Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Psychology 27 (2006) 737–761
JOURNAL OF Economic Psychology
www.elsevier.com/locate/joep

Evidence against prospect theories in gambles with positive, negative, and mixed consequences

Michael H. Birnbaum *

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

Received 11 April 2006; accepted 13 April 2006
Available online 15 June 2006

Abstract

Three experiments tested behavioral properties of risky decision making. In the first, 433 participants received different formats for display of gambles. Within each of three formats, five "new paradoxes" that violate cumulative prospect theory (CPT) were tested. Despite suggestions that theories might be descriptive with these procedures, all five paradoxes persisted in all three formats. A second experiment with 200 participants tested the same properties in gambles on losses. These data also violated CPT but were approximately compatible with the reflection hypothesis. In the third experiment, consequences were framed to produce "mixed" gambles. These data violated CPT, but there were also significant framing effects. Results contradicted four editing principles and implications of original prospect theory. The major findings agree with a transfer of attention exchange model that has also bested CPT in other studies. Combined with other results, there is now a substantial body of evidence with five paradoxes that refute CPT with positive, negative and mixed gambles, involving more than 8000 participants and 14 formats for displaying gambles.

JEL classification: C91; D81

PsycINFO classification: 2340

Keywords: Behavioral economics; Choice paradoxes; Cumulative prospect theory; Decision making; Expected utility; Rank dependent utility; Risky decision making; Utility theory

* Tel.: +1 714 278 2102/7653; fax: +1 714 278 7134.
E-mail address: mbirnbaum@fullerton.edu

0167-4870/$ - see front matter
doi:10.1016/j.joep.2006.04.001

1. Introduction

Recent studies reported evidence against rank dependent utility (RDU) and cumulative prospect theories (CPT) of risky decision making (Birnbaum, 1999, 2004b; Birnbaum & Navarrete, 1998). They found systematic violations of four behavioral properties implied by any form of RDU or CPT. These four behavioral properties are coalescing, stochastic dominance, lower cumulative independence, and upper cumulative independence. These studies also observed violations of a fifth property, restricted branch independence, which do not contradict RDU or CPT with a nonlinear probability weighting function. However, the observed pattern of violations of restricted branch independence contradicts the inverse-S probability weighting function assumed in CPT and needed to describe the standard Allais paradoxes.

Definitions of these five properties and empirical violations of CPT are displayed in Table 1. Proofs that CPT implies these behavioral properties are given in Appendix A. The "new paradoxes" of Table 1 are consistent with a transfer of attention exchange (TAX) model, a model fit by Birnbaum (1997) to previous data that successfully predicted the first four findings in advance, without estimating any new parameters from the new data (Appendix B).

These findings contradicted what had been a growing consensus in economics in favor of CPT (Starmer, 2000; Wu, Zhang, & Gonzalez, 2004). The argument in favor of CPT was that it could account for much of the data that had been assembled to refute expected utility (EU) and original prospect theories as descriptive models of risky decision making. Indeed, a share of the Nobel Prize in Economics in 2002 was awarded to Daniel Kahneman, in recognition of his empirical work that refuted EU but was consistent with the prospect theories.

Table 1
Properties of choice tested in the experiments, including five "new paradoxes" that violate CPT

Property	Expression	Example violation of CPT
Stochastic dominance[a]	$G^+ = (x,p;y^+,q';y,r-q') \succ G^- = (x,p-q;x^-,q;y,r)$	$G^+ = (\$96,.90;\$14,.05;\$12,.05) \prec G^- = (\$96,.85;\$90,.05;\$12,.10)$
Coalescing	$G^+ \succ G^- \iff GS^+ = (x,p-q;x,q;y^+,q';y,r-q') \succ GS^- = (x,p-q;x^-,q;y,q';y,r-q')$	$G^+ \prec G^-$ and $GS^+ = (\$96,.85;\$96,.05;\$14,.05;\$12,.05) \succ GS^- = (\$96,.85;\$90,.05;\$12,.05;\$12,.05)$
Lower cumulative independence[b]	$S = (x,p;y,q;z,r) \succ R = (x',p;y',q;z,r) \Rightarrow S'' = (x,p+q;y',r) \succ R'' = (x',p;y',q+r)$	$S = (\$44,.1;\$40,.1;\$2,0.8) \succ R = (\$98,.1;\$10,.1;\$2,.8)$ and $S'' = (\$44,.2;\$10,.8) \prec R'' = (\$98,.1;\$10,.9)$
Upper cumulative independence[b]	$S' = (z',r;x,p;y,q) \prec R' = (z',r;x',p;y',q) \Rightarrow S''' = (x',r;y,p+q) \prec R''' = (x',p+r;y',q)$	$S' = (\$110,0.8;\$44,.1;\$40,.1) \prec R' = (\$110,.8;\$98,.1;\$10,.1)$ and $S''' = (\$98,.8;\$40,.2) \succ R''' = (\$98,.9;\$10,.1)$
Restricted branch independence[b]	$S \succ R \iff S' \succ R'$ *CPT & inverse*-S $\Rightarrow S \prec R$ *and* $S' \succ R'$	$S \succ R$ and $S' \prec R'$

[a] $p+r=1; q<p; q'<r; x>x^->y^+>y \geqslant 0$.
[b] $p+q+r=1;\ z'>x'>x>y>x'>z>0$.

Kahneman (2003) recalled his collaboration with Amos Tversky that contributed to his winning this prize. He reviewed their development of the editing principle of combination (Kahneman & Tversky, 1979) as follows:

> "To amuse ourselves, we invented the specter of an ambitious graduate student looking for flaws, and we labored to make that student's task as thankless as possible. ...We were concerned that a straightforward application of our model implied that the prospect denoted ($100, .01; $100, .01) – two mutually exclusive .01 chances to gain $100 – is more valuable than the prospect ($100, .02). The prediction is wrong, of course, because most decision makers will spontaneously transform the former prospect into the latter and treat them as equivalent in subsequent operations of evaluation and choice. To eliminate the problem, we proposed that decision makers, prior to evaluating the prospects, perform an editing operation that collects similar outcomes and adds their probabilities. We went on to propose several other editing operations that provided an explicit and psychologically plausible defense against a variety of superficial counterexamples to the core of the theory. We had succeeded in making life quite difficult for that pedantic graduate student..." (Kahneman, 2003, p. 727).

This editing process, so eloquently described, implies coalescing. Although violated by original prospect theory and requiring the special process described above, coalescing is automatically satisfied by the newer version of prospect theory by Tversky and Kahneman (1992), CPT (Birnbaum & Navarrete, 1998).

As noted by Birnbaum (1997, 1999) and Birnbaum and Navarrete (1998), coalescing is a key property for distinguishing models of risky decision making, and it is violated by his configural weight models. If people violate coalescing, they can be induced to violate lower and upper cumulative independence and first order stochastic dominance, four properties implied by CPT. Birnbaum (1999, 2004b) summarized evidence strongly refuting the intuitions expressed by Kahneman (2003, p. 727) and which are assumed in CPT.

Tversky and Kahneman (1992) gave a "pessimistic assessment" in which they doubted that CPT would generalize to new situations. They speculated that experiments testing risky decision making might be highly dependent on uninteresting features of experiments such as how gambles are described to the participants. But if these "uninteresting" manipulations are the variables that determine human behavior in economic experiments, then they are important variables that behavioral scientists must understand if they hope to predict, control, and explain behavior. This paper will investigate several attempts to rescue CPT by means of procedures intended to help find data compatible with that theory.

Birnbaum (2004b) distinguished three variables of such procedural manipulations: form, format, and framing. *Form* is defined with respect to *branches* of a gamble. A *branch* is a probability (or event)–consequence pair that is distinct in the presentation to the decider. In the above example by Kahneman (2003), the first prospect is a three-branch gamble to win $100 with probability .01, to win $100 with probability .01, or to receive $0 with probability .98. The second is a two-branch gamble with one branch of .02 to win $100 and a second branch with probability .98 to win $0. The first is a "split" *form* of the gamble and the second is the "coalesced" *form*. A prospect can be presented in many possible *split forms* or in a unique *coalesced form*, in which all branches leading to the same consequences are combined. According to either the editing principle of combination or the representation of CPT, this variable of *form* (coalesced or split) should have no effect.

The term *format* is used to denote how gambles and choices between gambles are displayed to the participants. For example, one might display probability by means of pie charts representing wheels with spinners, as relative numbers of marbles of various colors in an urn, as percentages, frequencies, lists of lottery tickets, decimal fractions, or in other ways. Similarly, a choice may be arranged with gambles of a choice presented side by side or one placed above the other. Branches may be arranged vertically or horizontally, and they may be juxtaposed or not. Format and form have often been confounded in previous studies, but they need not be.

