The Journal of Risk and Uncertainty, 31:3; 263–287, 2005

A Comparison of Five Models that Predict Violations of First-Order Stochastic Dominance in Risky Decision Making*

MICHAEL H. BIRNBAUM mbirnbaum@fullerton.edu
Department of Psychology, CSUF H-830M, P.O. Box 6846, Fullerton, CA 92834-6846

Abstract

Five descriptive models of risky decision making are tested in this article, including four quantitative models and one heuristic account. Seven studies with 1802 participants were conducted to compare accuracy of predictions to new tests of first order stochastic dominance. Although the heuristic model was a contender in previous studies, it can be rejected by the present data, which show that incidence of violations varies systematically with the probability distribution in the gambles. The majority continues to violate stochastic dominance even when two of three branches have higher consequences in the dominant gamble, and they persist in mixed gambles even when probability to win is higher and probability to lose is lower in the dominant gamble. The transfer of attention exchange model (TAX) was the most accurate model for predicting the results.

Keywords: coalescing, cumulative prospect theory, event-splitting, expected utility, nonexpected utility, rank dependent utility, stochastic dominance

JEL Classification: C91, D81

1. Introduction

Birnbaum (1997, p. 94) created a method for producing significant majority violations of stochastic dominance. According to either his rank affected multiplicative weights (RAM) or transfer of attention exchange (TAX) models, people should violate stochastic dominance in choices such as the following: From which urn would you prefer to draw a marble at random, if the color of marble drawn determines your prize?

```
A:  90 red marbles to win $96     B:  85 green marbles to win $96
    05 blue marbles to win $14        05 black marbles to win $90
    05 white marbles to win $12       10 yellow marbles to win $12
```

When the probability of winning prize x or greater given gamble A is greater than or equal to the probability to win x or more in gamble B, for all x, and if this probability is

*Support was received from National Science Foundation Grants, SBR-9410572, SES 99-86436, and BCS-0129453.

strictly higher for at least one value of x, we say that gamble A stochastically dominates gamble B. In this example, A dominates B. The probability to win \$96 or more is .90 in A, and only .85 in B; the probability to win \$90 or more is the same, the probability to win \$14 or more is higher in A than B; and the probability to win \$12 or more is the same in both gambles.

Birnbaum (1997) noted that his models, as fit to data of Tversky and Kahneman (1992) predicted that people should choose B over A. He devised this test as a way to compare his class of older "configural weight" models against the class of rank dependent utility (RDU) and cumulative prospect theories (CPT), which allow no violations of stochastic dominance (Quiggin, 1993; Tversky and Kahneman, 1992), apart from those produced by "chance" or "error." Configural weight models allow weight to depend on rank (Birnbaum, 1974; Birnbaum and Stegner, 1979); but they differ from RDU in that these older models imply violations of stochastic dominance in specific circumstances.

Birnbaum and Navarrete (1998) tested this prediction. About 70% of 100 undergraduates chose B in this choice and others based on the same recipe. Subsequent studies confirmed that significantly more than 50% of undergraduates violate stochastic dominance in choices like this (Birnbaum, 2004a, 2004b; Birnbaum, Patton, and Lott, 1999; Birnbaum, 1999b; Birnbaum and Martin, 2003). The rate of violation is lower in more highly educated people, but it is still substantial (about 50%) among doctorates (Birnbaum, 1999b). According to the configural weight models, the violations are caused by failure of a simpler property, coalescing, described in Section 1.1.

The present series of studies compares four quantitative models, RAM, TAX, lower gains decomposition utility (GDU) of Marley and Luce (2001), and Viscusi's (1989) prospective reference theory (PRT) with a "heuristic" model, the *consequence counting heuristic*. According to this heuristic, the decision-maker counts consequences on the branches, and chooses the gamble with the greater number of branches having higher consequences. Note that B and A both have branches leading to prizes of \$96 and \$12, but B has a branch to win \$90, whereas A has a branch to win only \$14. If the decision maker ignored probabilities, and counted the number of branches with higher consequences in each gamble, then B might seem better than A, even though A dominates B.

The rest of this paper is organized as follows. Section 1.1 describes simple behavioral properties that underlie stochastic dominance, Section 2 presents the five models to be compared, and Section 3 shows how models fit to previous data can be used to predict results of new manipulations. Sections 4–7 describe five new studies of stochastic dominance in choices between gambles with non-negative consequences. These studies systematically manipulate features of choices testing stochastic dominance. In Section 8, choices that tested CPT and EU in the first five studies are analyzed. In Section 9, TAX, RAM, GDU, and PRT are fit by estimating parameters from the combined data. Section 10 reports studies 6 and 7, which test stochastic dominance in five-branch mixed gambles. These studies pit Payne's (2005) probability to win or lose against branch-splitting. Section 11 is the discussion, which concludes that empirical results systematically violate CPT and the heuristic model and that they provide no support for Payne's (2005) argument. Among the models that can account for violations of stochastic dominance, TAX was most accurate with the fewest free parameters.

1.1. The roots of stochastic dominance

Let G_0, = ($96, .9; $12, .1) represent a two-branch gamble with a probability of .9 to win $96 and a probability of .1 to win $12. The gamble G_1, = ($96, .9; $12, .05; $12, .05) represents a three-branch gamble with branches of .9 to win $96, .05 to win $12, and .05 to win $12. Although these gambles are the same in CPT, they have different values in RAM, TAX, GDU and PRT.

It is useful to decompose stochastic dominance into thee simpler premises (Birnbaum and Navarrete, 1998). If people satisfy *transitivity*, *coalescing*, and *consequence monotonicity*, then they will not violate first-order *Stochastic Dominance*.

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

Coalescing is the assumption that if a gamble has two (probability-consequence) *branches* yielding the identical consequence, those branches can be combined by adding their probabilities without affecting the utility. For example, if $G =$ ($100, .2; $100, .2; $0, .6), then $G \sim G' =$ ($100, .4; $0, .6), where $\sim$ denotes indifference. GDU satisfies *upper coalescing*, $G = (x, p; x, q; y, 1-p-q) \sim G' = (x, p+q; y, 1-p-q)$, where $x > y > 0$, as illustrated in the above example, but it violates *lower coalescing* which implies that $G'' = (x, p; y, 1-p) \sim G''' = (x, p; y, 1-p-q; y, q)$, where $x > y > 0$.

Violations of coalescing combined with transitivity are termed *event-splitting effects* (Humphrey, 1995; Luce, 1998; Birnbaum, 1999a, 1999b; Starmer, 2000; Starmer and Sugden, 1993). Kahneman and Tversky (1979) assumed an editing rule of "combination," prior to evaluation of gambles (see also Kahneman, 2003) which implies coalescing, and coalescing is implied by the rank dependent utility representation in CPT (Birnbaum and Navarrete, 1998).

Consequence monotonicity is the assumption that if one consequence in a gamble is improved, holding everything else constant, the gamble should be improved. For example, $F =$ ($96, .9; $14, .10) should be preferred to $G =$ ($96, .9; $12, .10), because $14 is preferred to $12.

Stochastic Dominance is the relation between non-identical gambles, such that $P(x > t|G) \geq P(x > t|F)$ for all t. This relation is also known as "First-order Stochastic Dominance" and it should not be confused with other relations described as types of "stochastic dominance.[1] "

The assertion that *choices satisfy (first-order) stochastic dominance* means that if G dominates F, then $G \succ F$. We need to decide whether observed violations are due to "chance or errors" or are "real." A very lenient standard for satisfaction (stringent for rejection) has been used in empirical tests; namely, if G dominates F, then the probability of choosing F over G should be less than or equal to 1/2. This lenient standard makes a very conservative test; nevertheless, even by this conservative test, stochastic dominance

[1]For example, Levy and Levy (2002) criticized CPT on the basis of a relation described as "stochastic dominance," but Wakker (2002) defended CPT by showing that it predicted violations of the relation tested by Levy and Levy. However, CPT, rank dependent utility, and rank- and sign-dependent utility must satisfy first-order stochastic dominance (Birnbaum and Navarrete, 1998).

has been rejected in a number of studies. A less conservative test uses a "true and error" model fit to replicated data to estimate the rate of "true" violation (Birnbaum, 2004b).

Birnbaum (1997) constructed the following recipe in which his RAM and TAX models should violate stochastic dominance: Start with binary gamble, $G = (x, p; y, 1 - p)$, where $x > x^- > y^+ > y > 0$. Split the upper branch of G [i.e., (x, p)] into $(x, p - r)$ and (x, r), and reduce the consequence on the splinter slightly, creating $G^- = (x, p-r; x^-, r; y, 1-p)$; finally, split the lower branch of G and increase the consequence on the splinter slightly, $G^+ = (x, p; y^+, q; y, 1 - p - q)$. RAM and TAX violate coalescing; splitting the upper branch of G to create G^- improves it, and the splitting of the lower branch of G to create G^+ makes that gamble worse. The first example of this paper illustrates this recipe with $x = \$96,\ x^- = \$90,\ y^+ = \$14,\ y = \$12$; and $p = 0.9,\ r = q = 0.05$.

