Research Article

TESTING CRITICAL PROPERTIES OF DECISION MAKING ON THE INTERNET

Michael H. Birnbaum

California State University, Fullerton, and Decision Research Center

Abstract—*This article reviews recent findings that violate a broad class of descriptive theories of decision making. A new study compared 1,224 participants tested via the Internet and 124 undergraduates tested in the laboratory. Both samples confirmed systematic violations of stochastic dominance and cumulative independence; new tests also found violations of coalescing. The Internet sample was older, more highly educated, more likely male, and also more demographically diverse than the lab sample. Internet participants were more likely than undergraduates to choose the gamble with higher expected value, but no one conformed exactly to expected value. Violations of stochastic dominance decreased as education increased, but violations of stochastic dominance and coalescing were still substantial in persons with doctoral degrees who had read a scientific work on decision making. In their implications, Internet research and lab findings agree: Descriptive decision theories cannot assume that identical consequences can be coalesced.*

Some people say that psychological science is based on research with rats, the mentally disturbed, and college students. We study rats because they can be controlled, the disturbed because they need help, and college students because they are available. The Internet now makes available a worldwide population. This new medium not only provides new research opportunities, but also raises new questions about sampling and experimental control (Krantz, Ballard, & Scher, 1997; Krantz & Dalal, in press). How do results from the Internet compare with those obtained in the laboratory?

This study explores this question with new tests that refute descriptive theories of decision making. Reviews of modern theories, including rank-dependent utility (RDU), rank- and sign-dependent utility (RSDU), and cumulative prospect theory (CPT), can be found in Quiggin (1993); Luce (1990, 1998); Luce and Fishburn (1991, 1995); Stevenson, Busemeyer, and Naylor (1991); Tversky and Kahneman (1992); Wakker and Tversky (1993); Weber (1994); and Wu and Gonzalez (1998). These modern theories account for phenomena that were not explained by earlier theories of Edwards (1954), Karmarkar (1979), and Kahneman and Tversky (1979). However, even these modern theories are now challenged by evidence with newly devised tests.

STOCHASTIC DOMINANCE

Not only is stochastic dominance considered rational, but it is also implied by many descriptive theories, including RSDU, RDU, CPT, and others (Becker & Sarin, 1987; Machina, 1982). A test of stochastic dominance is illustrated in Choice 5 of Table 1 (first row). Birnbaum (1997) proposed this choice as a test between theories that satisfy stochastic dominance and models that violate it. Two configural-weight models, the rank-affected multiplicative (RAM) and transfer-of-attention-exchange (TAX) models (with parameters estimated in previous studies), imply violation of stochastic dominance in this choice (Birnbaum, 1997, 1999; see the appendix).[1] Birnbaum and Navarrete (1998) found that 70% of undergraduates violated stochastic dominance by choosing $J = G–$ over $I = G+$, even though $G+$ dominates $G–$.

COALESCING AND EVENT SPLITTING

Coalescing also distinguishes decision theories (Birnbaum & Navarrete, 1998; Luce, 1998). Coalescing (see Table 2) assumes that branches with identical consequences can be combined (adding their probabilities), without affecting preference. Coalescing was assumed as an editing principle of prospect theory (Kahneman & Tversky, 1979), but it also follows from RSDU, RDU, CPT, and other theories (Luce, 1998).

Because stochastic dominance can be deduced from transitivity, consequence monotonicity, and coalescing (see Table 2), Birnbaum and Navarrete (1998) and Birnbaum, Patton, and Lott (1999) argued that violations of coalescing might cause violations of stochastic dominance, but these studies did not test coalescing directly. Starmer and Sugden (1993) and Humphrey (1995) reported event-splitting effects (violations of coalescing combined with transitivity); however, Luce (1998) described that evidence as "decidedly weak" (p. 91). The present study used strong, within-subjects tests to determine if event splitting can reverse violations of stochastic dominance.

According to coalescing, Choice 11 in Table 1 is the same as Choice 5, because $GS+$ (U in Table 1) is simply the split version of $G+$ (I) and $GS–$ (V) is the split version of $G–$ (J). Any theory assuming coalescing and transitivity implies $G+ \succ G–$ if and only if $GS+ \succ GS–$, where $\succ$ represents preference. The configural-weight RAM and TAX models with previously estimated parameters predict that people should prefer $G–$ over $G+$ in Choice 5 and $GS+$ over $GS–$ in Choice 11.