Kahneman and Tversky (1979) and Tversky and Kahneman (1992) theorized that form (coalesced or split) should have no effect, but format might have large effects. Perhaps the effect of form depends on format. If so, then there may be a format in which CPT can be retained as a descriptive model.

Consequence framing refers to how a given consequence is described; for example, one might describe the same (objective) situation in terms of gains or losses (Kahneman & Tversky, 1979). Would you prefer a 50–50 gamble to win $100 or $0, or would you rather have $45 for sure? Most choose $45 for sure. Now suppose I gave you $100 contingent on your accepting one of the following losing alternatives: would you prefer a 50–50 gamble to lose $100 or $0 or would you prefer to lose $55 for sure? Most choose the gamble in this framing, even though it is objectively the same as the gamble they rejected in the "win" framing. Consequence framing is studied in Experiment 3.

Event-framing refers to how the "events" that determine consequences are described and presented. Tversky and Kahneman (1986) conjectured that *event-framing* might be an important variable. Event-framing was confounded with coalescing and with other variables in their test of stochastic dominance. Birnbaum (2004a, 2004b, in press) found that event-framing manipulations (marble colors on corresponding branches) had miniscule effects; consequently, this variable is not pursued here.

Birnbaum (2004b) and Birnbaum and Martin (2003) tested a total of 11 formats to see if there was some format in which prospect theories would describe majority choices. In all 11 formats, violations of five implications of CPT in Table 1 were observed. However, the demonstration that 11 formats refute CPT does not imply that all formats would contradict CPT.

Harless (1992) compared a matrix format against a text ("tickets") format. A matrix format had previously been used by Savage (1972, p. 103) as a device to convince himself to satisfy his own "sure thing" axiom and thus avoid Allais paradoxes. Keller (1985) reported that such a format indeed reduced incidence of Allais paradoxes. In the matrix format used by Harless, however, juxtaposition of branches was confounded with event-splitting: when branches were juxtaposed, the larger prize of one choice was also split, and when they were not juxtaposed, branches yielding the same prize were coalesced (Harless, Fig. 1). However, in the text ("tickets") format used by Harless (1992, Fig. 3), branches were always coalesced, whether juxtaposed or not. Because the juxtaposition effect was theorized to be a "regret" effect, Harless concluded that "regret" effects depend on problem representation (i.e., format). Experiment 1 will use both split and coalesced forms in matrix formats like those of Harless (1992), to disentangle these variables.

Starmer and Sugden (1993) tested juxtaposition against event-splitting in a matrix format, concluding that event-splitting effects were prominent. Humphrey (1995, 2001) also reported event-splitting effects. However, Luce (2000, p. 183) considered those tests unconvincing because they did not employ within-subjects designs.

The remainder of this paper is organized as follows. The first experiment will use within-subjects tests of coalescing within each of three formats, as well as the other four properties in Table 1. The second experiment investigates these same behavioral properties in gambles with strictly non-positive consequences, and the third experiment tests them with gambles framed to produce "mixed" consequences. All three experiments find consistent evidence against CPT as a descriptive model of risky decision making. The discussion addresses the implications of these results for descriptive models of risky decision making.

2. Experiment 1: Format and form

Perhaps with the matrix format, such as used by Harless (1992) or Humphrey (2001), CPT might be rescued as a descriptive model, contrary to the conclusions of Birnbaum (2004b). We must, however, distinguish form from format in order to provide a proper test of CPT. It has not yet been determined whether the effects of form might be different in this format. Perhaps in matrix format, evidence against CPT would vanish. Experiment 1 of this study therefore explores three variations of the "tickets" used by Harless (1992) and Humphrey (2001) to see if prospect theories can be saved as descriptive of risky decision making in one or more of these previously unexamined formats.

2.1. Method of Experiment 1

2.1.1. Experiment 1: Three new formats

Participants viewed the materials via the WWW using Web browsers. They chose between gambles by clicking a button beside the gamble in each pair that they would prefer to play, knowing that three people would be selected to play one of their chosen gambles for real cash. In all three new formats, there were always 100 numbered tickets and the number on a randomly drawn ticket would determine the prize.

Participants were randomly assigned to one of three different format conditions by a JavaScript routine. Instructions were the same as in the "tickets" condition of Birnbaum (2004b), modified only to accommodate the new formats. Complete materials can be viewed from the following URLs:

http://psych.fullerton.edu/mbirnbaum/decisions/exp2a_new_tickets.htm
http://psych.fullerton.edu/mbirnbaum/decisions/exp2a_new_ticks_UN.htm
http://psych.fullerton.edu/mbirnbaum/decisions/exp2a_ticks_align_rows.htm

The *new tickets* format is shown in Fig. 1. This format specifies ticket numbers corresponding to each prize, like the format of Harless (1992), rather than the number of tickets to win each prize, as in Birnbaum (2004b). For example, instead of "5 tickets to win $14", the new tickets format states, "Tickets # 91–95 win $14."

The *unaligned matrix* format is illustrated in Figs. 2 and 3. Fig. 2 shows the coalesced form (choice 5) and Fig. 3 shows the split form (choice 11) of the same objective choice. Note that in the unaligned format, each branch has equal horizontal spacing within each choice, independent of the number of tickets associated with that branch. The unaligned matrix format is similar to that used by Savage (1972, p. 103) and Connolly (2004, personal communication), who suggested that CPT might be rescued in this format.

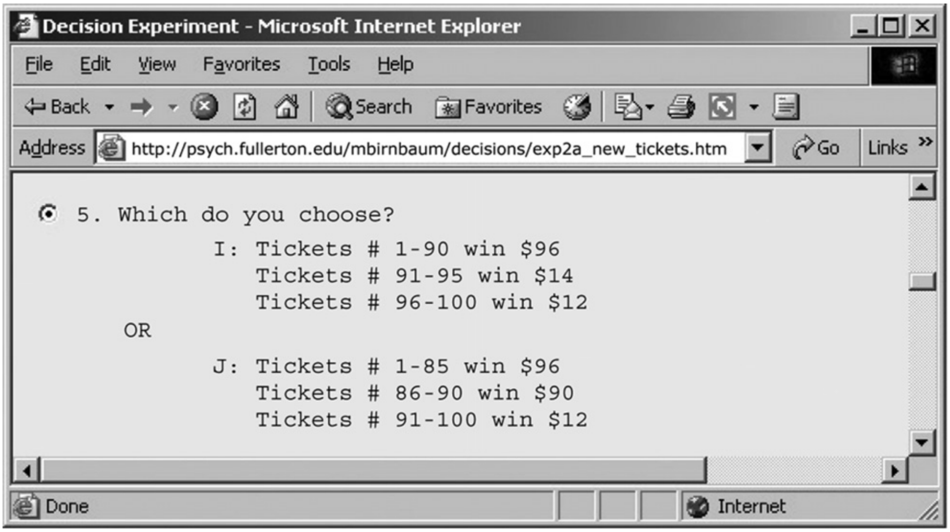

Fig. 1. New tickets format. The "split" form of this choice used additional lines, as in the tickets format of Birnbaum (2004b).

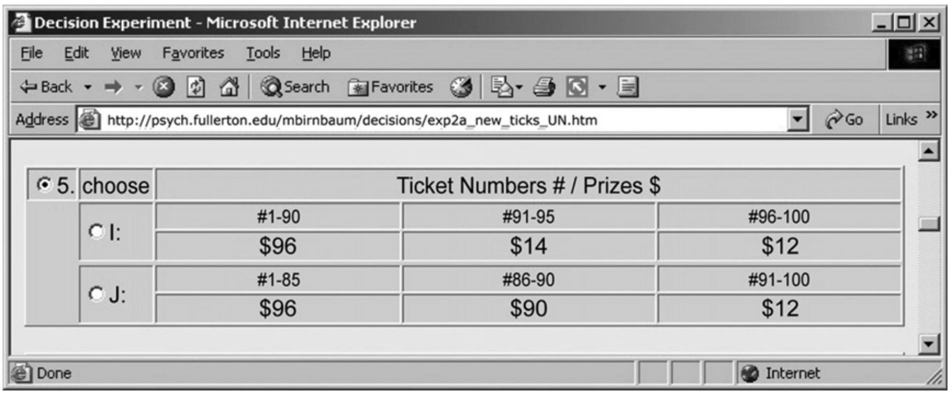

Fig. 2. Unaligned matrix format in coalesced form.

The *aligned* matrix format is illustrated in Figs. 4 and 5. In the aligned format, horizontal spacing was constrained to vary monotonically with the number of tickets in each branch. The HTML was, in fact, programmed to make the horizontal spacing proportional to numbers of tickets; however, proportionality cannot be guaranteed for all browsers, systems, monitors, and window settings. Fig. 4 illustrates how choice 5 is displayed in the coalesced form. Note that in Fig. 4, Tickets #86–90 in Gamble *J* appear to the left of tickets #91–95 in Gamble *I*, unlike the unaligned format shown in Fig. 2. Like Birnbaum's (2004b) *pie chart* formats, the aligned format should (in theory) reveal stochastic dominance visually.