Transitivity, coalescing, and consequence monotonicity imply satisfaction of stochastic dominance in this test (Birnbaum and Navarrete, 1998). Because people show systematic violations, at least one of these three assumptions must be empirically false. Because the class of RDU and CPT models assume or imply these three principles, they cannot explain systematic violations of stochastic dominance (Birnbaum and Navarrete, 1998).

1.2. Theoretical importance of violations of stochastic dominance

Systematic violations of first-order stochastic dominance refute a large class of descriptive models including rank-dependent utility (RDU) theory (Diecidue and Wakker, 2001; Quiggin, 1985; 1993), rank and sign dependent utility (RSDU) theory (Luce, 2000; Luce and Fishburn, 1991; 1995), cumulative prospect (CPT) theory (Gonzalez and Wu, 1999; Starmer and Sugden, 1989; Tversky and Kahneman, 1992; Tversky and Wakker, 1995; Wakker and Tversky, 1993; Wu and Gonzalez, 1996), lottery dependent utility theory (Becker and Sarin, 1987), aspiration level theory (Lopes and Oden, 1999; Payne, 2005), and generalized utility theory (Machina, 1982). See Weber (1994), Starmer (2000), Luce (2000), and Luce and Marley (2005) for reviews.

RAM, TAX, GDU, and PRT obey transitivity and consequence monotonicity, but they violate coalescing, and therefore they predict violations of stochastic dominance in Birnbaum's recipe. A fifth model, the consequence counting heuristic, will also be tested. One of the strategies that will be used is to compare models using parameters estimated from previous data to predict new results. Models will also be compared based on estimation of parameters from the data.

2. Five models of risky decision making

2.1. Rank affected multiplicative (RAM) model

The RAM model for gambles, $G = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n);\ x_1 > x_2 > \cdots > x_n > 0$, can be written as follows:

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

where RAM(G) is the utility of gamble G, $a(i, n)$ are the rank-affected branch weights, $s(p)$ is a function of p, and $u(x)$ is the utility of consequence x. The RAM model takes its name from the separable (multiplicative) relationship between functions of the probability of a branch and the rank of the branch's consequence. Note that if $s(p) = p$ and $a(i, n) = 1 \; \forall \; i, n$, RAM reduces to EU.

In the domain of pocket cash ($0 < x <$ \$150) RAM can be fit with $u(x) = x^{\beta}$, where $\beta = 1$. It is important to keep in mind, however, that in RAM (and TAX), the assumption of linear utility does not imply risk neutrality. Even with small cash prizes, a majority of undergraduates typically exhibits *risk aversion*, which refers to systematic preference for sure cash over gambles with the same or higher expected values. In both RAM and TAX models, it is the greater weight of branches leading to lower consequences that accounts for risk aversion. For undergraduates, the RAM weights can be approximated as equal to their ranks; i.e., $a(i, n) = i$. For example, in a two-branch gamble, the branch leading to the highest consequence has a weight of 1, the lower branch has a weight of 2. With these parameters, the gamble F = (\$100, 0.5; \$0, 0.5) has a certainty equivalent value of \$33.33, (i.e., $F \sim$ \$33.33). Thus, RAM predicts that people will exhibit risk aversion by choosing \$40 for sure rather than gamble F, even though the gamble's expected value is \$50.

To approximate data of Tversky and Kahneman (1992), let $s(p) = p^{\gamma}$, where $\gamma = .6$. The "prior" RAM model refers to Equation (1) with these parameters: $\beta = 1$; $\gamma = 0.6$; $a(i, n) = i \; \forall \, i, n$. The term, "prior" indicates that the parameters have been estimated from previous data, rather than fit to the data at hand. For binary gambles, $G = (x,\ p;\ 0,\ 1-p)$, prior RAM, like CPT, implies that certainty equivalents will be an inverse-S function of probability: $\text{RAM}(G) = u(x)p^{\gamma}/[p^{\gamma} + 2(1-p)^{\gamma}]$. Like CPT, this model implies risk-seeking for small p. For example, G = (\$100, 0.05; \$0, 0.95) $\sim$ \$7.87, which exceeds the expected value of G, \$5.00.

Unlike CPT, however, this model violates coalescing, and that is why RAM violates stochastic dominance in Birnbaum's (1997) recipe. For example, let G_0 = (\$96, 0.9; \$12, 0.1). Prior RAM predicts that this gamble has certainty equivalent = 0.651 · \$96 + 0.349 · \$12 = \$66.72. Splitting the lower branch (\$12, 0.1) increases the total relative weight of the lower splinters G' = (\$96, 0.9; \$12, 0.05; \$12, 0.05) $\sim$ 0.531·\$96+0.188·\$12+0.281·\$12 = \$56.62. Note that the relative weight of \$12 is now 0.469 instead of 0.349. Splitting the upper branch, however, we have G'' = (\$96, 0.85; \$96, 0.05; \$12, 0.1) $\sim$ 0.455 · \$96 + 0.166 · \$96 + 0.378 · \$12 = \$64.22. The effect of splitting outweighs small changes in the consequence on the middle splinter, so we have, G^- = (\$96, 0.85; \$90, 0.05; \$12, 0.1) $\sim$ \$63.23 $>$ G^+ = (\$96, 0.9; \$14, 0.05; \$12, 0.05) $\sim$ \$56.99. Thus, RAM predicted the violation of stochastic dominance in Birnbaum and Navarrete (1998).

It is important to keep in mind that rank weights in RAM (and TAX) are a function of ranks of consequences on the discrete branches. For example, in a three-branch gamble, there are exactly three ranks, 1, 2, and 3, for branches with the highest, middle, and lowest-consequences, respectively, independent of branch probabilities. With weights set equal to their ranks, the lowest consequence has 3 times the weight of the highest consequence and the middle consequence has twice the weight of the highest consequence, if the probabilities of the three branches were equal.

2.2. Transfer of attention exchange (TAX) model

In the transfer of attention exchange model, weights of branches of a gamble are dependent on branch probabilities; however, instead of each branch having a separable rank weight as in RAM, branches compete for attention in TAX. These transfers of attention are represented by transfers of weight from branch to branch. For people who are risk averse, attention is drawn from branches leading to higher consequences and transferred to branches leading to lower consequences. The "special" TAX model for $G = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$; $x_1 > x_2 > \cdots > x_n > 0$ assumes that all weight transfers for a given number of branches are the same proportion of (transformed) branch probability:

$$\mathrm{TAX}(G) = \frac{\sum_{i=1}^{n}\left[t(p_i) + \frac{\delta}{n+1}\sum_{j=1}^{i-1} t(p_j) - \frac{\delta}{n+1}\sum_{j=i+1}^{n} t(p_i)\right]u(x_i)}{\sum_{i=1}^{n} t(p_i)} \tag{2}$$

In this case, it is assumed that $\delta \geq 0$.[2]

To understand Equation (2), it helps to examine separately the three terms in the numerator affecting weight. First, apart from configural transfers of weight, branch weights depend on branch probabilities. If the configural weight parameter, δ, is 0, then the formula simplifies, and weight of a branch is a function of branch probability, $t(p)$, divided by the total transformed probability of all branches. When $\delta = 0$ and $t(p) = p$, the model reduces to EU.

Second, when $\delta > 0$, the weight of each branch is increased by the second term, which is a proportion of the weight of branches with higher-valued consequences. Third, each branch's weight is reduced by transferring a portion of its weight to all branches having lower consequences. Note that in the third term, probability is a constant, so this term can also be written, $\frac{\delta t(p_i)}{n+1}\sum_{j=i+1}^{n} 1 = \frac{\delta t(p_i)(n-i)}{n+1}$.

TAX can be fit with $u(x) = x$ for cash in the domain of pocket money ($0 < x < \$150$). As in the case of RAM, this utility function does not guarantee risk neutrality. If $\delta > 0$, TAX predicts risk aversion. With $t(p) = p^{.7}$, and $\delta = 1$, TAX approximates the data of Tversky and Kahneman (1992). When $\delta = 1$, it means that in a two-branch gamble, 1/3 of the probability weight of the branch with the higher valued consequence is transferred to the lower-valued branch. Thus, $F = (\$100,\ 0.5;\ \$0,\ 0.5) \sim \$33.33$, so people are predicted to be risk averse for 50-50 gambles. These parameters also imply risk-seeking for gambles with small probabilities to win [e.g., $G = (\$100,\ 0.05;\ \$0,\ 0.95) \sim \$7.53 > \$5.00$].

