, 2005) refute both versions of prospect theory as well as other models that share this property. Violations of gain–loss separability also bring into question the theory of loss aversion as attributed to the utility function in CPT. This is not to say that a kinked utility function has been disproved; rather, it is the assump-tions that lead to the kinked utility function that have been dis-proved. Wu and Markle (in press) have recently completed new tests confirming and extending their previous findings. Violations of gain–loss separability in two-branch gambles violate both ver-sions of prospect theory.

It might seem that the issue of whether loss aversion is produced by greater weight (attention) being given to losses or greater disutility to losses than to gains of equal absolute value is merely a semantic or mathematical issue. Both theories can imply the

Table 9
Summary of Six New Paradoxes That Refute Cumulative Prospect Theory, Rank-Dependent Utility, and Rank-and Sign-Dependent Utility Theories

Property name	Expression	Implications	References
Event-splitting effect (coalescing & transitivity)	$A = (x, p; z, 1 - p) > B = (y, q; z, 1 - q)$ $\Leftrightarrow A' = (x, p - r; x, r; z, 1 - p) > B' = (y, q; z, s; z, 1 - q - s)$	Violations refute RDU, RSDU, CPT	SS93, H95, B99c, B00, B01, BM03, B04a; B04b; B05a; B05b; B06, TK86; BN98; B99c; BPL99; B00; B01; B07b; BB07
Stochastic dominance	$P(x > t \| G^+) \geq P(x > t \| G^-) \forall t$ $\Rightarrow G^+ > G^-$	Violations refute RDU, RSDU, CPT	B04a; B04b; B05a; B05b; B06; B07b; BM03
Upper tail independence	$S = (x, p; y, q; z, 1 - p - q) >$ $R = (x, p; y', q'; z,' 1 - p - q')$ $\Leftrightarrow S_1 = (y', p; y, q; z, 1 - p - q >$ $R_1 = (y', p + q'; z', 1 - p - q')$	Violations refute GDU, RDU, RSDU, CPT	W94, B01, B05b
Lower cumulative independence	$S = (x, p; y, q; z, 1 - p - q) >$ $R = (x', p; y', q; z, 1 - p - q)$ $\Rightarrow S'' = (x, p + q; y', 1 - p - q) >$ $R'' = (x', p; y'; 1 - p)$	Violations refute RDU, RSDU, CPT	BN98, B99c, BPL99; B04b; B05a; B06
Upper cumulative independence	$S' = (z', 1 - p - q; x, p; y, q) <$ $R' = (z', 1 - p - q; x', p; y', q)$ $\Rightarrow S''' = (x', 1 - p - q; y, p + q) <$ $R''' = (x', 1 - q; y', q)$	Violations refute RDU, RSDU, CPT	BN98, B99c, BPL99; B04b; B05a; B06
Gain–loss separability	$A^+ > B^+$ and $A^- > B^- \Rightarrow A > B$	Violations refute RSDU, CPT, OPT	WM05, BB07

Note. B99c, B00, B01, B04a, B04b; B05a, B05b; B06, B07b = Birnbaum (1999c, 2000, 2001, 2004a, 2004b; 2005a; 2005b, 2006, 2007b); BB07 = Birnbaum and Bahra (2007). BM03 = Birnbaum and Martin (2003); BN98 = Birnbaum and Navarrete (1998); BPL99 = Birnbaum et al. (1999); H95 = Humphrey (1995); SS93 = Starmer and Sugden (1993); TK86 = Tversky and Kahneman (1986); W94 = Wu (1994); WM05 = Wu and Markle (2005). RDU = rank-dependent utility; RSDU = rank-and sign-dependent utility; CPT = cumulative prospect theory; GDU = gains decomposition utility; OPT = original prospect theory. $z' > x' > x > y > y' > z \geq 0$.

behavioral properties that people would rather win than lose and that they tend to avoid fair mixed bets. However, the utility interpretation ("losses loom larger") assumes gain–loss separability, whereas the configural weight theory (losses get greater weight) of this phenomenon implies violations of this property. Given the sparse data on mixed gambles and gain–loss separability, it cannot yet be determined whether both factors (utility and weight) contribute to this effect, but the violations of gain–loss

Table 10
Summary of Five Properties Implied by SWU and OPT Whose Violations Contradict or Fail to Confirm Predictions of CPT

Property name	Expression	Empirical results	References
RBI	$S = (x, p; y, q; z, 1 - p - q) >$ $R = (x', p; y', q; z, 1 - p - q)$ $\Leftrightarrow S' = (z', 1 - p - q; x, p; y, q) >$ $R' = (z', 1 - p - q; x', p; y', q)$	Pattern of violations refutes CPT with inverse-S.	BMc196, BC97, BN98, B99c, BPL99, B04a, B04b; B05a, B05b; B06, B07b
4-DI	$S = (z', r; x, p; y, p; z, 1 - r - 2p) >$ $R = (z', r; x', p; y', p; z, 1 - r - 2p)$ $\Leftrightarrow S' = (z', r'; x, p; y, p; z, 1 - r' - 2p) >$ $R' = (z', r'; x', p; y', p; z, 1 - r' - 2p)$	Violations refute RAM; pattern violates CPT with inverse-S.	BC97
3-LDI	$S = (x, p; y, p; z, 1 - 2p) >$ $R = (x', p; y', p; z, 1 - 2p)$ $\Leftrightarrow S2 = (x, p'; y, p'; z, 1 - 2p') >$ $R2 = (x', p'; y', p'; z, 1 - 2p')$	Failed to confirm predictions of CPT.	B05b
3-2 LDI	$S0 = (x, 1/2; y, 1/2) > R0 = (x', 1/2; y', 1/2)$ $\Leftrightarrow S = (x, p; y, p; z, 1 - 2p) >$ $R = (x', p; y', p; z, 1 - 2p)$	Failed to confirm predictions of CPT.	B05b
3-UDI	$S' = (z', 1 - 2p; x, p; y, p) > R' = (z', 1 - 2p; x', p; y', p) \Leftrightarrow S'_2 = (z', 1 - 2p'; x, p'; y, p') > R'_2 = (z', 1 - 2p'; x', p'; y', p')$	Violations refute RAM; pattern refutes CPT with inverse-S.	B05b