Birnbaum (2004b) noted that the tests of restricted branch independence in his study were not optimally "tuned." In the new tickets format, values of consequences were

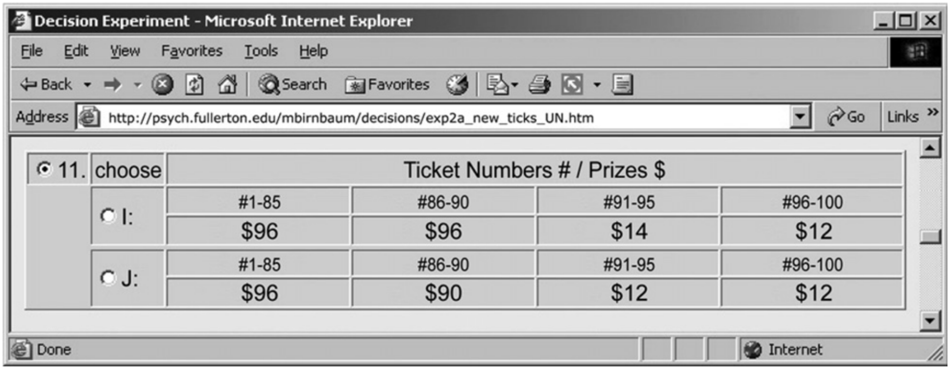

Fig. 3. Unaligned matrix format in split form.

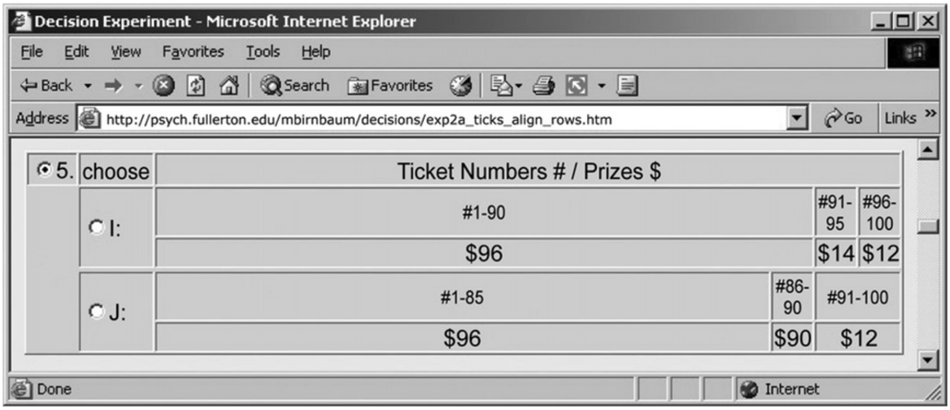

Fig. 4. Aligned matrix format in coalesced form.

adjusted to provide a more optimal test; given those recent data. Specifically, the values $40 and $44 were replaced with $43 and $47, respectively.

Participants were 433 people (85% were recruited from the "subject pool" of California State University, Fullerton, and 15% were recruited via the WWW). About half of the undergraduates were tested in labs, and half participated via the Web at times and places of their own choosing. No systematic differences were observed between these groups, once participant demographics were factored out, so data are combined in the analyses presented here. Participants were randomly assigned to one of three conditions: 141, 141, and 151 completed the new tickets, aligned matrix, and unaligned matrix formats, respectively.

2.2. Results of Experiment 1

Table 2 presents tests of stochastic dominance and coalescing. According to CPT, people should satisfy stochastic dominance by choosing G^+ over G^- in choices 5 and 7, because G^+ dominates G^- by first order stochastic dominance. Instead, Table 2 shows that

84%, 81%, and 72% violated stochastic dominance in the new tickets, unaligned, and aligned formats on choice 5, respectively and 76%, 73%, and 72% on choice 7, respectively. All six percentages are significantly greater than 50%. (With $n = 141$, percentages above 58% or below 42% deviate significantly from 50% by a two-tailed binomial test with $\alpha = .05$ significance level.)

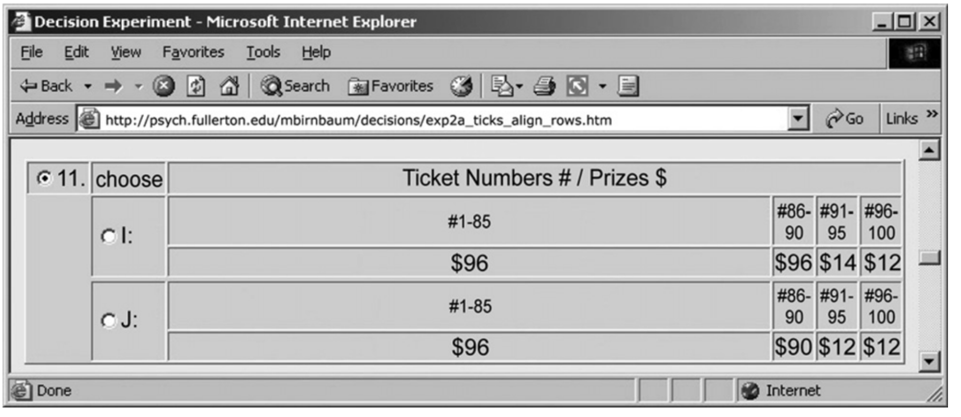

Fig. 5. Aligned format in split form.

Table 2
Violations of stochastic dominance and coalescing

No.	Choice type		Condition				
	G^+	G^-	Tickets	New tickets	Unaligned	Aligned	Negative (reflected)
5	#1–90 win $96 #91–95 win $14 #96–100 win $12	#1–85 win $96 #86–90 win $90 #91–100 win $12	71	84	81	72	72
11	#1–85 win $96 #86–90 win $96 #91–95 win $14 #96–100 win $12	#1–85 win $96 #86–90 win $90 #91–95 win $12 #96–100 win $12	06	08	12	09	15
7	#1–94 win $99 #95–97 win $8 #98–100 win $6	#1–91 win $99 #92–94 win $96 #95–100 win $6	67	76	73	72	71
13	#1–91 win $99 #92–94 win $99 #95–97 win $8 #98–100 win $6	#1–91 win $99 #92–94 win $96 #95–97 win $6 #98–100 win $6	13	06	12	12	16

Table entries are percentages of violation of stochastic dominance in Experiments 1 and 2.
Notes: The dominant gamble (G^+) was presented first in choices 5 and 11 and second in choices 7 and 13. All choice percentages in the table are significantly different from 50%. Note that choices 5 and 11 are the same, except for coalescing, as are choices 7 and 13. The tickets format is from Birnbaum (2004b), with 342 participants. In the negative condition, where wins are replaced by losses, gambles G^+ and G^- switch roles.

Note that choice 5 in Table 2 is objectively the same as choice 11 in Table 2, except for coalescing. According to original prospect theory with the editing operation of combination or by CPT, people should make the same decision in choice 5 as in choice 11, except for error. Similarly, they should make the same decision in choice 7, which is the coalesced form of choice 13.

Table 2 shows, however, that when the same choices are presented in appropriately split forms (in which corresponding branches have the same ticket numbers), the vast majority satisfies stochastic dominance in each format. In choices 11 and 13, there are about 10% violations, averaged over the six tests; all of these percentages are significantly less than 50%. Data from Birnbaum's (2004b) tickets condition ($n = 342$) are presented in Table 2 for comparison.

The new data for choices 5 and 7 (coalesced form) show slightly higher rates of violation of stochastic dominance in the new formats, perhaps because of the greater percentage of undergraduates in this sample compared to Birnbaum (2004b). (The column labeled "Negative" shows results from Experiment 2, which are described later).

Table 2 answers the key question. It shows that even in matrix formats, there are significant majorities who violate stochastic dominance when choices are presented in coalesced form (Figs. 2 and 4). Thus, the key to this result is not the format but instead the form; that is, majority violations of stochastic dominance are observed in all new formats with coalesced gambles and are nearly eliminated by using the appropriately split form of the choices (Figs. 3 and 5).

Table 3 presents a more detailed breakdown of stochastic dominance and coalescing, showing the number of people who showed each response pattern on choices 5 and 11. As shown in Table 3, there were 109, 110, and 91 participants who violated stochastic dominance in the coalesced form *and* satisfied it in the split form versus 1, 7, and 4 participants who had the opposite reversal of preferences in new tickets, unaligned, and aligned matrix formats, respectively. The new data show 72% of participants with the pattern predicted by TAX, averaged over the three new conditions, close to the 65% found in the previous tickets data of Birnbaum (2004b). Significantly more than half of participants violated coalescing by reversing preferences, which violates CPT. Similar results were obtained for choices 7 and 13 (not shown). (Row totals in Table 3 do not always add to the exact number of participants because a few people skipped one or more items.)