Prior TAX implies violations of coalescing and therefore violations of stochastic dominance in special choices (Birnbaum, 1997, 2004a). For example, $G^+ = (\$96,\ 0.9;\ \$14,\ 0.05;\ \$12,\ 0.05) \sim \$45.77 < \$63.10 \sim G^- = (\$96,\ 0.85;\ \$90,\ 0.05;\ \$12,\ 0.1)$, violating dominance. In this case, the relative weights in G^+ are 0.395, 0.276, and 0.328 for highest, middle, and lowest branches respectively; and in G^- they are 0.367, 0.259,

[2]This model is equivalent to that in Birnbaum and Navarette (1998); however, the branches have been ordered here from best to worse, to conform to the convention used in CPT, GDU, and other models. Consequently, $\delta > 0$ in this formulation corresponds to $\delta < 0$ in Birnbaum and Navarrrete (1998), and other previous articles that used the opposite convention.

and 0.373, respectively. Note that the total relative weight of the probability of 0.9 in the dominant gamble is 0.395, but the sum of weights of splinters with the same decumulative probability in G^- is $0.367 + 0.259 = 0.626$. Splitting the upper branch of a gamble makes it better, and splitting the lower branch makes it worse.

2.3. *Gains decomposition utility*

Marley and Luce (2001) presented a lower gains decomposition utility (GDU) model in which each gamble can be decomposed into a series of binary gambles to win the lowest consequence or to win a gamble on higher consequences. Binary gambles in this GDU model are represented as follows:

$$GDU(x, p; y) = W(p)u(x) + [1 - W(p)]u(y) \tag{3}$$

where $x > y > 0$. To calculate values of three-branch gambles, the gains decomposition rule (Luce, 2000, pp. 200–202) is applied as follows:

$$GDU(G) = W(p + q)GDU(x, p/(p + q); y) + [1 - W(p + q)]u(z) \tag{4}$$

Expression 4 can be iterated to make predictions for gambles with more than three branches.

To fit the model, let $u(x) = x^{\beta}$, and approximate the weights by the expression developed by Prelec (1998) and by Luce (2000):

$$W(p) = \exp[-\gamma(-\ln p)^{\delta}] \tag{5}$$

This GDU model is similar to TAX for binary gambles when $\beta = 1$, $\gamma = 1.382$, and $\delta = 0.542$. The "prior" GDU model refers to these assumptions and parameter values.[3]

2.4. *Prospective reference theory*

Prospective reference theory (Viscusi, 1989) can be written as follows:

$$PRT(G) = \gamma \sum_{i=1}^{n} p_i u(x_i) + (1 - \gamma) \sum_{i=1}^{n} u(x_i)/n \tag{6}$$

[3]The GDU model tested here uses the same assumptions as in Luce (2000, pp. 200–202). A more general form is axiomatized in Marley and Luce (2001). Marley and Luce (2005) and Luce and Marley (2005) derive relationships among special cases of a generic rank weighted utility model that includes special TAX, GDU, CPT, and PRT as special cases. They axiomatized this class of models and showed that it is equivalent to the generic TAX model, though the special TAX model has not yet been axiomatized. They are currently working with models that use gains decomposition but which do not satisfy Expression 3. These newer models violate idempotence, the assumption that a gamble that always yields the same prize is indifferent to that prize; i.e., $G = (x, p; x, 1 - p) \sim x$, which is implied by Expression 3. The newer models can be tested by investigation of idempotence and its implications, which is not done here. All of the models investigated here satisfy idempotence.

Here the PRT utility of a gamble is a weighted average of its EU and a simple average of the utility of its consequences. This model will be fit with the assumption that $u(x) = x^\beta$.

In order to calculate "prior" predictions, parameters were selected so that this model matches prior TAX for two gambles. Note that in binary gambles with $p = 1/2$, γ drops out of the equation. That is, for $G = (x, 0.5; y, 0.5)$, $PRT(G) = \gamma[u(x) + u(y)]/2 + (1 - \gamma)[u(x) + u(y)]/2$, so $PRT(G) = [u(x) + u(y)]/2$. We can therefore select β such that $F = (\$100,\ 0.5;\ \$0,\ 0.5) \sim \$33.33$, which implies $[100^\beta + 0^\beta]/2 = 33.33^\beta$; therefore, $\beta = 0.631$. Once β is fixed, we can select γ such that $G = (\$100,\ 0.05;\ \$0,\ 0.95) \sim \$7.53$, which implies $\gamma = 0.676$. With these parameters, PRT matches TAX for two binary gambles and like TAX, it implies violations of stochastic dominance in Birnbaum's recipe.

PRT is a special case of RAM in which $a(i, n) = 1\ \forall\ i, n$, and where $s(p) = \gamma p + (1 - \gamma)/n$. PRT is also a special case of TAX in which $\delta = 0$ and $t(p) = \gamma p + (1 - \gamma)/n$. It thus differs from RAM and TAX in that its weights are unaffected by rank and because it uses a linear probability weighting function instead of the power functions used in those models. Whereas the utilities of binary gambles of the form, $H = (x, p; 0, 1 - p)$, are inverse-S functions of p in CPT, RAM, TAX, and GDU, this relationship is positively accelerated in PRT when $u(x) = x^\beta$ if $\beta < 1$.

2.5. Consequence counting heuristic

The *consequence counting heuristic* assumes that people choose between gambles with equal numbers of branches by selecting the gamble with the greater number of branches whose consequences are higher than corresponding consequences on the ranked branches of the other gamble. This heuristic implies that people choose $G-$ over $G+$ because the gambles are equal in their best and worst consequences but differ in the middle branch, where $G-$ has the higher consequence.

3. Predicting violations of stochastic dominance

Consider choices constructed from the following recipe: $G- = (\$96, p, \$90, .05; \$12, .95 - p)$ versus $G+ = (\$96, .90, \$14, .05; \$12, .05)$, which was used in Studies 1 and 2. Using the equations, one can calculate utilities for the gambles for the prior models, and ask, for what value of p is $U(G-) = U(G+)$? In other words, for what value of p will half of the choices satisfy stochastic dominance?

To convert utility differences into choice probabilities, the following logistic function was used, which has one free parameter:

$$P(A, B) = \frac{1}{1 + \exp\{-\alpha[U(A) - U(B)]\}} \tag{7}$$

where α is the parameter relating utility differences to choice proportions. This model implies that if $U(A) - U(B) > 0$ then the probability of choosing A over B exceeds 1/2. In this paper, the value of α is chosen so that each model predicts a choice percentage of 70%

for Birnbaum's (1997) original recipe ($p = 0.85$), as found by Birnbaum and Navarrete (1998). According to these parameters, prior TAX model predicts majority violations when $p \geq 0.5$; prior RAM predicts majority violations when $p \geq .78$, prior PRT requires $p \geq 0.6$, and prior GDU requires $p \geq 0.8$. Therefore, these models can be compared based on the accuracy of their predictions for the effects of manipulating p.

4. Studies 1 and 2: Manipulation of probability

Studies 1 and 2 were designed to determine which of these models does the best job of predicting the incidence of violations as a function of how probabilities are split between branches. For example, consider the following choice:

```
A:  90 red marbles to win $96     B:  75 green marbles to win $96
    05 blue marbles to win $14        05 black marbles to win $90
    05 white marbles to win $12       20 yellow marbles to win $12
```

According to the prior TAX, there should be 64% violations of stochastic dominance whereas RAM predicts "only" 39% violations in this case. According to the counting heuristic, violations of stochastic dominance should be independent of how probabilities are split, so there should be no effect of p. Predictions of EU, RDU, or CPT would all be less than 50%, since these models satisfy stochastic dominance.

4.1. Method of studies 1–2

Participants made 20 or 21 choices between pairs of gambles in Studies 1 and 2. They viewed materials via the Internet, and made each choice by clicking a button beside the gamble they preferred. They were told that 3 lucky participants would be selected to play one of their chosen gambles, with cash prizes as high as $110, so they should choose carefully. Prizes were awarded as promised. The probability mechanism was described as urns containing 100 otherwise identical marbles, from which one marble would be drawn at random, and the color of the selected marble would determine the prize.

There were 330 participants recruited via the Web and from the psychology department's "subject pool" at California State University, Fullerton; there were 98 and 232 in studies 1 and 2.

The seven tests of stochastic dominance are listed in Table 1. Study 2 added Choice #21, not used in Study 1.