LOWER AND UPPER CUMULATIVE INDEPENDENCE

Birnbaum (1997) also devised two cumulative-independence conditions that test modern theories. Any theory that assumes comonotonic branch independence, consequence monotonicity, transitivity, and coalescing implies both lower and upper cumulative independence (see Table 2). Whereas RDU, RSDU, and CPT must satisfy these properties, RAM and TAX models violate them.

Address correspondence to Michael H. Birnbaum, Department of Psychology H-830M, California State University, Fullerton, P.O. Box 6846, Fullerton, CA 92834-6846; e-mail: mbirnbaum@fullerton.edu; Internet: http://psych.fullerton.edu/mbirnbaum/home.htm.

1. To compute predictions for CPT and the configural-weight TAX models, use Netscape to load the on-line calculator at URL http://psych.fullerton.edu/mbirnbaum/taxcalculator.htm. Additional information on model fitting (including source listings of computer programs) is available from this Web site.

Table 1. *Choices used to test models of risky decision making*

Choice number	Choice type		Choice		Percentage choosing the gamble on the right: Internet	Lab
5	*G*+	*G*–	*I*: .05 to win $12 .05 to win $14 .90 to win $96	*J*: .10 to win $12 .05 to win $90 .85 to win $96	58	73
6	*S*	*R*	*K*: .80 to win $2 .10 to win $40 .10 to win $44	*L*: .80 to win $2 .10 to win $10 .10 to win $98	69	58
7	*G*–	*G*+	*M*: .06 to win $6 .03 to win $96 .91 to win $99	*N*: .03 to win $6 .03 to win $8 .94 to win $99	54	36
8	*S*''	*R*''	*O*: .80 to win $10 .20 to win $44	*P*: .90 to win $10 .10 to win $98	75	69
9	*S*'''	*R*'''	*Q*: .20 to win $40 .80 to win $98	*R*: .10 to win $10 .90 to win $98	47	34
10	*S*'	*R*'	*S*: .10 to win $40 .10 to win $44 .80 to win $110	*T*: .10 to win $10 .10 to win $98 .80 to win $110	73	72
11	*GS*+	*GS*–	*U*: .05 to win $12 .05 to win $14 .05 to win $96 .85 to win $96	*V*: .05 to win $12 .05 to win $12 .05 to win $90 .85 to win $96	10	15
12	*R*'	*S*'	*W*: .05 to win $12 .05 to win $96 .90 to win $106	*X*: .05 to win $48 .05 to win $52 .90 to win $106	50	49
13	*GS*–	*GS*+	*Y*: .03 to win $6 .03 to win $6 .03 to win $96 .91 to win $99	*Z*: .03 to win $6 .03 to win $8 .03 to win $99 .91 to win $99	95	92
14	*R*'''	*S*'''	*a*: .05 to win $12 .95 to win $96	*b*: .10 to win $48 .90 to win $96	74	81
17	*R*	*S*	*g*: .90 to win $3 .05 to win $12 .05 to win $96	*h*: .90 to win $3 .05 to win $48 .05 to win $52	49	61
20	*R*''	*S*''	*m*: .95 to win $12 .05 to win $96	*n*: .90 to win $12 .10 to win $52	28	31

Note. Choice type refers to the notation used in the text and in Table 2.

Choices 6 and 8 of Table 1 test lower cumulative independence. Suppose $S \succ R$ in Choice 6. By comonotonic independence, we can increase the common branch in both gambles from $2 to $10, implying ($10, .8; $40, .1; $44, .1) $\succ$ ($10, .8; $10, .1; $98, .1). Increasing $40 to $44 should make *S* even better. Therefore, ($10, .8; $44, .1; $44, .1) $\succ$ ($10, .8; $10, .1; $98, .1); by coalescing, ($10, .8; $44, .2) $\succ$ ($10, .9; $98, .1); that is, $S'' \succ R''$ in Choice 8.