Combining choices 5 and 7 over conditions, 63% showed two violations of stochastic dominance.

Table 3
Analysis of coalescing (event-splitting effects) in choices 5 and 11

Condition	Choice pattern			
	G^+GS^+	G^+GS^-	G^-GS^+	G^-GS^-
Tickets (342)	95	3	224	16
New tickets (141)	21	1	109	10
Unaligned (151)	22	7	110	11
Aligned (141)	34	4	91	9
Total new (433)	77	12	310	30
Negative (reflected)	88	23	250	34

Each entry shows the number of choices of each pattern in the different studies. (Row totals may not equal the number of participants, due to occasional skipping of an item.)

Table 4
Tests of upper cumulative independence and lower cumulative independence

No.	Type	First gamble	Second gamble	Condition				
				Tickets 342	New tickets	Unaligned	Aligned	Neg (1 − *p*)
10	S' vs R'	#1–80 win $110 #81–90 win $44 #91–100 win $40	#1–80 win $110 #81–90 win $98 #91–100 win $10	74	65	71	70	70
9	S''' vs R'''	#1–80 win $98 #81–100 win $40	#1–90 win $98 #91–100 win $10	33	23	23	21	41
12	R' vs S'	#1–90 win $106 #91–95 win $96 #96–100 win $12	#1–90 win $106 #91–95 win $52 #96–100 win $48	46	45	49	51	37
14	R''' vs S'''	#1–95 win $96 #96–100 win $12	#1–90 win $96 #91–100 win $48	78	86	79	75	68
6	S vs R	#1–10 win $44 #11–20 win $40 #21–100 win $2	#1–10 win $98 #11–20 win $10 #21–100 win $2	63	37	46	43	55
8	S'' vs R''	#1–20 win $44 #21–100 win $10	#1–10 win $98 #91–100 win $10	70	77	72	69	66
17	R vs S	#1–5 win $96 #6–10 win $12 #11–100 win $3	#1–5 win $52 #6–10 win $48 #11–100 win $3	54	71	60	48	54
20	R'' vs S''	#1–5 win $96 #6–100 win $12	#1–10 win $52 #11–100 win $12	31	25	29	32	24

Note: Each entry is the percentage of participants in each condition who chose the gamble shown on the right. In the new tickets format, the values $40 and $44 were changed to $43 and $47, respectively. In the negative condition, losses replaced wins, and the percentages shown are percentages choosing the gamble shown on the left.

Table 4 summarizes tests of upper cumulative independence: $S' \prec R' \Rightarrow S''' \prec R'''$. We see in choice 10 that a significant majority prefers R' in all conditions (65%, 71%, and 70% in new tickets, unaligned, and aligned, respectively); however significantly less than half prefers R''' in choice 9 (23%, 23%, and 21%, respectively), violating upper cumulative independence. Choices 12 and 14 also violate upper cumulative independence (with position of S and R reversed and different probabilities). The same pattern of violation is found because the percentage choosing the "safe" gamble is significantly greater in choice 14 than in choice 12 in all three conditions.

Choices 6 and 8 of Table 4 test lower cumulative independence, $S \succ R \Rightarrow S'' \succ R''$. In each of the three new conditions, the majority preferred S in choice 6 (37%, 46%, and 43% chose R); however, significant majorities prefer R'' in choice 8 (77%, 72%, and 69%, respectively). Choices 17 and 20 also violate lower cumulative independence, with position counterbalanced.

Choices 6 and 10 form a test of restricted branch independence, $S \succ R \Longleftrightarrow S' \succ R'$. Note that the common branch is either 80 to win \$2 (choice 6) or 80 to win \$110 (choice 10). Whereas the majority chose the safe gamble in choice 6 in all three conditions (i.e., 37%, 46%, and 43% chose *R*), significant majorities chose the risky gamble in choice 10 (65%, 71%, and 70%). Choices 12 and 17 also test restricted branch independence, yielding a similar pattern in two of the three new conditions.

The choice proportions in the new formats are quite close to previous "tickets" data of Birnbaum (2004b), except in choice 6, where the percentage choosing *R* was less than 50% in all three new conditions, whereas this percentage exceeded 50% in all 12 conditions in Birnbaum (2004b). Two factors may be responsible for this relative shift: first, in the new tickets condition, the *S* gambles in choices 6, 8, 9, and 10 were altered by changing \$40 and \$44 to \$43 and \$47, respectively. Second, the three new conditions used a greater percentage of undergraduates, who tend to be more risk averse than are college graduates (Birnbaum, 1999).

Tables 5 and 6 analyze upper cumulative independence in choices 10 and 9 and lower cumulative independence in choices 6 and 8, respectively. The number of people who have preference reversals that violate the property is significantly greater than the number of people who switch preferences in the direction consistent with the property.

Table 7 analyzes restricted branch independence in choices 6 and 10. There are significantly more people who switch from *S* to R' than make the opposite switch. This result

Table 5
Tests of upper cumulative independence

Condition	*n*	Choice pattern			
		$S'S'''$	$S'R'''$	$R'S'''$	$R'R'''$
Tickets	340	71	16	156	96
New tickets	141	38	11	71	21
Aligned	141	36	5	74	23
Unaligned	151	34	9	81	25
Total (new)	433	108	25	226	69
Negative (reflected)	200 × 2	77	42	157	123

Choices 10 and 9.
$S' = (\$110, .8; \$44, .1; \$40, .1)$, $R' = (\$110; .8; \$98, .1; \$10, .1)$, $S''' = (\$98, .8; \$40, .2)$, $R''' = (\$98, .9; \$10, .1)$.

Table 6
Tests of lower cumulative independence

Condition	Number	Choice pattern			
		SS''	SR''	RS''	RR''
Tickets	342	52	74	51	163
New tickets	141	20	67	11	39
Aligned	141	32	48	12	49
Unaligned	151	29	52	13	55
Total (new)	433	81	167	36	143
Negative (reflected)	200 × 2	82	96	54	165

Choices 6 and 8.
$S = (\$44, .1; \$40, .1; \$2, .8)$, $R = (\$98, .1; \$10, .1; \$2, .8)$, $S'' = (\$44, .2; \$10, .8)$, $R'' = (\$98; .1; \$10, .9)$.

Table 7
Tests of branch independence

Condition	n	Choice pattern			
		SS'	SR'	RS'	RR'
Tickets Birnbaum (2004b)	342	50	77	37	177
New tickets	141	34	54	14	37
Aligned matrix	141	28	51	13	46
Unaligned matrix	151	28	53	14	52
Total (new data)	433	90	158	41	135
Negative (reflected)	200 × 2	74	104	45	174

Choices 6 and 10.
$S = (\$44, .1; \$40, .1; \$2, .8)$, $R = (\$98, .1; \$10, .1; \$2, .8)$, $S' = (\$110, .8; \$44, .1; \$40, .1))$, $R' = (\$110; .8; \$98, .1; \$10, .1)$.

agrees with previous findings but is opposite the prediction of the inverse-S decumulative weighting function used in CPT, which predicts that people should have shown the opposite reversal.

3. Experiment 2: Reflection

Experiment 2 used the same choices as the first, except that all prizes were converted from gains to hypothetical losses. As noted in Starmer's (2000) review, it is often (but not always) found that gambles composed of chances to lose (or, at best, to break even) have the opposite preference order from that of gambles composed of chances to win or break even, when losses are simply converted to wins. This empirical generalization is known as the *reflection hypothesis* (Kahneman & Tversky, 1979), and it has been reported to be a good approximation for choices that satisfy prospect theories, such as choices between binary gambles (Tversky & Kahneman, 1992).

In this paper, *weak reflection* refers to the generalization that $P(A, B) \geqslant 1/2 \Longleftrightarrow P(-A, -B) \leqslant 1/2$, where A and B are gambles consisting strictly of gains, $P(A, B)$ is the probability of choosing A over B, and $-A$ and $-B$ are the same gambles, except each positive consequence is replaced by a loss of the same absolute value. *Strong reflection* refers to the property that $P(A, B) = 1 - P(-A, -B)$.

The reflection hypothesis cannot hold in general for mixed gambles; however, because the assumption leads to self-contradiction for certain symmetric mixed gambles, unless all such gambles are indifferent to each other (Appendix C).

The reflection hypothesis has not been previously tested with the types of choices used in this study. There are three simple possibilities for the results. First, it is possible that reflection will be satisfied, in which case choices between gambles on losses will also systematically violate CPT. Second, gambles on losses might satisfy CPT, in which case they must violate the reflection hypothesis. Third, both CPT and the reflection hypothesis might fail to hold for these choices.