Note. Consequences are ordered such that $z' > x' > x > y > y' > z > 0$. B99c, B04a, B04b, B05a, B05b; B06, B07b = Birnbaum (1999c, 2004a, 2004b, 2005a, 2005b, 2006, 2007b); BB97 = Birnbaum and Beeghley (1997); BC97 = Birnbaum and Chavez (1997); BMc196 = Birnbaum and McIntosh (1996); BN98 = Birnbaum and Navarrete (1998); BPL99 = Birnbaum et al. (1999). SWU = subjectively weighted utility; OPT = original prospect theory; CPT = cumulative prospect theory; RBI = restricted branch independence; 4-DI = 4-distribution independence; RAM = rank-affected multiplicative weights; 3-LDI = 3-lower distribution independence; 3-2 LDI = 3-2 lower distribution independence; 3-UDI = 3-upper distribution independence.

separability show that the utility function approach by itself does not work.

The Marley and Luce (2001) version of GDU satisfies upper tail independence, so violations of this property (Wu, 1994) are evidence against the lower GDU model as well as of the RDU/RSDU/CPT family. See also Birnbaum (2007b).

Table 10 lists five independence properties that are systematically violated by RDU, RSDU, and CPT if the weighting function is nonlinear. These properties must be satisfied by PRT, SWU, "stripped" prospect theory (including EU), OPT, and any other theory assuming the editing rule of cancellation. If people cancelled common branches, then they would satisfy all five of these properties. In these tests, data either fail to confirm predicted violations of CPT or show the opposite pattern of violations from what is predicted by CPT with its inverse-S-weighting function. Empirical violations of RBI, 4-DI, and 3-UDI are of the opposite type from what is predicted by CPT with its inverse-S-weighting function. These phenomena also refute PRT, SWU, OPT, and any other theory that assumes cancellation, but they are compatible with predictions of TAX.

Summing up, prior CPT predicts only 17 of the 40 modal choices (42%) analyzed in Tables 1, 2, 5, 6, and 7 (it is correct for Choice Problems 1.2, 3.2, 4, 5.1, 6.2, 7.2, 8.2, 9.2, 10.1, 10.3, 10.5, 12.3, 12.7, 12.8, 15.2, 16.1, and 17.2).

According to OPT with or without the editing principle of cancellation, there should be no violations of either branch independence or distribution independence (see Tables 6 and 7). Systematic violations of the properties of Table 10 refute SWU, PRT, and OPT, with or without cancellation. According to OPT with the editing principle of combination, there should be no violations of coalescing and no violations of the tests of Tables 1 and 2, summarized in Table 9. OPT satisfies gain–loss separability, so violations of that property in two-branch gambles, as reported by Wu and Markle (2005), refute OPT. In summary, no version of OPT, with or without its editing principles, has been found that is compatible with the data summarized here.

An important property not listed in Tables 9 or 10 is branch-splitting independence (Birnbaum, 2007b; Birnbaum & Navarrete, 1998). This is the assumption that splitting the same probability-consequence branch in the same way should have the same effect on any gamble containing that branch. For example, splitting a branch of .2 to win $50 into two branches of .1 to win $50 should either improve the gamble or diminish it whether $50 is the highest outcome in the gamble or the lowest. Because CPT cannot violate coalescing, this property is moot under CPT as well as in any other model that satisfies coalescing. TAX and RAM imply violations of both coalescing and of branch-splitting independence: Splitting the higher branch should improve a gamble, and splitting the lower branch should make it worse. Idempotent, SWU, including PRT, will also violate both coalescing and branch-splitting independence. Lower GDU implies upper coalescing but violates lower coalescing. Additive SWU and stripped OPT imply violations of both upper and lower coalescing, but they satisfy branch-splitting independence. Birnbaum (2007b) found large violations of upper coalescing, which refute lower GDU, and he found small, but significant violations of branch-splitting independence. Combined with evidence of violations of branch and distribution independence, those data contribute to the case against additive SWU, and stripped PT.

The three "classic" variations of the Allais paradox, which can be viewed as cases inside the probability simplex, confound RBI with coalescing. When these two properties are teased apart, it is found that the data refute both OPT and CPT with or without the editing principles of cancellation and combination. The data fall in the one cell of Table 3 in which neither version of prospect theory can lay claim. Namely, data show that the Allais paradoxes are produced by violations of coalescing and that violations of branch independence actually work in the opposite direction to these Allais paradoxes (Birnbaum, 2004a, 2007b).

Considering all of the evidence, there is a strong case against both versions of prospect theory with or without the editing principles of cancellation, combination, and dominance detection. If people satisfied combination, then there would be no violation of the first five properties in Table 9, and if cancellation held, then there would be no violation of the five properties in Table 10.

If there is a dominance detector, then it fails to work effectively for cases such as Choice Problem 2 in Table 1, which are predicted to fail according to RAM and TAX, but the detector "works" in Choice Problems 3.2 and 4 in which TAX and RAM predict dominance to be satisfied by the majority. One might assert that the dominance detector only works when dominance is "transparent" and otherwise does not work, but one would still need to define *transparent* in some noncircular way (Birnbaum et al., 1999).