Upper cumulative independence is illustrated with Choices 10 and 9 of Table 1. Suppose *S*' = ($40, .1; $44, .1, $110, .8) $\prec$ *R*' = ($10, .1; $98, .1; $110, .8) in Choice 10. Reduce the (common) prize from $110 to $98 in both gambles. Reducing $44 to $40 makes *S*' even worse, so $S''' \prec R'''$ in Choice 9, by coalescing.

BRANCH INDEPENDENCE

Branch independence postulates that if two gambles have a common branch (the same prize with the same, known probability), then that common consequence can be changed without changing the preference order of the gambles. For example, *S* and *R* in Choice 6 (Table 1) have a common branch of .8 probability to win $2. Branch independence assumes that $2 can be changed to $110 in both gambles, so the preference between *S*' and *R*' in Choice 10 should match that in Choice 6; that is, $S \succ R$ if and only if $S' \succ R'$.

The theory that decision makers cancel common branches prior to choice (Kahneman & Tversky, 1979) implies branch independence; however, without this assumption, RDU, RSDU, and CPT violate it. The inverse-S weighting function and CPT model of Tversky and Kahneman (1992) predict that $R \succ S$ and $S' \succ R'$, opposite of what has been observed (Birnbaum & Chavez, 1997; Birnbaum & McIntosh, 1996; Birnbaum et al., 1999).

Because violations of stochastic dominance and cumulative independence potentially refute a large class of descriptive theories, it is vital to know if laboratory studies hold up outside the lab with people who are not college students and who make choices with real monetary

Table 2. *Testable properties of decision-making theories*

Property name	Illustration
Transitivity	$A \succ B$ and $B \succ C \Rightarrow A \succ C$
Consequence monotonicity	$(x^+, p; y, q; z, r) \succ (x, p; y, q; z, r) \Leftrightarrow x^+ \succ x$
Coalescing	$(x, p; x, q; z, r) \sim (x, p + q; z, r)$
Stochastic dominance	$P(x > t \mid A) \geq P(x > t \mid B)\ \forall\ t \Rightarrow A \succ B$ or $A \sim B$
Restricted branch independence	$S = (x, p; y, q; z, r) \succ R = (x', p; y', q; z, r) \Leftrightarrow S' = (x, p; y, q; z', r) \succ R' = (x', p; y', q; z', r)$
Lower cumulative independence	$S = (z, r; x, p; y, q) \succ R = (z, r; x', p; y', q) \Rightarrow S'' = (x', r; y, p + q) \succ R'' = (x', r + p; y', q)$
Upper cumulative independence	$S' = (x, p; y, q; z', r) \prec R' = (x', p; y', q; z', r) \Rightarrow S''' = (x, p + q; y', r) \prec R''' = (x', p; y', q + r)$

Note. The following notation is used in this table: Let $A = (x, p; y, q; z, r)$ represent a gamble to win x with probability p, y with probability q, and z with probability r, where $p + q + r = 1$. $A \succ B$ means A is preferred to B, and ~ represents indifference. $P(x > t \mid A)$ represents the probability of winning more than t given A. Branch independence is *restricted* when the number of distinct branches and their probabilities are fixed. *Comonotonic* branch independence is the special case in which consequences maintain the same ranks. In tests of (noncomonotonic) branch independence and cumulative independence, $0 < z < x' < x < y < y' < z'$.

consequences. If event splitting can reverse the violations, it pinpoints coalescing as the property that must be revised in descriptive theory.

METHOD

Participants completed the experiment on-line by visiting the World Wide Web site (now retired) at URL http://psych.fullerton.edu/mbirnbaum/exp2a.htm. Instructions in the page included the following:

Would you rather play:
A: fifty-fifty chance of winning either $100 or $0 (nothing),
OR
B: fifty-fifty chance to win either $25 or $35.
. . . [If people choose A] . . . half the time they might win $0 and half the time $100. But in this study, you only get to play a gamble once, so the prize will be either $0 or $100. Gamble B's bag has 100 tickets also, but 50 of them say $25 and 50 of them say $35. Bag B thus guarantees at least $25, but the most you can win is $35 . . . .
For each choice below, click the button beside the gamble you would rather play . . . after people have finished their choices . . . [1% of participants] will be selected randomly to play one gamble for real money . . . . [If you are selected], you will get to play the gamble you chose on the trial selected. . . . Any one of the 20 choices might be the one you get to play, so choose carefully.