3.1. Method of Experiment 2

In Experiment 2, the 20 choices were the same as those in Experiment 1, except that all prizes were changed from gains to hypothetical losses. Each gamble was displayed in the reversed text format used by Birnbaum (2004b), except a branch would be written as ".50 probability to lose \$100" rather than ".50 probability to win \$100." In this format, branches are listed in descending order of absolute value of the consequences. There were 200 undergraduates who made the 20 decisions twice, separated by three intervening tasks that required about 15 min. Complete materials for the study, including instructions, can be viewed from the following URL: http://psych.fullerton.edu/mbirnbaum/decisions/exp2a_neg.htm.

3.2. Results of Experiment 2

The choice percentages in Experiment 2, with non-positive consequences have been reflected in Tables 2–7 for ease of comparison. For example, the 72% in the first row of Table 2 indicates that, consistent with reflection but in violation of CPT, 72% of the 400 choices (200 participants by two replicates) chose *I* over *J* in the following choice, even though *J* dominates *I*:

I: .90 probability to lose \$96,
.05 probability to lose \$14,
.05 probability to lose \$12.

J: .85 probability to lose \$96,
.05 probability to lose \$90,
.10 probability to lose \$12.

Comparing choice percentages between experiments, the correlation with Birnbaum's (2004b) tickets condition is −0.96. The largest discrepancy between $1 - P(-A, -B)$ and $P(A, B)$ was 0.133 on choice 15, a test of risk aversion. In that choice, 53.5% chose a .01 chance to win \$100 over a sure win of \$1, but in the negative condition, 59.8% chose a .01 chance to lose \$100 over a sure loss of \$1. According to the strong reflection hypothesis, the second figure should have been 46.5%, assuming no error in the first. According to weak reflection, it should have been less than 50%. Aside from this discrepancy, strong reflection provided a good approximation. Gambles on losses show the same patterns of violations of CPT as found with gains (Tables 2–7).

The negative condition included replication and thus allows a test of reliability. It was found that the average percentage agreement between decisions made on the first and second replicates was 68.6%. By splitting this measure into agreement on the first 10 decisions and the second 10 decisions, it was found that self-agreement percentages correlated 0.40 between these two split-half measures, indicating significant individual differences in internal consistency.

When data are sub-divided into those 85 who agreed with their own choices on 15 or more decisions (75% or more) and those 115 who agreed on fewer than 15 of 20 decisions, it was found that the two groups had similar choice percentages ($r = 0.957$). For example, 79.4% of the more consistent group violated stochastic dominance on choice 7 compared to 64.8% violations in the less consistent group. Of 85 consistent participants, 50 (59%) violated stochastic dominance on all four tests, 15 had three violations, 7 had two, 5 had one, and 7 had no violations. Of the 115 less consistent participants, 33 (29%) had four violations, 31 had three, 29 had two, 15 had one, and 5 had none.

Applying the true and error model (Birnbaum, 2004b) to these data, the more reliable group had an estimated "true" rate of violation of stochastic dominance of 0.854 with an "error" rate of 0.130. The less consistent group had an estimated "true" rate of violation of 0.825 and an "error" rate of 0.284. This result does not support the notion that those who are more careful have lower rates of "true" violations of stochastic dominance. Instead, the model indicates that the two groups had similar estimated rates of "true" violation but estimated error rates differing by a factor more than 2. These findings with negative consequences are similar to results previously obtained with positive consequences (Birnbaum, 1999, 2004b, 2005a, in press).

4. Experiment 3: Consequence framing and cash segregation

Experiment 3 tests the paradoxes of Table 1 in gambles framed to appear as positive, negative, and mixed gambles. This framing was implemented by subtracting \$25, \$50, or \$100 from all consequences of Experiment 1, with the provision that winning participants would receive an endowment of \$25, \$50, or \$100 plus the result of one of their chosen gambles, respectively. If people added the endowment of c to each consequence, they would make the same decisions in all conditions of Experiment 3 as they did in Experiment 1. However, if people ignore the endowment and treat the framed gambles as "mixed" gambles, they might make different decisions.

Cash segregation. The property of *cash segregation* holds that $F \succ G$ if and only if $c \oplus F' \succ c \oplus G'$, where F' and G' are the same as F and G, except that the same objective cash value, c, has been subtracted from each consequence, and $\oplus$ represents the joint receipt of cash and a gamble or two gambles. The same value of cash, c, is given to the decider as a cash endowment. Segregation was postulated by Kahneman and Tversky (1979) as an editing rule, and a more general form has received a theoretical study in theories of joint receipt by Luce (2000, pp. 146–148). Kahneman and Tversky (1979) reported consequence-framing effects, which are violations of segregation.

As in the case of reflection, we can distinguish *weak* and *strong* versions of *cash segregation*. *Weak cash segregation* is defined as $P(F, G) > 1/2 \Longleftrightarrow P(c \oplus F', c \oplus G') > 1/2$, where $P(F, G)$ is the probability of choosing F over G; F' is the same as F, except that cash value, c, has been subtracted from each of its consequences, and the notation $c \oplus F'$ represents the joint receipt of endowment of c plus the result of gamble F'. Define *strong cash segregation* as the assumption that $P(F,G) = P(c \oplus F', c \oplus G')$.

4.1. Method of Experiment 3

In Experiment 3, instructions stated that each participant had a 3% chance to receive the cash consequence of one of their choices plus an endowment, c. There were three conditions, $c = \$25, \$50, \$100$, in which each prize of Experiment 1 was translated by subtracting the corresponding amount, and winners would receive that same amount as an endowment plus the (translated) consequence of one of their chosen gambles. By construction, each choice in Experiment 3 is objectively equivalent (in its take home prize) to its corresponding choice in Experiment 1, so participants could not lose their own money. In the \$100 condition, however, consequences exceeding \$100 were reduced to \$100 before translation (i.e., \$110 and \$106 were replaced by \$100 prior to translation). This meant that no consequences in this condition were framed to exceed zero. Choices 15 and 18 were altered in each condition to match choices 5 and 7, providing two additional tests of stochastic dominance.

Probabilities were formatted in terms of the number of marbles of each color in an urn containing 100 marbles, where the color of a randomly drawn marble would determine the prize.

The order of conditions was counterbalanced during the course of the study. Participants were undergraduates who served as one option for a course assignment in lower division psychology. Some participants served in only one or two of the three conditions. There were 170, 144, and 144 participants who completed both repetitions in the \$25, \$50, and \$100 conditions, respectively. Of these, 75% were female, and 93% were 22 years of age or younger. Included were 92 who completed all six tasks – two repetitions of each of the three conditions.

4.2. Results of Experiment 3

Table 8 displays tests of stochastic dominance in coalesced and split forms, as in Table 2. The average rates of violation exceeded 70% in the \$25, \$50, and \$100 conditions, respectively (averaged over choices 5, 15, 7, and 18). In split form, average violation rates

Table 8
Tests of stochastic dominance and coalescing in mixed framed gambles

No.	Choice		Condition		
	G^+	G^-	\$25	\$50	\$100
5	90 (\$71, \$46, −\$4) 05 (−\$11, −\$36, −\$86) 05 (−\$13, −\$38, −\$88)	85 (\$71, \$46, −\$4) 05 (\$65, \$40, −\$10) 10 (−\$13, −\$38, −\$88)	76	75	73
15	Same as No. 5	Same as No. 5	79	71	77
7	94 (\$74, \$49, −\$1) 03 (−\$17, −\$42, −\$92) 03 (−\$19, −\$44, −\$94)	91 (\$74, \$49, −\$1) 03 (\$71, \$46, −\$4) 06 (−\$19, −\$44, −\$94)	79	73	74
18	Same as No. 7	Same as No. 7	79	69	78
	GS^+	GS^-			
11	85 (\$71, \$46, −\$4) 05 (\$71, \$46, −\$4) 05 (−\$11, −\$36, −\$86 05 (−\$13, −\$38, −\$88)	85 (\$71, \$46, −\$4) 05 (\$65, \$40, −\$10) 05 (−\$13, −\$38, −\$88) 05 (−\$13, −\$38, −\$88)	11	12	09
13	91 (\$74, \$49, −\$1) 03 (\$74, \$49, −\$1) 03 (−\$17, −\$42, −\$92) 03 (−\$19, −\$44, −\$94)	91 (\$74, \$49, −\$1) 03 (\$74, \$49, −\$1) 03 (−\$19, −\$44, −\$94) 03 (−\$19, −\$44, −\$94)	17	15	10

Table entries are percentages of violation of stochastic dominance.
Notes: The dominant gamble (G^+) was presented first in choices 5, 11, and 15 and second in choices 7, 13, and 18. All choice percentages in the table are significantly different from 50%. Values in parentheses show the framed values of the consequences in the \$25, \$50, and \$100 conditions, respectively.

were below 15%, in all three conditions (choices 11 and 13). These results are similar to those in Table 2.