An economist might suggest the following defense of CPT: It should not be required that CPT predict the results of manipulating the form of a choice (split vs. coalesced) because form of a choice is not an economic variable. No rational economic person would knowingly violate stochastic dominance, so CPT should not be required to predict such violations, even though people exhibit them. From this viewpoint, the "correct" way to present a choice is in the canonical split form, in which very few people violate stochastic dominance (e.g., Choice Problem 3.2). If we adopt this position, then it follows that the inverse-S-weighting function is false because when Allais paradoxes are dissected, the canonical split form produces violations of RBI that rule out the inverse-S-weighting function. Because the Allais paradoxes confound coalescing and branch independence, this economic position rules out CPT as a theory of the Allais paradox. The "overwhelming" evidence of inverse-S-weighting function, cited by Wakker (2001), is then rendered irrelevant by this position because such evidence comes from choices in which coalescing is used, and is therefore outside the new, restricted domain of the economic theory. Finally, this theory requires different weighting functions in CPT for two-branch and three-branch gambles. This modified theory does not seem particularly attractive as a descriptive model, and it is unclear why it should be preferred to EU as an economic model.

Two models are partially consistent with the data, RAM and idempotent, lower GDU. Despite success in predicting violations of properties that rest on coalescing and the pattern of RBI, prior RAM cannot describe violations of 4-DI and 3-UDI. Therefore, TAX is more accurate than RAM because it predicts these results correctly. TAX has also been found to be more accurate than RAM in predicting tests of stochastic dominance (Birnbaum, 2005a).

GDU also gives a good fit, except for tests that rest on upper coalescing. GDU cannot account for violations of upper tail independence, nor can it account for direct tests of upper coalescing by Birnbaum (2007b). GDU can, however, account for violations of stochastic dominance, lower and upper cumulative independence, and the pattern of violations of branch independence and distribu-

tion independence. Luce, Ng, Marley, and Aczél (in press) are presently exploring gains decomposition without the assumption of idempotence, which may yield a more accurate model.

The model that best describes all of these new paradoxes is TAX. With prior parameters, it correctly predicted all 10 of the modal choices in Table 1, 11 of the 12 choices in Table 2, all eight choices in Table 5, all four in Table 6, and all six in Table 7. The only case in which prior TAX was wrong was in Choice Problem 9.3, in which consequences were in the millions of dollars. It is worth repeating that predictions for TAX had been calculated from simplified parameters and used to predict tests of lower cumulative independence, upper cumulative independence, stochastic dominance, 3-LDI, 3-2 LDI, and 3-UDI. These parameters were used to design tests that differentiate CPT and TAX, and the new tests confirmed the predicted results by prior TAX. It is also worth repeating that although the simplifying assumption that $u(x) = x$ was used here to illustrate the predictions, this assumption is not part of the model, and it is not optimal. Furthermore, the finding that $u(x) = x$ works does not imply that only $u(x) = x$ works.

TAX requires a nonlinear utility function to fit the data in Choice Problems 9.1–9.3 of Table 2, in which large consequences (millions of dollars) were used (Birnbaum, 2007b). Aside from cases involving large consequences, its success in predicting new results, including the recipe for violations of stochastic dominance (Birnbaum, 1997) and the breakdown of the Allais paradoxes (Birnbaum, 2004a), cannot be dismissed as post hoc model fitting. Those predictions were used to design new choices and were successful in predicting new results without fitting anything to those data.

The TAX model describes all 11 of the new paradoxes reviewed here that refute CPT, and it also accounts for those classic phenomena involving two-branch gambles and three-branch gambles in the probability simplex triangle that were successfully described by CPT, using no more parameters than CPT. Because this model was successful in predicting new as well as old phenomena, its success might be due to more than luck.

There is no claim that the success of the linear utility function used here to simplify the presentation refutes other utility functions. This simplifying assumption was used to make it clear that the properties reviewed here can all be understood by assumptions concerning configural weighting. Similarly, whereas the violations of gain–loss separability can be described by a very simple version of TAX without any loss aversion attributed to the utility function, no claim is made that the findings reviewed here refute a kinked or nonlinear utility function.

Brandstätter et al. (2006) argued that CPT is more accurate than TAX when TAX is restricted by the assumptions that $u(x) = x^\beta$ with $\beta = 1$, and CPT is allowed to use up to three best-fit values of $\beta < 1$. But this conclusion dissolves if researchers allow both TAX and CPT to use best-fit parameters. For example, the TAX model with $\beta = 1$, $\gamma = 0.7$, and $\delta = 1$ predicted only 73 of the 100 modal choices correctly between random binary gambles in Erev, Roth, Slonim, and Barron (2002). However, if researchers use the median best-fit estimates reported by Birnbaum and Navarrete (1998), $\hat{\beta} = 0.41$, $\hat{\gamma} = 0.79$, and $\hat{\delta} = 0.95$, then these parameters reproduce all but 4 of the 100 modal choices correctly in the same data. And when β is estimated from those data, TAX fits 99 of 100 modal choices correctly, the same as the optimal fit of CPT to the same data. The conclusion has already been stated that binary gambles are not going to allow researchers to compare TAX and

CPT because these models are virtually identical for binary gambles: When both models are allowed the same number of free parameters, they achieve virtually equal fit to those data. Similarly, both TAX and CPT can reproduce the modal choices in Kahneman and Tversky (1979) perfectly, if they are allowed to estimate their parameters from those data. Those choices all fall inside the probability simplex in which these models cannot be distinguished very well. When comparing predictions of models, one should not force assumptions on the values of parameters, as that procedure can easily lead to wrong conclusions (Birnbaum, 1973a; 1974b).