Games were played as promised, and 11 participants won $90 or more; 8 won smaller prizes.

Stimuli

Gambles were displayed as in the following example:

◉ 1. Which do you choose?

◯ A: .50 probability to win $0
.50 probability to win $100

OR

◯ B: .50 probability to win $25
.50 probability to win $35

Design

The 20 choices are listed in Tables 1 and 3. There were two tests each of stochastic dominance, event splitting, lower cumulative independence, upper cumulative independence, and branch independence, with position (first or second gamble) counterbalanced.

Questionnaire

The form requested the participant's e-mail address (so prizewinners could be contacted), country, age, gender, and education. Subjects were also asked the yes/no question "Have you ever read a scientific paper (i.e., a journal article or book) on the theory of decision making or the psychology of decision making?" Participants were invited to write comments in a box.

Recruitment, Procedure, and Reliability

Lab sample

The 124 undergraduates came from the subject pool and served as one option toward completing an assignment in introductory psychology. In the lab, the experimental Web page was displayed on several computers. Experimenters checked that participants could use the mouse to click and to scroll through the page. After completing the form by clicking the "submit" button, each lab participant repeated the same task on a fresh page.

Reliability of lab data

The mean number of agreements between the first and second repetitions was 16.4 (82% agreement). The median was 17, and 80% of participants made 15 or more identical choices (≥75% agreement). The within-person correlation between choices, averaged over participants, was .63, which is higher than .24, the average correlation between different people. When analyzed separately, data from the two replicates led to the same conclusions, so they are combined in the following presentation, except where noted.

Internet sample

The Internet sample consisted of 1,224 participants from 44 nations who completed the experiment on-line. The Web site was announced by e-mail sent to members of the Societies for Judgment and Decision Making and for Mathematical Psychology. It was suggested to major search engines and to Web sites that list contests and games with prizes.

Table 3. *Choices used to assess risk attitudes and consequence monotonicity*

Choice number	Choice		Percentage choosing the gamble on the right: Internet	Lab
1	*A*: .50 to win $0 .50 to win $100	*B*: .50 to win $25 .50 to win $35	48	58
2	*C*: .50 to win $0 .50 to win $100	*D*: .50 to win $45 .50 to win $50	60	69
3	*E*: .50 to win $4 .30 to win $96 .20 to win $100	*F*: .50 to win $4 .30 to win $12 .20 to win $100	6	8
4	*G*: .40 to win $2 .50 to win $12 .10 to win $108	*H*: .40 to win $2 .50 to win $96 .10 to win $108	96	94
15	*c*: $1 for sure	*d*: .99 to win $0 .01 to win $100	58	55
16	*e*: $3 for sure	*f*: .99 to win $0 .01 to win $100	43	50
18	*i*: $90 for sure	*j*: .01 to win $0 .99 to win $100	33	38
19	*k*: $96 for sure	*l*: .01 to win $0 .99 to win $100	26	32

Demographic Characteristics of the Samples

The lab sample ranged from 18 to 28 years old; 91% were 22 and under. The Internet sample ranged from 18 to 86 years old, with 77% over 22, 50% over 28, and 20% above 40. Of the lab sample, 91% had 3 years of college or less (none had degrees). In the Internet sample, 60% were college graduates, including 333 who reported postgraduate studies (134 had doctoral degrees). Only 1% of the Internet sample had less than 12 years of education. Seventy-three percent of lab subjects and 56% of the Internet sample were female. Of the lab sample, 13% indicated having read a scientific work on decision making, compared with 31% of the Internet sample.

All lab subjects were residents of the United States, whereas the Internet sample represented 44 different nations. Countries with 8 or more people were Australia (23), Canada (66), Germany (42), The Netherlands (62), Norway (12), Spain (8), the United Kingdom (36), and the United States (896).