There were four "warm-up," choices and 10 "filler" choices. These 14 extra choices were intended to break up the pattern of trials and perhaps prevent participants from adopting a special strategy for this experiment, if every choice had involved stochastic dominance. Fillers were the same as corresponding trials in Birnbaum (1999b), except formatted in terms of the marbles. Complete materials of all studies and raw data can be viewed at the following URL: http://psych.fullerton.edu/mbirnbaum/archive.htm

Table 1. Percentage of violations of stochastic dominance in studies 1 and 2.

No.	Choice *G*+	Choice *G*−	Total $n = 330$	Study 1 $n = 98$	Study 2 $n = 232$
5[a]	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**85** red marbles to win $96 05 blue marbles to win $90 **10** white marbles to win $12	74*	81*	72*
18	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**85** red marbles to win $97 05 blue marbles to win $91 **10** white marbles to win $13	70*	72*	69*
13	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**75** red marbles to win $97 05 blue marbles to win $91 **20** white marbles to win $13	56*	56	55
11[a]	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**65** red marbles to win $96 05 blue marbles to win $90 **30** white marbles to win $12	50	49	50
7	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**55** red marbles to win $97 05 blue marbles to win $91 **40** white marbles to win $13	50	51	49
15[a]	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**45** red marbles to win $96 05 blue marbles to win $90 **50** white marbles to win $12	42*	46	40*
21[a]	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**25** red marbles to win $96 05 blue marbles to win $90 **70** white marbles to win $12	35*		35*

[a]Choices marked *a* denote cases where the dominant gamble was presented in the first position. Asterisks show percentages significantly different from 50%. Bold entries show probabilities manipulated in studies 1–2.

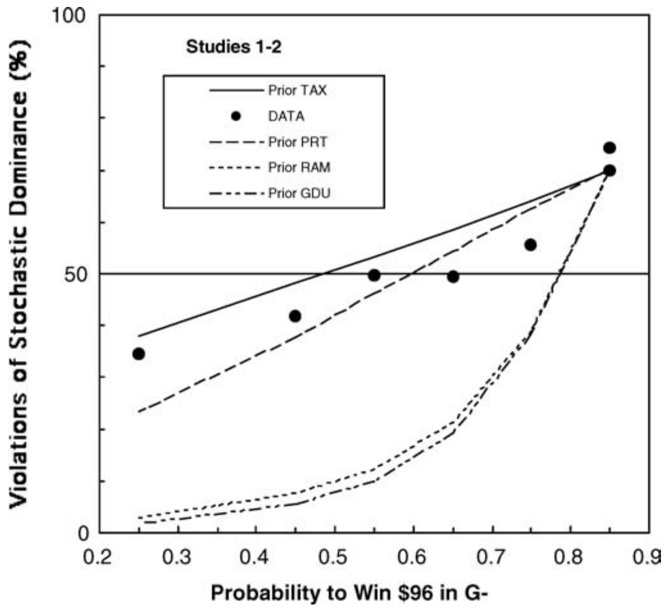

Figure 1. Empirical percentage of violations of stochastic dominance as a function of the probability to win $96 in the dominated gamble, in choices: ($96, .9; $14, .05; $12, .05) versus ($96, .95 – r; $90, .05; $12, r), as a function of .95 – r. Predicted curves show fit of prior models, calibrated to coincide at 70% violations when probability to win $96 is .85 in the dominated gamble.

4.2. Results of studies 1–2

Table 1 shows the percentage of violations of stochastic dominance in each choice. If people ignored probability, the incidence of violations would have been the same from row to row. Instead, the percentage of violations systematically changes from 74.2% to 34.6%. This large and significant decrease refutes the consequence counting heuristic.

For $n = 232$ (Study 2), the binomial distribution with $p = 1/2$ has a mean of 50% and a standard deviation of 3.28%; therefore percentages outside the interval from 43.4% to 56.6% deviate significantly from 50%. The term "significant" and asterisks (*) in Table 1 indicate results outside this 95% confidence interval. In Study 2, the percentage choosing the dominated gamble changed from significantly more than 50% (first two rows) to significantly less than 50% (last two rows) as the probability to win the highest consequence in the dominated gamble is decreased from 0.85 to 0.25.

The binomial test of correlated proportions is a more sensitive test of within-subject *changes* in choice proportions. Differences in percentages in Table 1 are significant by this test.[4]

[4]By the test of correlated proportions, the proportion of violations in Study 2 changes significantly between Choices 5 and 13 ($z = 3.8^*$) and between Choices 18 and 13 ($z = 3.5^*$). Choice 18 was also significantly different from Choice 7 ($z = 4.6^*$), as were Choices 5 and 11 ($z = 5.0^*$), Choices 5 and 15 ($z = 6.81^*$), and Choices 11 and 21 ($z = 4.4^*$). Combining Studies 1 and 2, even the modest difference between Choices 11 and 13 was

Predictions of prior models are shown in Figure 1, along with the observed data. All were calibrated to agree in the prediction of 70% for the rightmost case in Figure 1 ($p = 0.85$ on the abscissa). The prior TAX and PRT models are more accurate than prior RAM or GDU in predicting where the percentage of violations crosses 50% on the ordinate.

5. Study 3: Probability monotonicity

Study 3 investigated another variation of Birnbaum's (1997) recipe, first tested by Martin (1998), and summarized in Birnbaum and Martin (2003). This variation redistributes the .9 probability to win \$96 and \$90 in the dominated gamble, as in the following example:

```
A:  90 red marbles to win $96     B:  15 red marbles to win $96
    05 blue marbles to win $14        75 blue marbles to win $90
    05 white marbles to win $12       10 white marbles to win $12
```

In this variation, the highest consequence in the dominated gamble has transferred probability to the middle branch, but the sum of probabilities of the two highest branches is constant.

Probability monotonicity holds that if probability is transferred from a branch with a lower consequence to a higher valued branch, holding everything else constant, it should improve the gamble. Probability monotonicity follows from transitivity, coalescing, restricted branch independence, and consequence monotonicity. Based on their prior parameters, RAM, TAX, PRT, and GDU disagree on predictions for this manipulation, which is tested in the third experiment. In this case, prior RAM implies a substantial violation of probability monotonicity and the other three models imply small effects. RAM makes this prediction because the shifted probability from the highest to middle branch receives the greater weight of 2 (for the middle branch) as opposed to the weight of 1 it received in the highest branch. This effect is strong enough to outweigh the small change in the value of the consequence. As in Studies 1–2, the heuristic model predicts no effect of redistribution of probability.

Instances of this manipulation were tested by Martin (1998), who found a greater change than predicted by prior TAX, as noted by Birnbaum and Martin (2003, p. 97). Study 3 extends this effect in a parametric variation, to compare the performance of the models for this manipulation.

The method in Study 3 is basically the same as in Studies 1 and 2, except for 7 trials designed to test probability monotonicity, listed in Table 2. There were 428 participants who made 21 choices, including the same 14 warm-ups and "filler" trials as in Study 2.

significant ($z = 2.23^*$), and the proportion of violations in Choice 13 (55.5%) is significantly greater than 50% ($z = 1.98^*$). Significant changes from row to row with the same consequences (e.g., choices 5, 11, 15, and 21) show that people do respond to the probability distribution in this type of choice. Choices 18, 13, and 7 lead to the same conclusion.

Table 2. Percentage of violations of stochastic dominance in Study 3.

No.	Choice G+	Choice G−	Data n = 428
5	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**85** red marbles to win $96 **05** blue marbles to win $90 10 white marbles to win $12	74*
18[a]	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**85** red marbles to win $97 **05** blue marbles to win $91 10 white marbles to win $13	65*
13[a]	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**65** red marbles to win $97 **25** blue marbles to win $91 10 white marbles to win $13	61*
15	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**55** red marbles to win $96 **35** blue marbles to win $90 10 white marbles to win $12	61*
7[a]	90 black marbles to win $97 05 yellow marbles to win $15 05 purple marbles to win $13	**35** red marbles to win $97 **55** blue marbles to win $91 10 white marbles to win $13	57*
21	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**25** red marbles to win $96 **65** blue marbles to win $90 10 white marbles to win $12	53
11	90 red marbles to win $96 05 blue marbles to win $14 05 white marbles to win $12	**15** red marbles to win $96 **75** blue marbles to win $90 10 white marbles to win $12	60*

[a] In choices marked *a*, the dominant gamble was in the second position. Bold entries show probabilities (numbers of marbles out of 100) manipulated in Study 3. All choice percentages are significantly greater than 50% except for choice 21.

Choice proportions from Study 3 are plotted in Figure 2 as a function of the probability to win $96 in the dominated gamble, $G-$. Predictions of prior models are shown as solid or dashed lines. The effect is greater than predicted by the prior TAX model and opposite predictions of prior RAM. Unlike previous findings, this manipulation did not reverse the modal preference, so the effect of the manipulation was less than that reported in Birnbaum and Martin (2003). Perhaps the smaller effect was produced by use of a greater relative proportion of trials of this type here.