Priority Heuristic

Although the properties reviewed here were not designed to test the priority heuristic (Brandstätter et al., 2006), every one of these "new paradoxes" violates predictions of that model systematically. According to the priority heuristic, people compare two nonnegative gambles by examining their worst consequences. If these differ by more than 10% of the maximal prize in either gamble, then people supposedly choose the gamble with the higher lowest consequence. However, if the difference is less than this threshold, then they next compare the probabilities to receive the lowest consequences and choose on the basis of this factor if the difference exceeds 0.1. Only if this difference is less than 0.1 do they go on to compare highest consequences, which is decisive in the case of two-branch gambles. With three or more branches, people are supposed to use the probability to win the highest consequence as their final comparison.

The priority heuristic fails to account for any of the new paradoxes. Of the 40 choices in Tables 1, 2, 5, 6, and 7, this heuristic is correct in only 16 cases (40% correct: Choice Problems 1.2, 4, 6.2, 7.2, 8.2, 9.1, 9.2, 10.3, 10.5, 12.3, 12.6, 12.7, 12.8, 14.2, 16.1, 17.2). This heuristic not only failed to predict even half of the modal choices in Birnbaum and Navarrete (1998) but also failed to predict even half of the choices used in a partial replication of that study by Brandstätter, Gigerenzer, and Hertwig (2008). See Birnbaum (2007a; 2008a, 2008b) and Bimbaum and LaCroix (2007) for further discussion of the priority heuristic.

Replication Study

It might be argued that the examples described here have been selected from previous research, and therefore the selection might have capitalized on chance. To check the stability of the results, the choices listed in Table 11 were presented to a new sample of 223 undergraduates, each of whom made each choice twice; the two presentations were separated by intervening tasks requiring about 10 min. The first two columns in Table 11 show the table number and choice number of the item replicated, the third and fourth columns show the choice percentage in the review (Rev) and in the new replication (Rep), respectively. Choice percentages in the replication are quite similar to previous results ($r = .92$). Except for three choices (8.1, 10.3, and 14.2), all of the choice proportions in the replication are significantly different from 0.5, by the z test of correlated proportions (this test compares the number who made two choices of one type with the number who made two choices of the opposite type).

Because there were two presentations of each choice, one can estimate the "true" probabilities of preferring the second gamble ($\hat{p}$) and the "error" rates ($\hat{e}$) for each choice, which are displayed

Table 11
Choice Percentages in Review (Rev) and in Replication (Rep) Study, With Estimates of "True" Choice Probabilities and "Error" Rates, as Well as Choice Patterns

Table no.	Choice no.	Rev	Rep	$\hat{p}$	$\hat{e}$	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	Choice 2
1	1.1	62	66	0.73	0.15	0.26	0.02	0.49	0.23	1.2
1	1.2	26	33	0.24	0.17					
1	2,3.1	73	69	0.77	0.16	0.22	0.00	0.76	0.01	3.2
1	3.2	6	9	0.01	0.08					
1	3.3	57	67	0.78	0.20	0.24	0.00	0.47	0.29	4
1	4	35	34	0.29	0.12					
1	6.1	67	55	0.59	0.19	0.40	0.02	0.40	0.18	6.2
1	6.2	33	29	0.20	0.15					
2	7.1	70	57	0.61	0.19	0.39	0.01	0.52	0.08	7.2
2	7.2	42	19	0.09	0.11					
2	8.1	62	52*	0.54	0.19	0.47	0.00	0.34	0.19	8.2
2	8.2	26	32	0.19	0.21					
2	10.1	38	29	0.16	0.19	0.27	0.54	0.00	0.19	10.2
2	10.2	64	65	0.74	0.19	0.26	0.01	0.36	0.37	10.4
2	10.3	54	51*	0.52	0.17					
2	10.4	43	41	0.37	0.16	0.12	0.50	0.00	0.38	10.5
2	10.5	78	80	0.89	0.11	0.11	0.73	0.00	0.16	10.1
6	13.1	40	34	0.23	0.21	0.42	0.34	0.00	0.24	13.2
6	13.2	62	57	0.61	0.19					
6	14.1	34	43	0.36	0.23	0.54	0.09	0.00	0.37	14.2
6	14.2	51	47*	0.45	0.19					
7	15.1	24	16	0.10	0.08	0.88	0.01	0.02	0.09	15.2
7	15.2	19	15	0.09	0.08					
7	16.1	31	32	0.28	0.11	0.73	0.00	0.12	0.15	16.2
7	16.2	34	20	0.12	0.10					
7	17.1	56	55	0.59	0.19	0.45	0.00	0.23	0.32	17.2
7	17.2	33	35	0.29	0.15					

Note. Asterisk indicates cases in which the choice percentage in the replication study was not significantly different from 50%. In the analysis of Choice Problems 10.5 with 10.1 (of Table 2), 10.1 was treated as the first choice problem and 10.5 as the second, therefore, $\hat{p}_{12} = 0.73$ means that 73% of participants are estimated to truly prefer the first gamble in Problem 10.1 and prefer the second gamble in 10.5.

in the fifth and sixth columns of Table 11. The next four columns in the table show the estimated probabilities of each combination of choices, with the second choice of each pair listed in the last column. For example, the first row in Table 1 shows that for Choice Problems 1.1 and 1.2, which tested coalescing, $\hat{p}_{11} = 0.26$, which means that 26% of participants preferred the first gamble on both choices (A over B and A′ over B′, AA′). However, $\hat{p}_{21} = 0.49$ and $\hat{p}_{12} = 0.02$, which indicate that 49% had the BA′ preference pattern predicted by TAX, and only 2% had the opposite reversal.

Choices 2 (3.1) and 3.2 test stochastic dominance and coalescing. The replication data indicate that 76% have the "true" preference pattern of violating stochastic dominance in the coalesced form and satisfying it in the canonical split form.