For the 92 participants who completed all three conditions with two repetitions, there are 24 tests of stochastic dominance. Of these, there were 43 people with 20–24 violations (12 had 24 violations), 34 with 15–19 violations, 9 with 10–14 violations, 6 with 5–9 violations, and no one had fewer than five violations. Testing each person separately by a one-tailed binomial test with $\alpha = .05$, one can reject the hypothesis of 50% or fewer violations for 57 of the 92 participants (62%), who had 18 or more violations. Only 11 people had 12 or fewer violations out of 24 tests. Dividing the data between first and second replicates, there is a correlation of 0.78 between the two split-half counts of violation of stochastic dominance, indicating systematic individual differences in observed rates of violation.

Table 9 summarizes tests of upper and lower cumulative independence, as in Table 4. Note that the choice percentages for two-branch gambles (choices 9, 14, 8, and 20) are close to their corresponding values in Table 4 in all three framing conditions. In addition,

Table 9
Tests of upper and lower cumulative independence in Experiment 3

No.	Type	Choices (number of marbles and consequences)								Choice percentage (second gamble)		
		First gamble				Second gamble						
		#	\$25	\$50	\$100	#	\$25	\$50	\$100	\$25	\$50	\$100
10	S' vs R'	80	85	60	0	80	85	60	0	42	74	77
		10	19	−6	−56	10	73	48	−2			
		10	15	−10	−60	10	−15	−40	−90			
9	S''' vs R'''	80	73	48	−2	90	73	48	−2	22	30	28
		20	15	−10	−60	10	−15	−40	−90			
12	R' vs S'	90	81	56	0	90	81	56	0	56	39	24
		05	71	46	−4	05	27	2	−48			
		05	−13	−38	−88	05	23	−2	−52			
14	R''' vs S'''	95	71	46	−4	90	71	46	−4	74	83	77
		05	−13	−38	−88	10	23	−2	−52			
6	S vs R	10	19	−6	−56	10	73	48	−2	35	76	50
		10	15	−10	−60	10	−15	−40	−90			
		80	−23	−48	−98	80	−23	−48	−98			
8	S'' vs R''	20	19	−6	−56	10	73	48	−2	71	79	80
		80	−15	−40	−90	90	−15	−40	−90			
17	R vs S	05	71	46	−4	05	27	2	−48	72	37	42
		05	−13	−38	−88	05	23	−2	−52			
		90	−22	−47	−97	90	−22	−47	−97			
20	R'' vs S''	05	71	46	−4	10	27	2	−48	31	18	15
		95	−13	−38	−88	90	−13	−38	−88			

Note: Each entry is the percentage of participants in each condition who chose the second gamble. Cases of violation of weak cash segregation are shown in bold; these are also cases where the number of branches leading to positive or negative consequences differs within choices.

the three-branch gambles in the $100 condition have similar choice percentages in Table 9 to their corresponding values in Table 4.

However, there are violations of both weak and strong cash segregation ("framing effects") in all four cases of three-branch gambles in Table 9. For choice 10 in the $50 and $100 conditions, 74% and 77% chose the second gamble; however, in the $25 condition, the majority (58%) chose the first gamble, which had no branches leading to negative consequences rather than the second gamble, which had one branch leading to a loss (−$15). In the $50 condition, however, the second gamble in choice 10 has only one negative branch (−$40) but the first gamble has two (−$6 and −$10). Of the 92 who completed both replications of all three conditions, there were 41 and 38 participants who switched from choosing the first gamble in the $25 condition to choosing the second one in the $50 condition in the first and second replicates, respectively. Only 7 and 6 people made the opposite reversals in the two respective replicates ($z = 4.91$ and 4.82). The most frequent response pattern (by 25 of 92 participants) was to choose the first gamble on both replicates in the $25 condition and choose the second gamble on both replicates of the $50 condition; only 2 had the opposite preference pattern ($z = 4.43$).

In choice 6, a similar reversal is observed between the $25 and $50 conditions. In this case, 65% chose the first gamble which has two positive and one negative branches in the $25 condition, but 76% chose the second gamble in the $50 condition, which has one positive branch compared to all negative branches in the first gamble. The percentage choosing the second gamble changed from 35% to 76% (both figures significantly different from 50%). There were 47 and 52 participants who showed this reversal in first and second replicates, whereas only 11 and 4 made the opposite reversals, respectively ($z = 4.73$, 6.41). The modal pattern (by 32 people) showed this reversal on both replicates, whereas only 2 had the opposite pattern ($z = 5.14$). In the $100 condition, all branches in both gambles are negative, and the choice proportion regresses to 50%.

In choice 17, a reversal also occurs between the $25 and $50 conditions. In the $25 condition, 72% choose the second gamble with one negative branch over the second gamble, which has two negative branches. In the $50, condition, however, both gambles have two negative branches and only 37% chose the second gamble. There were 42 and 53 who showed this reversal in first and second replicates including 31 who showed the same reversal in both replicates, compared to 19, 7, and 2 who showed the opposite patterns ($z = 2.94$, 5.94, 5.05).

Choice 12 shows a similar reversal between the $25 and $100 conditions. In the $25 condition, 56% chose the second gamble with all positive branches over the first gamble with one branch leading to a negative consequence. In the $100 condition only 24% chose the second gamble when both gambles have two negative branches. There were 39, 40, and 25 who showed this reversal in first and second replicates and both, compared to 8, 6, and 2 who showed the opposite patterns ($z = 4.52$, 5.01, 4.43). In sum, violations of cash segregation (framing effects) are found in cases when the number of branches leading to positive or negative consequences differs in the two gambles.

Because of these shifts due to differing numbers of positive and negative branches, the violations of restricted branch independence in Table 9 (choices 6 versus 10 and choices 12 versus 17) appear less impressive than their corresponding tests in Table 4, especially in the $50 condition.

Apart from these framing effects, however, note that all twelve tests of upper and lower cumulative independence in Table 9 show the same relations as in Table 4. The percentages choosing the risky gamble in choice 10 exceed those in choice 9 in all three conditions, contrary to upper cumulative independence. The same relation holds for choices 12 and 14 in all three conditions. Contrary to lower cumulative independence, the percentage choosing the risky gamble is greater in choice 8 than in choice 6, and greater in choice 20 than in choice 17 in all three conditions.

5. Discussion

The data of all three new conditions of Experiment 1, the non-positive gambles of Experiment 2, and gambles framed as "mixed" in Experiment 3 show majority violations of stochastic dominance and systematic violations of coalescing. These experiments provide convincing within-subjects results that should settle any lingering doubts about previous tests of these properties. Although these studies were designed to "help" CPT, the evidence against that theory persists with these new format and framing manipulations.

According to CPT, people should satisfy coalescing. According to the editing principle of combination (Kahneman, 2003; Kahneman & Tversky, 1979), people should satisfy coalescing. However, the evidence in all experiments is quite clear: people do not satisfy coalescing. The same choice can produce a choice percentage above 70% in coalesced form and less than 15% in split form (Tables 2 and 8).

According to CPT, people should satisfy first order stochastic dominance. According to the editing principle of dominance detection (Kahneman & Tversky, 1979), people should satisfy dominance when they detect it. Again, the data of all experiments is quite clear: observed violations occur in 70% or more of a typical test of stochastic dominance in three-branch gambles. Even in aligned matrix format, where stochastic dominance should be apparent in the display, the majority of participants violate stochastic dominance. From the true and error model, it is estimated that more than 80% of undergraduates "truly" violate stochastic dominance, and a minority of about 20% satisfy it in these tests. Apparently, people don't detect dominance in these three-branch gambles.

Significant violations of lower cumulative independence and upper cumulative independence were found in all three experiments, which also refute CPT with any utility function and any weighting function.

In addition, in all four conditions of Experiments 1 and 2 and in the $100 condition of Experiment 3, the pattern of violation of branch independence is opposite that predicted by cumulative prospect theory with its inverse-S probability weighting function. Patterns of violation of restricted branch independence, however, were significantly altered by consequence framing in "mixed" gambles of Experiment 3, apparently due to the contrasting numbers of positive and negative branches within those choices. Similar results have been obtained with mixed gambles by Birnbaum (in press): people have a tendency to choose gambles with the greater number of branches leading to positive consequences and smaller number of branches leading to negative consequences.

Systematic violations of restricted branch independence also refute original prospect theory and the editing principle of cancellation (Kahneman & Tversky, 1979).