The large effect of probability again contradicts the heuristic model. Study 4 manipulates consequences, with probabilities fixed, to see if changing consequences has a greater effect than predicted by quantitative models, which might indicate some partial use of the heuristic.

6. Study 4: Manipulation of two consequences

If people follow a heuristic of counting the number of branches that favor one gamble over another, then perhaps the following manipulation would produce a greater change in choice

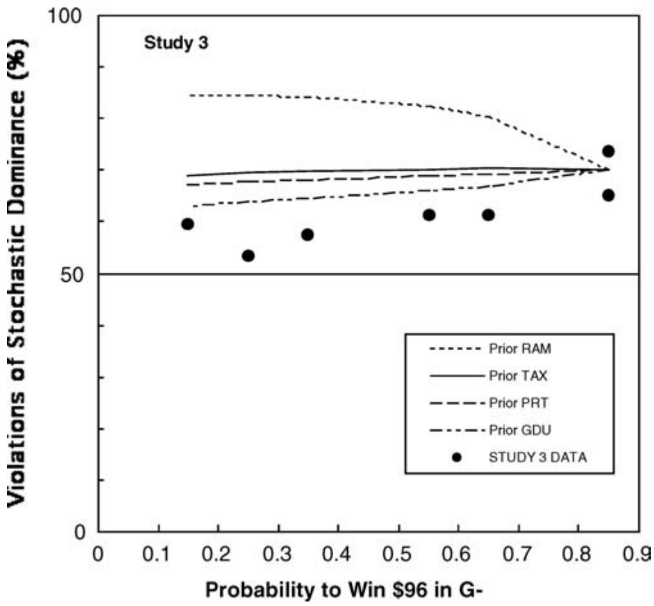

Figure 2. Violations of stochastic dominance, plotted as a function of the probability to win $96 in the dominated gamble, in choices: $G+ = (\$96, .9; \$14, .05; \$12, .05)$ versus $G- = (\$96, .90-r; \$90, r; \$12, .1)$, as a function probability to win $96 in $G-$. Here, prior RAM violates probability monotonicity, contrary to the data.

proportions than predicted by quantitative models that ignore this feature.

```
A:  90 red marbles to win $96      B″:  85 red marbles to win $94
    05 blue marbles to win $14          05 blue marbles to win $90
    05 white marbles to win $12         10 white marbles to win $10
```

In Studies 1–3, there were two branches with equal consequences and the middle branch favored the dominated gamble. However, in the choice between A and B'', Gamble A now has two branches with higher consequences than the corresponding prizes in B'', and B'' has only one. If people follow a heuristic that attends to this count, then the percentage of violations of stochastic dominance should be markedly reduced compared to the choice between A and B. RAM, TAX, PRT, and GDU predict only a small reduction in the value of B''.

This study also manipulated the size of the middle branch of G–. This consequence took on the values of $90, $70, $40, or $20. The purpose of this manipulation was to see if the magnitude of the effect exceeds that predicted by the models, which might be evidence that the heuristic could provide something "extra" to improve the models.

The method was the same as in previous studies, except for the 7 trials presented in Table 3. There were 319 new participants.

The effects of changing consequences on two branches in $G-$ to make them worse than those of $G+$ had very small effects on the choice proportions, not significantly different

Table 3. Violations of stochastic dominance in Study 4.

No.	Choice G+	Choice G−	Data $n = 319$	Prior Predictions TAX	RAM	GDU	PRT
21	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win \$96 05 blue to win \$90 10 white to win \$12	66*	70	70	70	70
7[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win \$97 05 blue to win \$91 10 white to win \$13	73*	70	70	70	70
5	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win **\$94** 05 blue to win \$90 10 white to win **\$10**	65*	68	65	65	68
18[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win **\$95** 05 blue to win \$91 10 white to win **\$11**	68*	68	65	65	68
15	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win \$96 05 blue to win **\$70** 10 white to win \$12	70*	64	60	60	67
13[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win \$97 05 blue to win **\$41** 10 white to win \$13	68*	55	***43***	***44***	62
11	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win \$96 05 blue to win **\$20** 10 white to win \$12	53	49	***32***	***33***	57

[a]In choices marked *a*, the dominant gamble, $G+$, was presented in the second position. All percentages are significantly greater than 50%, except for choice 11. Bold values show consequences manipulated in Study 4. Predicted percentages shown in italics show cases where prior models fail to predict modal choices and the difference exceeds 8%.

from the changes predicted by prior quantitative models (compare choices 21, 7, 5, and 18 in Table 3). In addition, the change due to manipulation of the middle branch (choices 21, 7, 15, 13, and 11) was even less than that predicted by TAX, RAM, or GDU. For example, when the middle branch was $20, TAX, PRT, RAM and GDU predict 49%, 57%, 32% and 33% violations of stochastic dominance, respectively, compared with an observed rate of 53%. The idea of the consequence counting heuristic is that people pay too much attention to the values of consequences, so once again, empirical data provide no support for the idea that the heuristic adds anything to the quantitative models.

7. Study 5: Manipulation of three consequences

Study 4 showed that changes in one or two of the consequences on the branches had effects similar to those predicted by TAX. Study 5 manipulates all three consequences in

$G-$, reducing them to see if the prior models can predict the point where the majority of participants switches from violating to satisfying stochastic dominance. For example, consider:

```
A:  90 black marbles to win $97    B''': 85 red marbles to win $70
    05 yellow marbles to win $15         05 blue marbles to win $60
    05 purple marbles to win $13         10 white marbles to win $2
```

According to prior TAX, 44% of participants should prefer B''' to A, whereas prior PRT, RAM and GDU predicts that only 38%, 11%, or 11% should violate stochastic dominance here, respectively.

The method of Study 5 was the same as in Studies 2–4, except for the 7 choices in Table 4. There were 411 new participants.

Results are shown in Table 4 along with predictions of prior models. All models predict decreasing violations as the consequences of $G-$ are reduced. The data show only

Table 4. Violations of stochastic dominance in Study 5.

	Choice		data	Prior Predictions			
No.	$G+$	$G-$	$n = 411$	TAX	RAM	GDU	PRT
5	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win \$96 05 blue to win \$90 10 white to win \$12	70*	70	70	70	70
21[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win \$97 05 blue to win \$91 10 white to win \$13	69*	70	70	70	70
18	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win **\$90** 05 blue to win **\$84** 10 white to win **\$6**	48	***64***	51	52	***62***
7[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win **\$90** 05 blue to win **\$80** 10 white to win **\$10**	57*	63	50	51	62
11	90 red to win \$96 05 blue to win \$14 05 white to win \$12	85 red to win **\$85** 05 blue to win **\$75** 10 white to win **\$4**	47	***58***	36	37	***56***
15[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win **\$80** 05 blue to win **\$70** 10 white to win **\$5**	42*	***53***	25	26	***50***
13[a]	90 black to win \$97 05 yellow to win \$15 05 purple to win \$13	85 red to win **\$70** 05 blue to win **\$60** 10 white to win **\$2**	31*	44	11	11	38

[a]Choices marked *a* had G+ in the second position. Bold entries show values manipulated in Study 5. Asterisks denote choice percentages significantly different from 50%. Predicted percentages shown in italics show cases where the prior models fail to predict the modal choice and the discrepancy exceeds 8%.

48% violations when all three consequences were reduced by \$6 (choice 18), but 57% violations when the reductions were \$7, \$11, and \$3 from the highest, middle, and lowest branches, respectively (choice 7 in Table 4). This result is significantly greater than 50%, slightly lower than the prediction of 63% by prior TAX and PRT, and higher than the prediction of 50% by RAM. The observed data for other cases also fall intermediate among predictions of TAX, PRT, RAM and GDU. The data again provide no evidence that people pay more attention to the consequences than implied by the prior quantitative models.

8. Analysis of filler trials

Studies 1–5 each had 14 choices in common. Eight of these "filler" trials provide tests of CPT and EU, and three choices assessed risk aversion with $p = 0.01,\ 0.5$, and 0.99; these 11 choices are listed in Table 5. Results were very nearly the same in all five studies, so data in Table 5 are combined over all 1488 participants.

As shown in Birnbaum and Navarrete (1998), CPT implies two properties called upper and lower cumulative independence. In choice 10, we see that 70% chose $R' \succ S'$, so the percentage who choose $R''' \succ S'''$ in choice 9 should be at least 70% according upper cumulative independence; however, only 31% made this choice. Both choice percentages are significantly different from 50% violating CPT, RDU, and EU. Choices 12 and 14 provide another test of upper cumulative independence, with the same conclusion. Percentages in choices 8 and 6 violate lower cumulative independence, as do the results in choices 20 and 17.[5]

With gambles defined as in Footnote 5, restricted branch independence requires that $S \succ R \Leftrightarrow S' \succ R'$. CPT with any weakly inverse-S shaped probability weighting function implies that violations (of branch independence) should be of the form, $S \prec R$ and $S' \succ R'$. However, the data show significantly more violations of the opposite type than of the type that would be consistent with this prediction of CPT. Note that the percentage choosing the risky gamble in choice 10 exceeds that in choice 6, and the same pattern is observed in choices 12 and 17.