The replication used a new version of Choice 3.3, I′ = ($90, 0.8; $13, 0.1; $12, 0.1) and J′ = ($89, 0.7; $88, 0.1; $11, 0.2) in order to provide a test of priority heuristic as well as of the contrast counting and similarity heuristics. In this case, the probabilities of the lowest and highest consequences are 0.1 lower and 0.1 higher in I′ than in J′, respectively, so the priority heuristic and similarity heuristics imply that people should choose I′, as does CPT. In addition, both lowest and highest consequences are higher in I′, so the contrast counting heuristic also predicts that people should choose I′. Despite these changes, the observed percentage of violations of stochastic dominance is 67% in this new choice, with an estimated "true" probability of 0.78.

The replication of Choice 4 of Table 1 yielded 34% violations, very close to the previous value of 35% in Table 1. Comparing Choices 2 (3.1) and 4, the "true and error" model estimates that 47% violated dominance in Choice 2 and satisfied it in Choice 4, with an additional 29% who violated it in both problems. No one was estimated to satisfy dominance in Choice 2 and violate it in Choice 4 (see Figure 7).

In tests of upper tail independence (Choices 6.1 and 6.2), upper cumulative independence (Choices 7.1 and 7.2), and lower cumulative independence, the replication study again confirmed previous results. The model estimated that 40%, 52%, and 34% showed the preference reversals predicted by TAX against only 2%, 1%, and 0% who showed the opposite pattern of reversal in these three properties.

In Choices 10.1 through 10.5, the replication results agreed well with previous results dissecting the Allais paradox. In this case, 73% of participants are estimated to show the predicted Allais paradox, reversing preferences between Choices 10.1 and 10.5, with no one having the opposite pattern. The replication study shows that coalescing is violated by 54% and 50%, as predicted by TAX (10.1 vs. 10.2 and 10.4 vs. 10.5, respectively), with no one having the opposite pattern. In Choices 10.2 versus 10.4, 36% are estimated to have the pattern of violation of RBI predicted by TAX, with only 1% estimated to show the pattern of reversal predicted by CPT with its inverse-S-weighting function.

Problems 13.1 and 13.2 used a revised test of RBI; S = ($44, 0.1; $40, 0.1; $2, 0.8) > R = ($96, 0.1; $4, 0.1; $2, 0.8) on 66% of the presentations, but S' = ($100, 0.8; $44, 0.1; $40, 0.1) > R' = ($100, 0.8; $96, 0.1; $4, 0.1) only 43% of the time. According to the true-and-error model, 34% of the participants had the SR' pattern of preference reversal and 0% showed the pattern predicted by CPT.

Choices 14.1 and 14.2 tested 4-DI. It was estimated that 9% violated 4-DI in the manner predicted by TAX and that no one had the opposite pattern. Although these trends agree in direction with those reported by Birnbaum and Chavez (1997), the effects are smaller than previously reported for the same choice.

Choices 15.1, 15.2, 16.1, and 16.2 tested 3-LDI and 3-2-LDI. The observed choice percentages are similar to previous values. The estimated percentage who reversed preferences as predicted in Choices 15 by prior CPT was 2%, and 1% were estimated to have the opposite reversal. In Problem 16, the true-and-error model estimated that 12% had the reversal pattern opposite that predicted by CPT, and no one showed the reversal predicted by that model with its prior parameters.

In Choices 17.1 and 17.2, observed choice percentages again replicated previous results. The true and error model estimated that 23% had the preference reversal opposite that predicted by CPT and that no one had the reversal predicted by the inverse-S-weighting function.

In summary, the replication study shows that the phenomena reviewed here, collected from different articles with different participants, can be largely replicated in a single study. The only case where the modal choice in the replication study did not match previous results was in Choice 14.2, where 47% chose R' over S', compared with 51% in the original study.

The findings summarized in this article are not equally strong for all of the new paradoxes for two main reasons. First, some tests are inherently more diagnostic than others. Violations of the properties in Table 9 refute CPT with any functions and parameters; therefore, these paradoxes are most telling. In contrast, violations of predicted patterns in Table 10 (e.g., pattern of violation of RBI) rule out CPT with an inverse-S-weighting function, but they do not rule out the general form of CPT; therefore, evidence of this type is inherently less strong. Failure to find a predicted pattern of violations on the basis of prior parameters (as in tests of 3-LDI, Problems 15.1, and 15.2) is the weakest type of evidence. One can always imagine that with some other choices, with other procedures, other participants, or using other parameters, predicted patterns of data might be observed.

Second, the weight of empirical evidence on these different properties varies across the properties. The number of studies testing each property, number of participants in each study, the number and variety of choices used in each test, and proportions of people who are estimated to show each type of violation differ from property to property. I think the evidence is most powerful for violations of coalescing, stochastic dominance, lower, and upper cumulative independence, for which the logical case, the amount of empirical evidence, and the magnitude of violations in individual participants combine to provide the strongest refutations of CPT. In contrast, the property of 4-DI had been tested in only one previous study with a dozen choice problems prior to the replication, and the replication data were not very impressive in the test selected for this review. Because violations of 4-DI refute the RAM model, the case against RAM is not strengthened much by

the replication study. In my opinion, however, the topic that deserves the most empirical attention is gain–loss separability. That property is crucial to both versions of prospect theory and to the issue of loss aversion. I am aware of only two studies on this topic, both finding evidence against the property.

Contextual Effects

The TAX model, as presented so far, is silent on certain phenomena that refute the EU model. For example, Birnbaum (1992a) examined choices between sure cash and binary gambles, while varying the distribution of sure cash values. When the distribution of sure cash values was positively skewed on the interval from $1 to $90 (many small values), the percentage who chose cash over a gamble exceeded the corresponding choice percentages obtained when this distribution was negatively skewed on the same interval. For example, only 33% chose $20 over the gamble G = ($48, 0.05; $0) in the negatively skewed context, whereas 65% chose $20 over the same gamble in the positively skewed context. The resulting cash-equivalent values for a gamble were higher in the negatively skewed context than in the positively skewed context. Such findings create problems for theories in which utility and probability weighting functions are treated as "absolutes," as they were in early versions of EU theory (Stewart, Chater, Stott, & Reimers, 2003).