Before these experiments were conducted, it seemed reasonable to imagine (especially in the aligned matrix format) that the association of tickets to prizes should allow deciders to "see" through tests of stochastic dominance. Examining Figs. 4 and 5, it seemed quite plausible that people might satisfy dominance in this format, and that the subtle difference between Figs. 4 and 5 might have little effect. Similarly, it seemed plausible that people would find it easy to detect common branches in this format and cancel them prior to comparing two gambles. However, the results showed that people behaved in much the same way with these formats as they had done in previous tests with different formats. The approximate stability of conclusions across display formats may be contrary to such prior intuitions, but they should be comforting to theorists. Apparently, one does not need to create new economic theory for each new way to display a choice between gambles.

Experiment 2 with gambles on negative consequences yielded two important new results: First, the strong reflection hypothesis was a good approximation to the data. Second, all five new paradoxes of Table 1 were replicated in gambles with strictly non-positive consequences.

Experiment 3 also found two new results: First, framed "mixed" and "negative" gambles continue to show systematic violations of stochastic dominance, coalescing, lower cumulative independence, and upper cumulative independence. Second, cash segregation was violated in cases where the numbers of branches leading to positive or negative consequences differed between two gambles in a choice. Put another way, there were framing effects that violate both strong and weak cash segregation in three branch gambles. As is the case for the editing rules of cancellation, combination, and dominance detection, the editing rule of cash segregation does not agree with empirical data.

As Kahneman (2003) described their development, the editing rules were postulated to cancel implications of prospect theory that were imagined at the time to be intuitively implausible. They were apparently not tested empirically, but simply stated as preemptive excuses for possible deviations. Although these four principles might work in some people some of the time, they do not appear to be in agreement with data when they are treated as scientific hypotheses and put to the test.

Combined with previous data, 14 different formats have now been tested with more than eight thousand participants. These studies consistently yield data in violation of prospect theories and their editing principles for gambles on positive consequences. The new data for gambles on losses and framed "mixed" gambles also show systematic violations of CPT. This heavy burden of contrary data should tip the scales and shift the burden of proof to those who would continue to defend either version of prospect theory as a descriptive model. Majority choices are better described by Birnbaum's (1999) special TAX model, which uses no more parameters than CPT.

Supporters of CPT might argue that CPT was not intended as an accurate descriptive model in the sense that it should have predicted the results of new experiments, but merely as a way of summarizing previous data that refuted EU theory. The argument for CPT was that it could organize data on Allais paradoxes and other violations of EU by using additional parameters to represent probability weighting functions for positive and negative consequences. However, by the same reasoning, one should prefer TAX over CPT because it reproduces not only the Allais paradoxes (Birnbaum, 2004a, in press), but also the five new paradoxes studied here, whereas CPT fails to account for the new paradoxes. And the TAX model uses no more parameters than CPT. So, if one prefers CPT to EU because it accounts for Allais paradoxes, that same person should prefer TAX to CPT because TAX accounts for the phenomena described by CPT as well as five new ones (Table 1), which violate both CPT and EU.

A model that is similar to TAX and which can also account for the phenomena studied here (Table 1) is the rank affected multiplicative weights model (RAM). For the properties tested in this article, RAM and TAX are virtually identical. These models differ in their predictions, however, for other properties such as distribution independence and probability monotonicity. Tests between these models have favored TAX over RAM (Birnbaum, 2005a, 2005b).

Another model that also shows promise as a descriptive theory for these results is gains decomposition utility (GDU) (Marley & Luce, 2001), which Marley and Luce (2005) have shown makes similar predictions to the TAX model for the properties tested here on positive gambles. A special case known as the idempotent lower gains decomposition model, however, has been refuted by systematic violations of upper coalescing (Birnbaum, in press), the property that coalescing of the two branches leading to the highest consequences should have no effect. Other versions of GDU that violate idempotence remain to be tested.

TAX also handles six other phenomena not examined here, but which also violate or fail to support CPT. Those phenomena are violations of gain-loss separability (Birnbaum & Bahra, in press; Wu & Markle, 2004), violation of 4-distribution independence (Birnbaum & Chavez, 1997), satisfaction of 3-lower distribution independence and 3–2 lower distribution independence, violations of 3-upper distribution independence (Birnbaum, 2005b), and violations of upper tail independence (Birnbaum, 2005b; Wu, 1994).

Birnbaum (2004a, in press) conducted experimental dissections of the Allais paradox and found that dissected Allais paradoxes are not consistent with either version of prospect theory. According to original prospect theory, people should satisfy restricted branch independence, with or without the editing principle of cancellation. According to CPT, people should violate branch independence and satisfy coalescing, with or without the editing rule of combination. Data show that people violate coalescing and restricted branch independence. Furthermore, the violations of branch independence are opposite those needed by CPT to account for the Allais paradoxes. The paradoxes of Allais, however, are consistent with the hypothesis that they are produced by violations of coalescing.

The intuitions behind TAX, original prospect theory, and CPT are different, but to me, no one of these models seems more parsimonious or intuitively appealing in some a priori sense than the other two. All three models can be fit with the same number of parameters. All three models provide fits to Allais paradoxes in tests that confound branch independence and coalescing. TAX and CPT make virtually identical predictions for choices in which there are no more than three distinct consequences (i.e., studies trapped inside the probability simplex). Because restricted branch independence is defined on six distinct consequences, and the recipe for violation of stochastic dominance requires at least four distinct consequences (Table 1), these properties that refute CPT fall outside the probability simplex.

I find the intuitions of TAX, in which attention (weight) depends on branch probability and attention is transferred from branches leading to good consequences to branches leading to unfavorable consequences to be more plausible than the idea of CPT that people satisfy coalescing and work with decumulative probabilities. Nor do I find the intuition of original prospect theory appealing that people satisfy restricted branch independence, especially since this intuition runs counter to evidence.

Those who prefer to retain the intuitions of prospect theories in descriptive theory need to find a way to either modify experimental procedures so that data conform to predictions or to modify the theory to account for the data. The present results show that formats and procedures intended to "help" people satisfy coalescing, restricted branch independence, and stochastic dominance failed to produce data compatible with those properties. So, the conclusion I find more appealing is to favor models that are more accurate descriptions of the greater number of empirical phenomena over models that fit only some of the data.

Acknowledgements

Support was received from National Science Foundation Grants, SBR-9410572, SES 99-86436, and BCS-0129453. I thank Terry Connolly for suggesting the tickets/tables formats and Daniel Kahneman for useful exchanges of views on issues considered in this paper.

Appendix A. Cumulative prospect theory and new paradoxes

For binary gambles, $F = (x, p; y, 1 - p)$ where $x > y > 0$, CPT can be written:

$$\text{CPT}(F) = W(p)u(x) + [1 - W(p)]u(y)$$

where $W(p)$ is a strictly monotonic weighting function of probability with $W(0) = 0$ and $W(1) = 1$. For three-branch gambles, $G = (x, p; y, q; z, r); x \geqslant y \geqslant z \geqslant 0; p + q + r = 1$, CPT can be written:

$$\text{CPT}(G) = W(p)u(x) + [W(p + q) - W(p)]u(y) + [1 - W(p + q)]u(z)$$

Consequence monotonicity will be satisfied by CPT if and only if $u(x)$ is a strictly increasing monotonic function.

CPT implies both lower and upper coalescing. Proof: Suppose $x = y$. Then $\text{CPT}(G) = W(p)u(x) + [W(p + q) - W(p)]u(x) + [1 - W(p + q)]u(z) = W(p + q)u(x) + [1 - W(p + q)]u(z) = \text{CPT}(G')$, where $G' = (x, p + q; y, r)$, which shows that CPT satisfies upper coalescing. Now, suppose the two lower consequences are equal; $y = z$, then $\text{CPT}(G) = W(p)u(x) + [W(p + q) - W(p)]u(y) + [1 - W(p + q)]u(y) = W(p)u(x) + [1 - W(p)]u(y) = \text{CPT}(G'')$, where $G'' = (x, p; y, q + r)$, thus satisfying lower coalescing.

CPT satisfies stochastic dominance in Birnbaum's (1997) recipe (Table 1); i.e., $\text{CPT}(G^+) > \text{CPT}(G^-)$, where $G^+ = (x, p; y^+, q'; y, r - q')$ and $G^- = (x, p - q; x^-, q; y, r)$. Proof: $\text{CPT}(G^+) = W(p)u(x) + [W(p + q') - W(p)]u(y^+) + [1 - W(p + q')]u(y) > W(p)u(x) + [W(p + q') - W(p)]u(y) + [1 - W(p + q')]u(y) = W(p)u(x) + [1 - W(p)]u(y) = W(p - q)u(x) + [W(p) - W(p - q)]u(x) + [1 - W(p)]u(y) > W(p - q)u(x) + [W(p) - W(p - q)]u(x^-) + [1 - W(p)]u(y) = \text{CPT}(G^-)$. Therefore, $\text{CPT}(G^+) > \text{CPT}(G^-) \iff G^+ \succ G^- \iff \text{CPT}(GS^+) > CPT(GS^-) \iff GS^+ \succ GS^-$.