RESULTS

Comparison of Choice Percentages

Tables 1 and 3 show the percentages of preferences for each gamble. The Internet and lab samples gave similar choice percentages (correlation = .94). Six choices showed differences of 10% or more: Choices 1, 5, 6, 7, 9, and 17; in these cases, the Internet sample was more likely to choose the gamble with higher expected value (EV). For example, in Choice 1, 58% of the lab sample (against 48% of the Internet sample) chose the “safe” gamble (to win $25 or $35) rather than the “risky” gamble (to win either $0 or $100), even though the risky gamble has the higher EV ($50 vs. $30).

Risk aversion refers to preference for a sure gain over a risky gamble with the same or higher EV. The results for risk aversion are consistent with previous findings (e.g., Tversky & Kahneman, 1992): Majorities of both samples were risk averse for gambles with medium or high probabilities to win (Choices 2, 18, and 19). Also, majorities of both samples were risk seeking when the probability to win was small (Choice 15).

Consequence Monotonicity

If two gambles are identical except for the value (values) of one or more consequences, then the gamble with the higher consequence (consequences) should be preferred. There were four direct tests of consequence monotonicity (Choices 3, 4, 11, and 13). There were also six choices that indirectly tested monotonicity. Consider Choices 1 and 2 of Table 3. If a person prefers *B* over *A* in Choice 1, then that person should also prefer *D* over *C* in Choice 2 because *A* is the same as *C*, and *D* dominates *B*. The term *indirect* is used because the test involves transitivity as well as monotonicity, not to mention memory.

Only 7 of the 1,224 Internet participants violated indirect monotonicity on Choices 1 and 2; for the lab sample, 5 choices of 248 were violations. For Choices 15 and 16, 12 Internet and 6 lab participants chose *c* over *d* and *f* over *e*. For Choices 18 and 19, there were 38 participants from the Internet and 10 in the lab who chose *i* over *j* and *l* over *k*. The mean rates of violation of indirect monotonicity were 1.6% and 2.8% for the Internet and lab samples, respectively.

For Choices 3, 4, 11, and 13 (direct tests of consequence monotonicity), there were 71, 42, 54, and 127 violations in the Internet sample and 21, 14, 36, and 21 violations in the lab sample, respectively. The average rates of violation in direct tests of monotonicity were therefore 6.0% and 9.3%. Interestingly, indirect tests, which were on

Table 4. *Percentages of each choice combination in tests of stochastic dominance and event splitting*

Sample	*G*+*GS*+	*G*+*GS*–	*G*–*GS*+	*G*–*GS*–
		Choices 5 and 11		
Internet	37.4	4.7	**51.6***	5.8
Lab	21.8	4.8	**62.5***	9.7
		Choices 7 and 13		
Internet	51.3	2.1	**43.7***	2.3
Lab	33.1	2.8	**58.5***	5.6

Note. Percentages are based on n = 1,224 for the Internet sample and n = 248 (2 replications × 124 judges) for the lab sample; percentages do not always sum to 100 because of occasional nonresponse. Entries in boldface indicate preference reversals predicted by the rank-affected multiplicative model and transfer-of-attention-exchange models—namely, that $G– \succ G+$ and $GS+ \succ GS–$. In each row, the asterisk indicates that violations of stochastic dominance are significantly more frequent when the gambles are presented in coalesced form than in split form. In three of the four rows, significantly more than half of the subjects violated stochastic dominance in the coalesced form.

adjacent trials, showed fewer violations than direct tests. When the data were analyzed separately for first and second replications in the lab sample, there were slightly fewer violations in the second replicate. If consequence monotonicity were considered an index of the quality of the data, then the Internet data would be judged higher in quality than the lab data.

Stochastic Dominance and Event Splitting

Table 4 shows the results for choices testing stochastic dominance and event splitting for the Internet and laboratory samples. Entries are percentages of each combination of preferences in Choices 5 and 11 and in Choices 7 and 13 of Table 1. If everyone satisfied stochastic dominance, then 100% would have chosen *G*+ and *GS*+. Instead, half or more of the choice combinations for Choices 5 and 11 were $G– \succ G+$ and $GS+ \succ GS–$.