To account for contextual effects or effects of choice configuration, the TAX model (or any model treated in this article) would require additional features beyond those stated above. In order to accommodate such effects within TAX, the utility and probability functions, $u(x)$ and $t(p)$ might be assumed to follow range-frequency theory (Birnbaum, 1974c, 1982, 1992a, 1992b; Parducci, 1965, 1995). Contextual effects might also affect the transformation from subjective differences to choice proportions (Birnbaum et al., 1971; Mellers & Birnbaum, 1982). This means that the functions should be subscripted for context.

Varey, Mellers, and Birnbaum (1990) asked people to examine squares containing numbers of white and black dots and to judge the proportions of each color. Varey et al. manipulated the actual numbers of white and black dots as well as the frequency distribution of actual proportions. They found that judgments of proportions followed different inverse-S relationships with actual proportions, consistent with the relative ratio, $t_k(w)/[t_k(w) + s_k(b)]$, where $s_k(b)$ and $t_k(w)$ are contextually affected, subjective numbers of black and white dots in context k, respectively. Figure 8 in Varey et al. (1990) shows how the relationship between judged and actual proportion depended on their contextual manipulations. It seems likely that similar contextual effects would be observed if people were asked to bet on a white or black ball drawn from urns represented by their stimuli or to choose between gambles defined by those stimuli and presented in those same contexts.

Roe, Busemeyer, and Townsend (2001) reviewed a number of studies in which the choice between two multiattribute alternatives A and B depends on the context provided by a third alternative, C, added to the choice set. It is possible to find a third alternative, C, such that the probability of choosing B from the set of three alternatives $\{A, B, C\}$ exceeds the probability of choosing B over A from the set of two, $\{A, B\}$. This is done by introducing C that is dominated by B on all dimensions and is not dominated by A (Huber, Payne, & Puto, 1982). In the set of three alternatives, A is

now third on the first dimension and first on the second dimension, whereas *B* is now first on the first dimension and second on the second dimension. Roe et al. (2001) developed a multialternative choice model to account for this effect and other contextual effects such as the similarity effect (Restle & Greeno, 1970).

The existence of contextual effects, however, need not rule out the idea that there is a context-free psychophysical function underlying judgments and choices. Birnbaum (1974c) showed that one can use range-frequency theory to fit judgments of the "size" of three-digit numbers in which the numbers were presented in different distributions. Despite large contextual effects due to manipulation of the frequency distribution, the range-frequency model yields a context-free scale of numerical magnitude. Furthermore, this scale of numerical magnitude agreed well with subjective scales of the same range of numbers, derived from a simple model of judged "ratios" and "differences" of numerical magnitude (Rose & Birnbaum, 1975). In this case, all three sets of data could be fit by the approximation, $v(x) = x^\beta$, where $\beta = 0.6$, where $v(x)$ represents the subjective magnitude of a number. Roe et al. (2001) concurred with this view; they showed that their contextual effects can also be described by a context-free psychophysical function combined with a theory of choice that accounts for the contextual effects.

Open Problems

The TAX model is an idempotent, rank-weighted utility model; this generic model has been axiomatized and related to other formulations by Marley and Luce (2001, 2005; see also Luce & Marley, 2005). It is not yet known, however, what additional assumptions force the special form of the TAX model, in which all of the weight transfers are equal proportions of the branch losing weight. This special model implies 3-LDI. When two upper branches have equal probability, the middle branch will gain as much from the highest branch as it gives up to the lowest branch, which keeps the ratio of weights independent of the value of that (common) probability. It has not yet been shown whether this theorem could be used as an axiom in combination with the idempotent, rank-weighted utility model to deduce the special TAX model. If these premises are not sufficient, then what assumptions would be required to imply this representation?

The properties reviewed here have been analyzed at both the group and individual level. What is characteristic of group data has been shown (in the original reports cited) to represent the behavior of individuals (see, e.g., Birnbaum & Navarrete, 1998). However, each study has involved a relatively small number of choices (typically from 20 to 130) for each participant and has tested only a small number of properties in each person tested. What has not yet been done is the assessement of the behavior of individuals over a large number of replications with a larger number of properties tested within the same person. With larger quantity of data per person, it becomes possible to analyze the data at the level of each individual and to ask, for example, whether there are some participants who are best represented by CPT and others who are best represented by TAX.

Until researchers have a clear reason to abandon it (as in Mellers et al., 1992), I think we should retain the principle that every individual is represented by the same model in which different people can have different parameters in that model. Because EU is a special case of TAX, it is compatible with TAX that some individuals might satisfy EU, but there should not be any participant who systematically violates EU and who satisfies CPT, except by random error. How would such a participant be identified? That person should not show any of the paradoxes of Table 9, should show violations of the properties in Table 10, should show Allais paradoxes, and these violations should all be consistent with the same decumulative probability weighting function. "Same" does not mean that all participants have the same weighting function, but rather that all of the violations of properties in Table 10 by a given person should agree with the same weighting function for that individual.

During a long experiment, it seems possible that people might learn to recognize stochastic dominance. Perhaps they might learn or could be taught how to convert a choice into its canonical split form, in which form stochastic dominance is largely satisfied. If so, then would that same person also come to satisfy upper and lower cumulative independence? That person might therefore still violate EU because violations of branch independence are observed in canonical split form. But would these people eventually come to satisfy CPT? Are there any individuals who systematically violate RBI with the *RS′* pattern of violations and show the standard Allais paradoxes? If so, then such findings could be taken as evidence against the special TAX model, and taken as evidence favoring the CPT model if the same weighting function predicted these results correctly as well as tests of distribution independence.