CPT implies lower cumulative independence. Proof: $S = (x, p; y, q; z, r) \succ R = (x', p; y', q; z, r) \iff W(p)u(x) + [W(p + q) - W(p)]u(y) + [1 - W(p + q)]u(z) > W(p)u(x') + [W(p + q) - W(p)]u(y') + [1 - W(p + q)]u(z)$. The common term can be subtracted from both sides, which yields, $W(p)u(x) + [W(p + q) - W(p)]u(y) > W(p)u(x') + [W(p + q) - W(p)]u(y')$. Therefore,

$$[W(p + q) - W(p)][u(y) - u(y')] > W(p)[u(x') - u(x)] \iff \frac{W(p + q) - W(p)}{W(p)} > \frac{u(x') - u(x)}{u(y) - u(y')}.$$

Suppose that $S'' = (x,p+q;y',r) \prec R'' = (x',p;y',q+r)$; this leads to contradiction:

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

because $u(x) > u(y)$. This contradiction shows that $S \succ R \Rightarrow S'' \succ R''$. This proof also reveals that violations represent a contradiction in the shape of the weighting function in CPT. Suppose we select consequences such that the ratio of differences in utility equals 1. If so, then an inverse-S weighting function implies that for small $p = q$, $\frac{W(2p)-W(p)}{W(p)-W(0)} < 1$; if so, then $R \succ S$ and $R'' \succ S''$. Instead, data show $R \prec S$ and $R'' \succ S''$; therefore, three-branch gambles violate the inverse-S shape, but two-branch gambles are consistent with it, and no version of CPT can reconcile the contradiction.

CPT implies upper cumulative independence. Proof: Suppose $S' = (z',r;x,p;y,q) \prec R' = (z',r;x',p;y',q)$. By CPT, $[W(p+r) - W(r)]u(x) + [1 - W(p+r)]u(y) < [W(p+r) - W(r)]u(x') + [1 - W(p+r)]u(y')$, so

$$\frac{1 - W(p+r)}{W(p+r) - W(r)} < \frac{u(x') - u(x)}{u(y) - u(y')};$$

but $S''' = (x',r;y,p+q) \succ R''' = (x',p+r;y',q)$ leads to contradiction, because

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

which implies,

$$\frac{1 - W(p+r)}{W(p+r) - W(r)} > \frac{u(x') - u(y)}{u(y) - u(y')} > \frac{u(x') - u(x)}{u(y) - u(y')}$$

because $u(x) > u(y)$. Now, suppose $p = q$. We again find that inverse-S weighting function implies $S' \succ R'$ and $S''' \succ R'''$. Instead, the data show $R' \succ S'$ and $S''' \succ R'''$, contradicting CPT and the inverse-S in three-branch gambles.

Appendix B. Transfer of attention exchange (TAX) model

For binary gambles, $F = (x,p;y,1-p)$ where $x > y > 0$ the special TAX model, assuming $\delta \geqslant 0$, can be written as follows:

$$\text{TAX}(F) = \frac{A'u(x) + B'u(y)}{A' + B'}$$

where $A' = t(p)[1 - \delta/3]$, $B' = t(1-p) + \delta t(p)/3$. Note that a certain fraction of the weight of the better consequence, $u(x)$, has been transferred to the less favored consequence, $u(y)$. It is this transfer of weight that accounts for risk aversion and for violations of restricted branch independence in this model. The "prior" TAX model assumes $u(x) = x$, for $0 < x < \$150$; $t(p) = p^{\gamma}$, where $\gamma = 0.7$; $\delta = 1$. [Note that $\delta = 1$ here corresponds to

$\delta = -1$ in Birnbaum (1999) and other articles that used the previous notational conventions.] These two parameters were selected to approximate the inverse-S relation between certainty equivalents of binary gambles in Tversky and Kahneman (1992), including risk aversion for 50–50 gambles.

For three-branch gambles, $G = (x,p;y,q;z,r); p + q + r = 1;\ x > y > z > 0$, the model can be written as follows:

$$\mathrm{TAX}(G) = \frac{Au(x) + Bu(y) + Cu(z)}{A + B + C}$$

where

$$A = t(p) - \delta t(p)/4 - \delta t(p)/4$$
$$B = t(q) - \delta t(q)/4 + \delta t(p)/4$$
$$C = t(r) + \delta t(p)/4 + \delta t(q)/4$$

With the parameters estimated from the data of Tversky and Kahneman (1992), TAX correctly predicts the results in Table 1. For example, $\mathrm{TAX}(G^-) = 63.1 > \mathrm{TAX}(G^+) = 45.8$, so the violation of stochastic dominance in Table 1 is predicted; however, in split form, $\mathrm{TAX}(GS^-) = 51.4 < \mathrm{TAX}(GS^+) = 53.1$, so the satisfaction of stochastic dominance in the split form and violation of coalescing are predicted.

Next, note that $\mathrm{TAX}(S) = 11.4 > \mathrm{TAX}(R) = 10.9$, but $\mathrm{TAX}(S'') = 16.2 < \mathrm{TAX}(R'') = 20.4$; therefore, the violation of lower cumulative independence is predicted. Similarly, $\mathrm{TAX}(S') = 65.0 < \mathrm{TAX}(R') = 69.6$, and $\mathrm{TAX}(S''') = 68.0 > \mathrm{TAX}(R''') = 58.3$, so the violation of upper cumulative independence is predicted. Finally, note that $\mathrm{TAX}(S) = 11.4 > \mathrm{TAX}(R) = 10.9$ and $\mathrm{TAX}(S') = 65.0 < \mathrm{TAX}(R') = 69.6$, so the pattern of violation of restricted branch independence is predicted. With the inverse-S probability weighting function, CPT predicts the opposite pattern of violation of restricted branch independence. And as noted in Appendix A, there is no way for CPT to account for violations of stochastic dominance, coalescing, lower cumulative independence, or upper cumulative independence.

Appendix C. Reflection cannot hold for all gambles

Definition of reflection: $\forall F, G$ as defined below:

$$G = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n) \succ F = (y_1, q_1; y_2, q_2; \ldots; y_m, q_m)$$
$$\iff -G = (-x_1, p_1; -x_2, p_2; \ldots; -x_n, p_n) \prec -F = (-y_1, q_1; -y_2, q_2; \ldots; -y_m, q_m)$$

If reflection holds for all mixed gambles, then it holds for symmetric gambles, such as the following: $A = -A$ and $B = -B$, where $-A$ indicates the same gamble as A, with all its consequences reflected. If reflection holds, then $A \succ B \iff A \prec B$, which is a self-contradiction, unless $A \sim B$.

For example, let $A = (\$100, .5; -\$100, .5)$; $B = (\$2, .5; -\$2, .5)$. According to reflection:

$$A = (\$100, .5; -\$100, .5) \prec B = (\$2, .5; -\$2, .5)$$
$$\iff -A = (-\$100, .5; \$100, .5) \succ -B = (-\$2, .5; \$2, .5) \iff A \succ B$$

Because its assumption leads to self-contradiction, the reflection hypothesis cannot hold for all mixed gambles, unless all such A and B are indifferent to each other. But empirically, it is often found that people tend to prefer the "safer" gamble in such cases; e.g., $B \succ A$ and $-B \succ -A$ in this example.

For purely non-negative gambles, however, this paper found that $P(A,B) \approx 1 - P(-A,-B)$ where $P(A,B)$ is the observed choice proportion choosing A over B. In sum, reflection cannot hold for all gambles, but it appears a good approximation for the strictly non-negative and non-positive gambles studied in Experiments 1 and 2.

References

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. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399–407.

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*(2), 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. (in press). Tests of branch splitting and branch-splitting independence in Allais paradoxes with positive and mixed consequences. *Organizational Behavior and Human Decision Processes*.

Birnbaum, M. H., & Bahra, J. (in press). Gain-loss separability and coalescing in risky decision making. *Management Science*.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71*(2), 161–194.

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

Harless, D. W. (1992). Actions vs. prospects: The effect of problem representation on regret. *American Economic Review, 82*, 634–649.

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

Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist, 58*(9), 723–730.

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

Keller, L. R. (1985). The effects of problem representation on the sure-thing and substitution principles. *Management Science, 31*, 738–751.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Savage, L. J. (1972). *The foundations of statistics* (2nd revised ed.). New York: Dover Publications, Inc.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332–382.

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

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

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

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

Wu, G., & Markle, A. B. (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. Available from http://gsbwww.uchicago.edu/fac/george.wu/research/abstracts.html.

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: Oxford University Press.