Comparing prior models in Table 5, TAX and PRT each have one failure to predict the majority choice (shown in bold font), whereas RAM and GDU each have three. The PRT

[5] Upper and lower cumulative independence are implied by any RDU or CPT model; they have been refuted previously (Birnbaum and Navarrete, 1998). Upper cumulative independence assumes $S' = (z', r; x, p; y, q) \prec R' = (z', r; x', p; y', q) \Rightarrow S''' = (x', r; y, q+p) \prec R''' = (x', p+r; y', q)$, where $z' > x' > x > y > y' > z > 0$. According to CPT the percentage who choose R'''in Choice 9 should be greater than or equal to the percentage who choose R' in Choice 10, contrary to the data in Table 5. In studies 1–5, the percentage who choose the risky gamble R' in Choice 10 significantly exceeded 50% and the percentage who choose the risky gamble R''' in Choice 9 was significantly less than 50%. Choices 12 and 14 also violate this property. Lower cumulative independence holds that, $S = (x, p; y, q, z, r) \succ R = (x', p; y', q; z, r) \Rightarrow S'' = (x, p+q; y', r) \succ R'' = (x', p; y', q+r)$. However, results for choices 6, 8, 17, and 20 show that more people switch from choosing the "safe" gamble, S, to the "risky" gamble, R'' than make the opposite switch, despite the improvement in the "safe" gamble. As shown by Birnbaum (1997), lower and upper cumulative independence follow from transitivity, coalescing, consequence monotonicity, and comonotonic restricted branch independence. For more detail on these tests and branch independence, see Birnbaum and Navarrete (1998).

Table 5. Tests of upper and lower cumulative independence, branch independence, and risk aversion (these "filler" trials were common to all five studies). Data are percentages of all 1488 participants who chose the second gamble, shown on the right.

No.	Type	Choice		Data 1488	Predictions of Prior Models TAX	RAM	GDU	PRT
10	S' *vs.* R'	10 to win \$40 10 to win \$44 80 to win \$110	10 to win \$10 10 to win \$98 80 to win \$110	70	56	57	58	65
9	S''' *vs.* R'''	20 to win \$40 80 to win \$98	10 to win \$10 90 to win \$98	31	38	38	41	42
12	R' *vs.* S'	05 to win \$12 05 to win \$96 90 to win \$106	05 to win \$48 05 to win \$52 90 to win \$106	39	49	***56***	***51***	***65***
14	R''' *vs.* S'''	05 to win \$12 95 to win \$96	10 to win \$48 90 to win \$96	73	65	65	63	60
6	S *vs.* R	80 to win \$2 10 to win \$40 10 to win \$44	80 to win \$2 10 to win \$10 10 to win \$98	58	***49***	***49***	***48***	60
8	S'' *vs.* R''	80 to win \$10 20 to win \$44	90 to win \$10 10 to win \$98	69	55	64	64	61
17	R *vs.* S	90 to win \$3 05 to win \$12 05 to win \$96	90 to win \$3 05 to win \$48 05 to win \$52	51	52	55	54	60
20	R'' *vs.* S''	95 to win \$12 05 to win \$96	90 to win \$12 10 to win \$52	35	48	44	43	42
2	RA	50 to win \$0 50 to win \$100	50 to win \$35 50 to win \$45	70	56	66	69	60
16	RA	100 to win \$3	99 to win \$ 0 01 to win \$100	32	49	***50***	***54***	60
19	RA	100 to win \$96	01 to win \$0 99 to win \$100	29	17	27	29	27

Choice types are described in Footnote 5, RA = test of risk attitudes. For $n = 1488$, a 95% confidence interval is $\pm 2.6\%$. All choice percentages significantly differ from 50% except choice 17. Percentages in italics show failures of prior models to predict modal choices, where the discrepancy is greater than 8%.

model fails to predict the violation of branch independence in choices 12 and 17, whereas TAX predicts a violation of branch independence (choices 6 and 10) that did not produce a reversal of modal choices.

9. Model fitting

Table 6 compares the fit of parameterized models. For example, "prior" TAX is labeled TAX(0), the "0" indicating that no parameters were estimated from the new data. Parameters were fit to minimize the sum of squared deviations between predicted and obtained choice

Table 6. A comparison of fit of models to combined data (percentages in Tables 1–5).

Model	Sum of squared errors	Serious errors	Alpha	Beta	Gamma	Delta	RAM WTS (high, med, low)
RAM(5)	2466	2	1.357	0.492	0.488		(1), 0.762, 0.338
TAX(3)	4868	1	0.845	0.533	0.436	(0)	
PRT(3)	6831	2	1.162	0.420	0.616		
RAM(4)	6938	3	0.063	(1)	0.534		(1), 2.17, 1.64
RAM(3)	7145	2	0.063	(1)	(0.6)		(1), 2.32, 1.37
TAX(2)	7544	1	0.059	(1)	(0.7)	0.750	
GDU(4)	7919	6	0.019	1.267	1.831	0.408	
TAX(0)	8031	5	0.049	(1)	(0.7)	(1)	
RAM(2)	8375	4	0.062	(1)	0.453		(1), (2), (3)
PRT(0)	9696	5	0.384	0.631	0.676		
CPT(4)	12000	20	5.787	0.103	0.209	4.222	
GDU(0)	12052	6	0.130	(1)	(1.382)	(0.542)	
PRT(2)	13813	4	0.031	(1)	0.498		
RAM(0)	14408	6	0.136	(1)	(0.6)		(1), (2), (3)
CPT(1)	17592	23	0.039	(0.88)	(0.61)	(0.72)	

Notes: A "serious error" is a case where the model failed to predict the majority choice, and the difference between predicted and obtained choice percentages is greater than or equal to 8%. There were 32 choices between 3-branch gambles and 7 choices between 2-branch gambles, making 39 choice percentages to be fit with up to 5 parameters. Models and parameters are described in Section 2. Values in parentheses are fixed.

percentages in Tables 1 to 5, where discrepancies in Table 5 received four times the weight of those in the other tables (they are based on all 1488 participants). Cases where a model failed to predict the modal choice and the discrepancy was greater than 8% are termed "serious errors." Parameters shown in parentheses were fixed. With three estimated parameters, TAX was the most accurate, followed by PRT, RAM and GDU, with CPT failing badly. RAM(5) provided the lowest sum of squares, but this superiority required a greater number of free parameters than the competitors.

Considering the number of estimated parameters, number of serious errors, and sum of squared discrepancies, TAX(2), with $\delta = .748$, is perhaps the best representation of the data. The smaller value of δ indicates that these data evidence less risk aversion than did previous data used to estimate the prior model. Asymptotic estimates of standard errors of the free parameters, α and δ , are 0.036 and 0.178, respectively. Bootstrapped estimates are slightly smaller. TAX(3) also provided a good fit to the data, with the configural weight parameter fixed, $\delta = 0$, and with free $\beta = 0.533$. Similarly, PRT(3), which is a special case of TAX, also provides a reasonable fit. Asymptotic estimates of standard errors for PRT parameters are 0.186, 0.036, and 0.043 for α, β, and γ; bootstrapped estimates are almost twice as large.

TAX(3) still outperformed PRT(3) even without its configural weight parameter (i.e., even with $\delta = 0$). Recall that PRT is a special case of TAX with $\delta = 0$, so these two models differ only in their specified probability functions. The slight advantage of TAX over PRT with the same number of parameters is apparently due to the use in TAX of a nonlinear function for probability weighting whereas PRT uses a linear function. Neither TAX(3) nor PRT(3) would be able to account for violations of restricted branch independence.

10. Studies 6 and 7: Probability to win or lose

Studies 6 and 7 tested a conjecture by Payne (2005) that the important variable in risky decision making is the probability to win or lose in mixed gambles. In Payne's (2005) study, however, overall probability to win was confounded with the number of branches leading to negative consequences. For example, Payne (2005) asked people to choose between:

$$G1 = (\$85, 0.3;\ \$65, 0.05;\ -\$25, 0.25;\ -\$55, 0.15;\ -\$65, 0.25)$$
$$G2 = (\$85, 0.3;\ \$65, 0.05;\ \$0, 0.25;\ -\$55, 0.15;\ -\$90, 0.25),$$

Payne reported that 59% chose $G2$, which has a probability of 0.60 to yield a non-negative consequence, whereas $G1$ has a probability of only 0.35 to do this. But $G1$ has three probability-consequence branches leading to negative consequences and $G2$ has only two branches leading to negative consequences. These variables, probability to win and number of branches leading to non-negative consequences, were also confounded in Payne's second study.