According to the TAX model, the pattern of violations of different properties should be linked according to the person's parameters of TAX. By testing a larger number of properties for each person, it would be possible to test whether the same parameters can explain all of that person's data. If it can be shown that different experimenters using different contexts and different participants find different values of parameters, then it would not be particularly surprising because neither TAX nor CPT assumes that utility functions and probability weighting functions are "absolutes." The advantage of collecting rich data for the same person in the same lab and in the same context is that researchers can test whether there is or is not a common set of parameters that accounts for multiple phenomena in the same person.

References

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

Abdellaoui, M. (2002). A genuine rank-dependent generalization of the von Neumann-Morgenstern expected utility theorem. *Econometrica, 70,* 717–736.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science, 53,* 1659–1674.

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.

Baltussen, G., Post, T., & van Vliet, P. (2006). Violations of cumulative

prospect theory in mixed gambles with moderate probabilities. *Management Science, 52,* 1288–1290.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16,* 215–233.

Battalio, R., Kagel, J. H., & Jiranyakul, K. (1990). Testing between alternative models of choice under uncertainty: Some initial results. *Journal of Risk and Uncertainty, 3,* 25–30.

Baucells, M., & Heukamp, F. H. (2006). Stochastic dominance and cumulative prospect theory. *Management Science, 52,* 1409–1423.

Becker, J., & Sarin, R. (1987). Lottery dependent utility. *Management Science, 33,* 1367–1382.

Bernoulli, D. (b. L. S. (1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22,* 23–36. (Translation of Bernoulli, D., 1738.)

Birnbaum, M. H. (1973a). The Devil rides again: Correlation as an index of fit. *Psychological Bulletin, 79,* 239–242.

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

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

Birnbaum, M. H. (1974b). Reply to the Devil's advocates: Don't confound model testing and measurement. *Psychological Bulletin, 81,* 854–859.

Birnbaum, M. H. (1974c). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15,* 89–96.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Eds.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H. (1992a). Should contextual effects in human judgment be avoided? [Review of E. C. Poulton, Bias in quantifying judgments]. *Contemporary Psychology, 37,* 21–23.

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

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

Birnbaum, M. H. (1999a). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods, 4,* 243–249.

Birnbaum, M. H. (1999b). 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. (1999c). 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. (2001). A Web-based program of research on decision making. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 23–55). Lengerich, Germany: Pabst Science Publishers.

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48,* 87–106.

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

Birnbaum, M. H. (2005a). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty, 31,* 263–287.

Birnbaum, M. H. (2005b). Three new tests of independence that differentiate models of risky decision making. *Management Science, 51,* 1346–1358.

Birnbaum, M. H. (2006). Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology, 27,* 737–761.

Birnbaum, M. H. (2007a). *Testing lexicographic semi-orders as models of decision making: Priority dominance, dimension integration, dimension interaction, and transitivity*. Unpublished manuscript, California State University, Fullerton.

Birnbaum, M. H. (2007b). Tests of branch splitting and branch-splitting independence in Allais paradoxes with positive and mixed consequences. *Organizational Behavior and Human Decision Processes, 102,* 154–173.

Birnbaum, M. H. (2008a). Evaluation of the priority heuristic as a descriptive model of risky decision making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review 115,* 253–262.

Birnbaum, M. H. (2008b). Postscript: Rejoinder to Brandstätter et al. (2008). *Psychological Review, 115,* 260–262.

Birnbaum, M. H., & Bahra, J. (2007). Gain-loss separability and coalescing in risky decision making. *Management Science, 53,* 1016–1028.

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

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

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

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104,* 97–112.

Birnbaum, M. H., & Jou, J. W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22,* 184–210.

Birnbaum, M. H., & LaCroix, A. R. (2008). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes, 105,* 122–133.

Birnbaum, M. H., & Martin, T. (2003). Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on decision research* (pp. 84–107). New York: Cambridge University Press.

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

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

Birnbaum, M. H., Parducci, A., & Gifford, R. K. (1971). Contextual effects in information integration. *Journal of Experimental Psychology, 88,* 158–170.

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

Birnbaum, M. H., Wong, R., & Wong, L. (1976). Combining information from sources that vary in credibility. *Memory & Cognition, 4,* 330–336.

Birnbaum, M. H., & Yeary, S. (2001). *Tests of stochastic dominance and cumulative independence in buying and selling prices of gambles.* Manuscript in preparation.

Birnbaum, M. H., Yeary, S., Luce, R. D., & Zhou, L. (2001). *Empirical evaluation of theories for buying and selling prices of binary gambles.* Manuscript in preparation.

Birnbaum, M. H., & Zimmermann, J. M. (1998). Buying and selling prices of investments: Configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes, 74*(2), 145–187.

Blavatskyy, P. R. (2005). Back to the St. Petersburg paradox? *Management Science, 51,* 677–678.

Bleichrodt, H., & Luis, P. J. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science, 46,* 1485–1496.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113,* 409–432.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky choice with heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willemsen (2008) and Rieger and Wang (2008). *Psychological Review, 115,* 281–289.

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. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty, 2,* 61–104.

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. (1998). Bounded rationality in individual decision making. *Experimental Economics, 1,* 163–183.

Carbone, E., & Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty, 20,* 161–176.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica, 51,* 1065–1092.

Chew, S. H., Epstein, L. G., & Segal, U. (1991). Mixture symmetry and quadratic utility. *Econometrica, 59,* 139–163.

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

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69,* 109–135.

Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2002). *Combining a theoretical prediction with experimental evidence to yield a new prediction: An experimental design with a random sample of tasks.* Unpublished manuscript, Columbia University and Faculty of Industrial Engineering and Management, Techion, Haifa, Israel.

Ert, E., & Erev, I. (in press). The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic. *Journal of Economic Psychology.*

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.