To compare probabilities of choosing *G*+ over *G*– against *GS*+ over *GS*–, one can use the test of correlated proportions. This binomial sign test compares the frequencies of reversal combinations (i.e., *G*–*GS*+ against *G*+*GS*–). For example, 632 of 1,224 Internet participants violated stochastic dominance by choosing *G*– over *G*+ on Choice 5 and switched preferences by choosing *GS*+ over *GS*– on Choice 11, compared with only 57 who showed the opposite combination of preferences. In this case, the binomial has μ = 344.5 [(632 + 57)/2], with σ = 13.12; therefore, the value of z is 21.91, which is significant.[2]

One can also test separately the (conservative) hypothesis that 50% of people violated stochastic dominance on Choice 5 by using the binomial sign test on the split of the 704 (57.5%) who chose *G*– over *G*+ against the 519 who chose *G*+ over *G*–; for this test, the value of z is 5.29. "Only" 46% of the Internet sample violated stochastic dominance in Choice 7; however, significantly more had the *G*–*GS*+ combination of preferences on Choices 7 and 13 than had the opposite switch, z = 21.49.

Significantly more than half of the lab sample violated stochastic dominance in both tests, averaging 68.3% violations. The lab sample also showed significant event-splitting effects. The lab sample had higher rates of violation of stochastic dominance than the Internet sample, a difference explored in terms of demographic correlates in a later section.

Table 5. *Percentages of choice combinations in tests of lower cumulative independence*

Sample	*SS*"	*SR*"	*RS*"	*RR*"
		Choices 6 and 8		
Internet	12.7	**18.0***	12.0	56.5
Lab	18.5	**23.0***	11.7	46.0
		Choices 17 and 20		
Internet	18.5	**30.3***	9.6	40.4
Lab	21.0	**39.5***	10.1	29.0

Note. Percentages are based on n = 1,224 for the Internet sample and n = 248 (2 replications × 124 judges) for the lab sample; percentages do not always sum to 100 because of occasional nonresponse. Entries in boldface designate preference shifts that violate lower cumulative independence; asterisks indicate cases in which the violation pattern is significantly more frequent than the opposite shift of preference (which would be consistent with the property).

Cumulative Independence

Table 5 shows the results for choices testing lower cumulative independence. Recall that a shift from $S \prec R$ to $S'' \succ R''$ would be consistent with the property; however, the opposite shift, $S \succ R$ and $S'' \prec R''$, would be a violation. Table 5 shows that violations (boldface type) significantly exceeded shifts that were consistent with the property in all four tests.

Results for choices testing upper cumulative independence (if $S' \prec R'$ then $S''' \prec R'''$) are shown in Table 6. It is a violation to prefer *R*' and *S*'''. Table 6 shows significantly more violations (boldface) than switches that were consistent in all four tests.

Branch Independence

Table 7 shows preferences for choices testing branch independence. Asymmetry of the preference shifts indicates systematic deviations. Consistent with previous tests (Birnbaum & Beeghley, 1997; Birnbaum & Chavez, 1997; Birnbaum & McIntosh, 1996; Birnbaum & Navarrete, 1998; Birnbaum et al., 1999), there were significantly more violations of the *SR*' type (boldface) than of the opposite type (*RS*') in three of four tests (the fourth was not significant).

Demographic Correlates

Because the Internet sample was large and diverse, it was possible to subdivide the sample by gender, education level, experience reading a scientific work on decision making, and nationality. The data were then analyzed as in Tables 4 through 7 within each of these subdivisions.

2. Throughout this article, "significant" indicates that $p < .05$. The critical value of z for a two-tailed test is 1.96; values greater than 1.96 are significant. For tests with $n < 30$, exact binomial probabilities were calculated.

Table 6. *Percentages of choice combinations in tests of upper cumulative independence*

Sample	*S'S'''*	*S'R'''*	*R'S'''*	*R'R'''*
	Choices 10 and 9			
Internet	19.2	6.9	**33.3***	39.7
Lab	23.0	3.6	**42.7***	29.4
	Choices 12 and 14			
Internet	42.2	7.8	**31.8***	17.5
Lab	42.3	6.5	**38.3***	12.9

Note. Percentages are based on n = 1,224 for the Internet sample and n = 248 (2 replications × 124 judges) for the lab sample; percentages do not always sum to 100 because of occasional nonresponse. Entries in boldface designate preference shifts that violate upper cumulative independence; asterisks indicate cases in which the violation pattern is significantly more frequent than the opposite shift of preference (which would be consistent with the property).