Therefore, Payne's results may be due to violations of coalescing rather than to probability to win or lose. The following choice pits these two theoretical interpretations against each other:

$$G3 = (\$100, 0.35;\ \$0, .37;\ -\$95, .04;\ -\$97, .04;\ -\$100, 0.20)$$
$$G4 = (\$100, .10;\ \$99, 0.10;\ \$96, 0.10;\ \$0, 0.40;\ -\$100, 0.30)$$

Here, probability to win a positive consequence in $G3$ is 0.35 and the probability to get a negative consequence is 0.28; whereas, in $G4$, there is only a 0.30 probability to win and a higher probability to lose, 0.30. Pitted against probability to win or lose, however, we see that $G3$ has three branches leading to negative consequences against only one in $G4$, and that $G4$ has three branches leading to positive consequences compared to only one in $G3$. But $G3$ dominates $G4$ by first order stochastic dominance. If probability to win or lose is important, people ought to satisfy stochastic dominance in this choice and choose $G3$; however, if splitting is important, they should choose $G4$.

Payne's interpretation, CPT, the model of Lopes and Oden (1999), expected utility, expected value, and stochastic dominance all imply that people should choose $G3$. Prior TAX, however, predicts the opposite. To apply TAX to mixed gambles, it was assumed that utility of money is proportional to money ($u(x) = x$ with $-\$100 \leq x \leq \$100$), and that the same prior TAX parameters apply to mixed gambles as to strictly positive ones. This TAX model predicts that people will choose $G4$, which has more branches leading to positive consequences and fewer leading to negative ones, because $G3$ and $G4$ have cash equivalent values of $-\$55.8$, and $-\$14.7$, respectively, so $G4$ should be preferred to $G3$.

In Study 6, the choice between $G3$ and $G4$ was included as the only test of its type among 20 other choices between mixed gambles, using procedures similar to those of Studies 1–5 with 114 undergraduates. Study 7 (with 200 different undergraduates) used two additional variations of this choice included among 18 other choices. The probability mechanism was described as drawing a ticket randomly from an urn containing 100 tickets. The two key trials were choices #4 and 21, as listed in the following URL, which contains complete instructions and materials: http://psych.fullerton.edu/mbirnbaum/psych466/exps/gls_2-branch.htm

In Study 6, it was found that 63% chose $G4$, consistent with TAX. Study 7 found 66.5% and 65.5% violated dominance by choosing the gamble with the smaller probability to win but the lower number of negative branches, respectively. All three percentages are significantly greater than 50%, showing that people systematically preferred the dominated gamble. In Study 7, 103 people chose the dominated gamble on both choices, compared to only 39 who chose the dominant gamble with the higher probability to win both times ($z = 5.37$).

These results suggest that Payne's (2005) results were likely due to the splitting manipulation that was confounded in his study, rather than to probability to win or lose. They also show that violations of stochastic dominance are not limited to gambles with positive consequences, but persist in mixed gambles, even when probability to win or lose is used to "help" stochastic dominance.

11. Discussion

Each of studies 1–5 provided two replications of the basic test of stochastic dominance that had been proposed as a test of CPT. Consistent with previous results, the percentages of violation in studies 1–5 are all close to 70%, the incidence reported by Birnbaum and Navarrete (1998) in a lab study. Apparently one can find good consistency in Web studies, once procedures for recruiting and testing participants are standardized, and decent sized samples are obtained.

Although CPT was correct in all 7 predictions for binary gambles (where CPT agrees with TAX), CPT cannot account for systematic violations of stochastic dominance. Out of 32 choices between three-branch gambles in Tables 1–5, the CPT model of Tversky and Kahneman (1992) made 23 erroneous predictions of the modal choice, and it was wrong in all three tests with mixed gambles in Studies 6–7. Considering the evidence presented here as well as the growing mass of other evidence contradicting CPT (Birnbaum, 1999b, 2004a, 2004b, 2005; Birnbaum and Navarrete, 1998; Wu, 1994; Wu and Markle, submitted); I think it is time to set CPT aside and move on to evaluate models that can describe the empirical phenomena of risky decision making.

These studies used five manipulations, two that varied how probability was distributed, two that manipulated branch consequences, and one that pitted probability to win against branch splitting. The effects of the manipulations in Studies 1–5 were not consistent with a heuristic of counting the number of branch consequences in one gamble that exceed corresponding consequences on branches of the other gamble. This heuristic erroneously implies no effect of probability in Studies 1–3. It also implies reversals that did not materialize in

Studies 4–5, when the dominated gamble had two lower valued branches and only one higher-valued branch than the dominating gamble. In studies 4 and 5, the effects of changing the values of consequences was not greater than effects predicted by quantitative models. So, no evidence emerged to support the idea that people use such a heuristic.

Heurisitics and biases are sometimes used as post hoc "excuses," which allow a theorist to ignore violations. They provide names for phenomena that give the illusion of understanding, and create the illusion that except for this "bias", the results would have conformed to the theory. For example, a defender of CPT might have said that violations of stochastic dominance do not refute CPT because they are due to a "counting heuristic," and are therefore to be treated as a blemish on the theory (that can be covered by make up), rather than as a fatal wound.

Some heuristic models seem "straw men" because they focus on some simple feature of a choice and ignore other relevant information. In the case of the counting heuristic, the heuristic does not attend to probabilities. Although such heuristics, like the editing rules of prospect theory (Kahneman and Tversky, 1979) or Payne's (2005) probability to win, seem implausible once they are stated as scientific hypotheses and refuted, it is useful, nevertheless to test and refute such simple models; lest they otherwise continue to haunt the field.[6]

Four models that predict violations of stochastic dominance were compared. Prior TAX and PRT outperformed prior RAM and GDU in Studies 1–2, where RAM and GDU predicted lower incidences of violations than were observed. In Study 3, RAM predicted a violation of probability monotonicity that failed to materialize, and here the other three models were more accurate. Although TAX, PRT, and GDU predict the correct direction of the effect in Study 3, the magnitude of the observed effect was larger than predicted by any of these prior models.

This result in Study 3 might indicate a problem with all of these models. Certainly by estimating parameters from the data, the fit of these models can be improved. However, the fact that the basic violation percentages (70% choosing A over B) are so consistent from study to study suggests that the people tested here are not that different from those tested previously. More convincing and informative than model fitting would be the creation of new "paradoxes" from any systematic error in a model, where the new paradoxes expose contradictions in the theory that cannot be avoided by changing functions or parameters.

In RAM, TAX, GDU, or PRT models, it is violations of coalescing that allow these models to predict violations of stochastic dominance. The manipulation of splitting or coalescing of branches has been shown to both create and nearly eliminate violations of stochastic dominance (Birnbaum, 1999b, 2004a, 2004b). Other manipulations intended to eliminate or reduce the incidence of violations, however, have not been very effective, despite a wide search for procedures in which stochastic dominance would be satisfied (Birnbaum, 2004b). For example, in a recent study, I asked participants to make a series of choices and write their reasons for each of six choices. One of the choices was the same as the first example

[6]Another counting heuristic that seems plausible when all consequences are equally likely is the median model. In this model, the participant evaluates a gamble by the consequence that has an equal number of branches with higher or lower consequences. But this model was tested and rejected as an explanation of violations of branch independence (Birnbaum and McIntosh, 1996).

in this paper. The incidence of violations of stochastic dominance was not reduced by this exercise, contrary to the idea that being asked to justify choices might help people notice and satisfy dominance.

PRT was nearly as accurate as TAX for these data. However, PRT cannot account for systematic violations of restricted branch independence, which have been observed in a number of studies (Birnbaum, 2004a, 2004b; Birnbaum and Navarrete, 1998). The idempotent lower GDU model tested here cannot account for violations of upper coalescing or violations of upper tail independence (Wu, 1994). Newer forms of GDU (Ng, Luce, and Marley, 2005) that extend the models of Meginniss (1976), and which violate idempotence, remain to be tested in future studies (see Footnote 3).

As noted by Fishburn (1978), models that violate idempotence can be made to violate a seemingly "transparent" recipe for stochastic dominance. Idempotence is the assumption that $G = (x, p_1; x, p_2; \ldots; x, p_n) \sim x$. Suppose idempotence is violated as follows: $G \succ x$. If so, then we can select a value $x' = x + \varepsilon$ for some small value of $\varepsilon > 0$ such that $G \succ x'$, violating stochastic dominance. In other words, people should prefer a gamble whose outcomes are all equal to x over a sure thing with a strictly higher consequence, x'. Kahneman and Tversky (1979) noted that their original prospect theory would violate dominance in this way, which they thought implausible, so they postulated that people detect and conform to dominance during an editing phase that precedes evaluation of the gambles.