Fox, C. R., & Hadar, L. (2006). Decisions from experience = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making, 1,* 159–161.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38,* 129–166.

Gonzalez, R., & Wu, G. (2003). *Composition rules in original and cumulative prospect theory.* Manuscript in preparation.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review, 109,* 137–155.

Green, J. R., & Jullien, B. (1988). Ordinal independence in nonlinear utility theory. *Journal of Risk and Uncertainty, 1,* 355–387.

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utiliity theories. *Econometrica, 62,* 1251–1290.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choices. *Psychological Science, 15,* 534–539.

Hey, J. D. (2005). Comparing theories: What are we looking for? In U. Schmidt & S. Traub (Eds.), *Advances in public economics: Utility, choice, and welfare: A Festschrift for Christian Seidl* (pp. 213–234). Dordrecht, the Netherlands: Springer.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62,* 1291–1326.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9,* 90–98.

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

Humphrey, S. J. (1998). More mixed results on boundary effects. *Economics Letters, 61,* 79–84.

Humphrey, S. J. (2000). The common consequence effect: Testing a unified explanation of recent mixed evidence. *Journal of Economic Behavior and Organization, 2000*(41), 239–262.

Humphrey, S. J. (2001a). Are event-splitting effects actually boundary effects? *Journal of Risk and Uncertainty, 22,* 79–93.

Humphrey, S. J. (2001b). Non-transitive choice: Event-splitting effects or framing effects? *Economica, 68,* 77–96.

Humphrey, S. J., & Verschoor, A. (2004). The probability weighting function: Experimental evidence from Uganda, India and Ethiopia. *Economics Letters, 84,* 419–425.

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, computational model of preference reversal phenomena. *Psychological Review, 112,* 841–861.

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist, 58,* 723–730.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the coarse theorem. *Journal of Political Economy, 98,* 1325–1348.

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

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

Karmarkar, U.S. (1979). Subjectively weighted utility and the Allais paradox. *Organizational Behavior and Human Performance, 24,* 67–72.

Karni, E., & Safra, Z. (1987). "Preference reversal" and the observability of preferences by experimental methods. *Econometrica, 55,* 675–685.

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

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review, 107,* 62–100.

Levy, M., & Levy, H. (2002). Prospect Theory: Much ado about nothing. *Management Science, 48,* 1334–1349.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum.

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

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. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review, 101,* 271–277.

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. (2001). Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology, 45,* 167–179.

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

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

Luce, R. D., & Marley, A. A. J. (2005). Ranked additive utility representations of gambles: Old and new axiomatizations. *Journal of Risk and Uncertainty, 30,* 21–62.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology, 29,* 1–72.

Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (2006). *Utility of gambling: Entropy-modified linear weighted utility.* Manuscript in preparation.

Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (in press). Utility of gambling II: Risk, paradoxes, and data. *Economic Theory.*

Machina, M. J. (1982). Expected utility analysis without the independence axiom. *Econometrica, 50,* 277–323.

Machina, M. J. (1987, May 1). Decision-making in the presence of risk. *Science, 236,* 537–543.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy, 60,* 151–158.

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

Marley, A. A. J., & Luce, R. D. (2005). Independence properties vis-à-vis several utility representations. *Theory and Decision, 58,* 77–143.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 582–601.

Mellers, B. A., Ordóñez, L., & Birnbaum, M. H. (1992). A change-of-process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes, 52,* 331–369.

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

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.

Parducci, A. (1995). *Happiness, pleasure, and judgment.* Mahwah, NJ: Erlbaum.

Payne, J. W. (2005). It is whether you win or lose: The importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty, 30,* 5–19.

Prelec, D. (1998). The probability weighting function. *Econometrica, 66,* 497–527.

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

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

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model.* Boston: Kluwer Academic.

Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology.* Reading, MA: Addison-Wesley.

Rieger, M. O., & Wang, M. (2008). What is behind the priority heuristic: A mathematical analysis and comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review, 115,* 274–280.

Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology, 103,* 171–173.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multiattribute decision field theory: A dynamic, connectionist model of decision making. *Psychological Review, 108,* 370–392.

Rose, B. J., & Birnbaum, M. H. (1975). Judgments of differences and ratios of numerals. *Perception & Psychophysics, 18,* 194–200.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica, 57,* 571–587.

Schmidt, K. U., & Zank, H. (2005). What is loss aversion? *Journal of Risk and Uncertainty, 30,* 157–167.

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. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypotheses. *Annals of Operations Research, 19,* 79–101.

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

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

Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General, 132,* 23–46.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32,* 101–130.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1,* 39–60.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286. (Reprinted 1994)

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

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

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*(4), 1039–1061.

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

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

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

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 613–625.

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

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

von Winterfeldt, D. (1997). Empirical tests of Luce's rank- and sign-dependent utility theory. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 25–44). Mahwah, NJ: Erlbaum.

Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. *Theory and Decision, 36,* 1–44.

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

Wakker, P. (2001). Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle. *Econometrica, 69,* 1039–1075.

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

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

Weber, E. U. (1994). From subjective probabilities to decision weights: The effects of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 114,* 228–242.

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

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

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

Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty, 16,* 115–139.

Wu, G., & Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. *Management Science, 45,* 74–85.

Wu, G., & Markle, A. B. (2005). *An empirical test of gain-loss separability in prospect theory.* Manuscript in preparation.

Wu, G., & Markle, A. B. (in press). An empirical test of gain-loss separability in prospect theory. *Management Science.*

Wu, G., Zhang, J., & Abdelloui, M. (2005). Testing prospect theories using probability tradeoff consistency. *Journal of Risk and Uncertainty, 30,* 107–131.

Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 399–423). Oxford, England: Oxford University Press.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica, 55,* 95–115.