Table 7. *Percentages of choice combinations in tests of branch independence*

Sample	*SS'*	*SR'*	*RS'*	*RR'*
	Choices 6 and 10			
Internet	12.0	**18.7***	14.1	54.3
Lab	18.5	**22.2***	7.7	50.0
	Choices 17 and 12			
Internet	32.0	**17.1**	17.8	32.4
Lab	37.5	**23.4***	11.3	27.8

Note. Percentages are based on n = 1,224 for the Internet sample and n = 248 (2 replications × 124 judges) for the lab sample; percentages do not always sum to 100 because of occasional nonresponse. Entries in boldface designate violations of branch independence of the type predicted by the rank-affected multiplicative and transfer-of-attention-exchange models (with parameters estimated from previous research). Asterisks indicate cases in which the previously observed *SR'* pattern of violation significantly exceeds the opposite pattern of violation, *RS'*.

These separate tests led to essentially the same conclusions (i.e., there were significant violations of stochastic dominance, event-splitting effects, and violations of upper and lower cumulative independence).

However, the incidence of violations of stochastic dominance correlated with education and gender. Table 8 shows the relationship between violations of stochastic dominance (Choice 5), violations of consequence monotonicity (Choice 11), education, and gender. In each group (each row), violations of stochastic dominance exceed the corresponding violations of consequence monotonicity, which violates coalescing. Females with bachelor's degrees had 65% violations of stochastic dominance on Choice 5, and equally educated males had 52.3% violations. For the 54 females with doctorates, there were 40.7% violations; for males with doctorates, the rate was 48.8%. Similar results (not shown) were observed for Choices 7 and 13. Of the 686 females, 60.3% violated stochastic dominance on Choice 5, and 55.1% violated stochastic dominance on Choice 7. Of the 526 males, 53.4% and 34.6% violated stochastic dominance on Choices 5 and 7, respectively. These rates were much higher than corresponding rates of violation of consequence monotonicity, which were 11.8% and 5.7% for females on Choices 11 and 13; for males, they were 8.4% and 2.5%, respectively.

Violations of stochastic dominance were also less frequent among those who had read a scientific work on decision making. Of the 837 people who had not read about decision making, 59.9% violated stochastic dominance on Choice 5; among the 382 who had read such a work, Choice 5 had 52.6% violations; this modest difference is significant, $\chi^2(1) = 5.41$.

Among those participants who had read about decision making were 95 who also held doctorates, including many members of the Society for Judgment and Decision Making. This expert group had 50% violations of stochastic dominance on Choice 5. This group included 46 participants with the preference combination *G–GS+* against only 7 with the combination *G+GS–*, which is a significant split, z = 5.36. Only 32 of these 95 violated stochastic dominance on Choice 7, but 30 of these switched preferences from *G–* in Choice 7 to *GS+* in Choice 13, compared with none who had the opposite

Table 8. *Violations of stochastic dominance and monotonicity related to gender and education*

Gender	Education[a]	Stochastic dominance (%) $G- \succ G+$	Monotonicity (%) $GS- \succ GS+$	Number of subjects
Female	<16	62.3 (75.3)	12.6 (14.3)	318 (91)
Female	16	65.0	13.1	206
Female	17–19	55.6	12.0	108
Female	20	40.7	1.9	54
Male	<16	60.1 (65.2)	9.8 (15.2)	163 (33)
Male	16	52.3	10.3	195
Male	17–19	47.7	2.3	88
Male	20	48.8	11.2	80

Note. Lab sample results are shown in parentheses. Percentages for stochastic dominance and monotonicity represent violations based on Choices 5 and 11, respectively.
[a]Education <16 indicates less than bachelor's degree; 16 = bachelor's degree; 17–19 = postgraduate studies; 20 = doctorate.

switch, which represents a significant event-splitting effect in this expert group. The expert group had 41.6% violations, averaged over the two tests, compared with 68.3% for the lab sample of undergraduates.