None of the models compared here (TAX, RAM, GDU, or PRT) violate idempotence, so they do not violate transparent dominance in Fishburn's recipe. A goal for future research comparing these models with the class of non-idempotent models will be to devise tests that are perhaps less transparent than that of Fishburn's recipe.

In sum, results show that violations of first-order stochastic dominance are robust and that they cannot be explained by a heuristic that ignores branch probabilities. Given the same number of parameters, the TAX model is more accurate than PRT, RAM, or GDU in predicting violations and satisfactions of stochastic dominance. Unlike RAM, the TAX model does not imply large violations of probability monotonicity in Study 3, though it does not always satisfy this property. Unlike idempotent lower GDU, TAX violates both upper and lower coalescing. Whereas in PRT, weights are independent of rank, TAX has rank-affected weights, so it accounts for violations of restricted branch independence. In fact, TAX attributes risk aversion and violations of branch independence to the same mechanism. This aspect of TAX gives it the advantage of an extra parameter over PRT, which is a special case of both TAX and RAM.

References

Becker, Joao L. and Rakesh K. Sarin. (1987). "Lottery Dependent Utility," *Management Science* 33, 1367–1382.

Birnbaum, Michael H. (1974). "The Nonadditivity of Personality Impressions," *Journal of Experimental Psychology* 102, 543–561.

Birnbaum, Michael H. (1997). "Violations of Monotonicity in Judgment and Decision Making," In Anthony A. J. Marley (ed.), *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce*. Mahwah, NJ: Erlbaum, pp. 73–100.

Birnbaum, Michael H. (1999a). “Paradoxes of Allais, Stochastic Dominance, and Decision Weights,” In James Shanteau, Barbara Mellers, and David Schum (eds.), *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Norwell, MA: Kluwer Academic Publishers, pp. 27–52.

Birnbaum, Michael H. (1999b). “Testing Critical Properties of Decision Making on the Internet,” *Psychological Science* 10, 399–407.

Birnbaum, Michael H. (2004a). “Causes of Allais Common Consequence Paradoxes: An Experimental Dissection,” *Journal of Mathematical Psychology* 48, 87–106.

Birnbaum, Michael 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, Michael H. (2005) “Three New Tests of Independence that Differentiate Models of Risky Decision Making,” *Management Science* 51, 1346–1358.

Birnbaum, Michael H. and Teresa Martin. (2003). “Generalization Across People, Procedures, and Predictions: Violations of Stochastic Dominance and Coalescing,” In Sandra L. Schneider and James Shanteau (eds.), *Emerging Perspectives on Decision Research*. New York: Cambridge University Press, pp. 84–107.

Birnbaum, Michael H. and William R. McIntosh. (1996). “Violations of Branch Independence in Choices Between Gambles,” *Organizational Behavior and Human Decision Processes* 67, 91–110.

Birnbaum, Michael H. and Juan B. Navarrete. (1998). “Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence,” *Journal of Risk and Uncertainty* 17, 49–78.

Birnbaum, Michael H., Jamie N. Patton and Melissa K. Lott. (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, Michael H. and Stephen E. Stegner. (1979). “Source Credibility in Social Judgment: Bias, Expertise, and The Judge’s Point of View,” *Journal of Personality and Social Psychology* 37, 48–74.

Camerer, Colin F. and Robyn M. Hogarth. (1999). “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Theory,” *Journal of Risk and Uncertainty* 19, 7–42.

Diecidue, Enrico and Peter P. Wakker. (2001). “On The Intuition of Rank-Dependent Utility,” *Journal of Risk and Uncertainty* 23, 281–298.

Fishburn, Peter C. (1978). “On Handa’s ‘New Theory of Cardinal Utility’ and the Maximization of Expected Return,” *Journal of Political Economy* 86, 321–324.

Gonzalez, Richard and George Wu. (1999). “On The Shape of The Probability Weighting Function,” *Cognitive Psychology* 38, 129–166.

Humphrey, Steven J. (1995). “Regret Aversion or Event-Splitting Effects? More Evidence Under Risk and Uncertainty,” *Journal of Risk and Uncertainty* 11, 263–274.

Kahneman, Daniel. (2003). “Experiences of Collaborative Research,” *American Psychologist* 58, 723–730.

Kahneman, Daniel and Amos Tversky. (1979). “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica* 47, 263–291.

Levy, Moshe and Haim Levy. (2002). “Prospect Theory: Much Ado About Nothing,” *Management Science* 48, 1334–1349.

Lopes, Lola L. and Gregg C. Oden. (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. Duncan. (1998). “Coalescing, Event Commutativity, and Theories of Utility,” *Journal of Risk and Uncertainty* 16, 87–113.

Luce, R. Duncan. (2000). “Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches,”

Luce, R. Duncan and Peter C. Fishburn. (1991). “Rank- and Sign-Dependent Linear Utility Models for Finite First Order Gambles,” *Journal of Risk and Uncertainty* 4, 29–59.

Luce, R. Duncan and Peter C. Fishburn. (1995). “A Note on deriving Rank-Dependent Utility Using Additive Joint Receipts,” *Journal of Risk and Uncertainty* 11, 5–16.

Luce, R. Duncan and Anthony A. J. Marley. (2005). “Ranked Additive Utility Representations of Gambles: Old and New Axiomatizations,” *Journal of Risk and Uncertainty* 30, 21–62.

Marley, Anthony A. J. and R. Duncan Luce. (2001). “Rank-Weighted Utilities and Qualitative Convolution,” *Journal of Risk and Uncertainty* 23, 135–163.

Marley, Anthony A. J. and R. Duncan Luce. (in press). "Independence Properties Vis-à-vis Several Utility Representations," *Theory and Decision* in press.

Martin, Teresa. (1998). "Comparing Rank Dependent, Subjective Weight, and Configural Weight Utility Models: Transitivity, Monotonicity, Coalescing, Stochastic Dominance, and Event Splitting Independence," Master's Thesis, California State University, Fullerton.

Meginniss, James R. (1976). "A New Class of Symetric Utility Rules for Gambles, Subjective Marginal Probability Functions, and a Generalized Bayes' Rule," *Proceedings of The American Statistical Association, Business and Economic Statistics Section*, 471–476.

Ng, Che T., R. Duncan Luce and Anthony A. J. Marley. (manuscript). "On The Utility of Gambling: Extending The Approach of Meginniss (1976)," *manuscript available from R. Duncan Luce, School of Social Science, University of California, Irvine, CA 92697-5100*.

Quiggin, John. (1985). "Subjective Utility, Anticipated Utility, and The Allais Paradox," *Organizational Behavior and Human Decision Processes* 35, 94–101.

Quiggin, John. (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*. Boston: Kluwer.

Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: Wiley.

Starmer, Chris. (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, Chris and Robert Sugden. (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, Chris and Robert Sugden. (1993). "Testing for Juxtaposition and Event-Splitting Effects," *Journal of Risk and Uncertainty* 6, 235–254.

Tversky, Amos and Daniel Kahneman. (1986). "Rational Choice and the Framing of Decisions," *Journal of Business* 59, S251–S278.

Tversky, Amos and Daniel Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.

Tversky, Amos and Peter Wakker. (1995). "Risk Attitudes and Decision Weights," *Econometrica* 63, 1255–1280.

Viscusi, W. Kip. (1989). "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty* 2, 235–264.

Wakker, Peter. "The Data of Levy and Levy (2002), "Prospect Theory: Much Ado About Nothing?" Support Prospect Theory," *Management Science* 49, 979–981.

Wakker, Peter and Amos Tversky. (1993). "An Axiomatization of Cumulative Prospect Theory," *Journal of Risk and Uncertainty* 7, 147–176.

Weber, Elke U. (1994). "From Subjective Probabilities to Decision Weights: The Effects of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin* 114, 228–242.

Wu, George. (1994). "An Empirical Test of Ordinal Independence," *Journal of Risk and Uncertainty* 9, 39–60.

Wu, George and Richard Gonzalez. (1996). "Curvature of the Probability Weighting Function," *Management Science* 42, 1676–1690.

Wu, George and Alex B. Markle. (2004). "An Empirical Test of Gain-Loss Separability in Prospect Theory," *Working Manuscript, 06-25-04.* Available from George Wu, University of Chicago, Graduate School of Business, 1101 E. 58 th Street, Chicago, IL 60637, http://gsbwww.uchicago.edu/fac/george.wu/research/abstracts.html.