Violations of cumulative independence did not appear to be systematically related to gender, education, or other demographic characteristics. For example, among the 95 participants in the expert group, 31 had the preference order *SR*" for Choices 17 and 20, violating the property, against 8 who had the opposite pattern, *RS*", a significant violation of lower cumulative independence. For Choices 12 and 14, these 95 showed 25 preference combinations *R'S'*", in violation of upper cumulative independence, against only 5 combinations *S'R*'". These splits are both statistically significant (by exact binomial sign tests), indicating significant violations among the expert group, but the rates of violation are similar to the overall rates in Tables 5 and 6.

There were 328 subjects from nations outside the United States. They were more highly educated on average than the U.S. sample (e.g., 62 had doctorates). Their data are similar to the data for the U.S. participants, once education is considered. Correlations with gender and education were also observed. For example, of the 59 foreign females with bachelor's degrees, 66% violated stochastic dominance on Choice 5, compared with 44% of the 64 foreign males with bachelor's degrees. There were 41 with doctoral degrees who had read on decision making; these had an average of 44% violations of stochastic dominance against 8.5% violations of consequence monotonicity.

DISCUSSION

The results show systematic violations of stochastic dominance, lower cumulative independence, and upper cumulative independence. These violations contradict the implications of a class of decision models including RDU, RSDU, and CPT, but they are consistent with configural-weight RAM and TAX models. Although there are differences between the Internet and lab results, the two sets of results lead to the same conclusions concerning these properties.

Violations of stochastic dominance are largely eliminated by event splitting, suggesting that coalescing is the key to the violations of stochastic dominance, cumulative independence, and ordinal independence observed here and in previous research (Birnbaum & Navarrete, 1998; Wu, 1994). Event splitting also violates this class of RDU, RSDU, and CPT models. The appendix presents a comparison of fit, which shows that a TAX model of configural weighting predicts more choices correctly than a CPT model with the same number of estimated parameters.

The Internet and lab samples yielded similar conclusions, indicating that the findings are not unique to lab studies of college students. Internet research has two potential problems, sampling and control. In the lab, investigators can ensure that subjects do not use calculators to compute expected value, for example, or can require them to do so. With an Internet study, there is less control over the conditions. Investigators can ask people to follow instructions, and can ask them if they did, but it is not possible to know for certain. One might hope that variations of conditions would simply introduce random error that would be overcome in large samples. Ultimately, investigators must rely on honesty, indirect checks, or the hope that deviations of protocol do not matter to the case at hand.

With respect to sampling, the Internet is not really a single population, but many. The demographics of Internet users are changing, and variations in recruitment potentially have powerful effects. This study used methods intended to reach a highly educated population knowledgeable in decision making. The fact that the recruited participants included 95 who had doctorates and had studied decision making suggests that the recruitment succeeded.

Although one can use methods intended to reach certain groups, Internet experimenters do not have complete control over recruitment. Another well-meaning person can place a link to a study in a place that recruits people the original investigator did not intend to sample.

If demographic or individual difference variables affect behavior, then one can measure these and study their correlations with the results. In the present study, rates of violation of stochastic dominance were found to correlate with gender, education, and experience reading a scholarly work on decision making. The Internet sample was less likely to violate stochastic dominance than the lab sample, but the Internet sample also had a lower percentage of females, had more education, and was older; Internet participants also were more likely to have read work on decision making. Thus, the differences in results fit those expected from the demographic differences between the samples.

Education, which correlated with violations of stochastic dominance, probably correlated with unmeasured variables that might have been the actual causal agents. Individuals with more education are probably also higher in intelligence and wealth. Therefore, fewer violations of stochastic dominance among the highly educated might be due to higher intelligence, for example, rather than to the causal effects of education. Experiments with differential education will test if specific training teaches people to satisfy stochastic dominance.

In summary, the Internet data clarify and reinforce the results from the lab. The data indicate that descriptive theory should not assume coalescing. Dropping coalescing allows violations of rationality (e.g., stochastic dominance), but evidence shows that people violate implications of coalescing even when there are monetary consequences. Thus, these and other data (Birnbaum & Navarrete, 1998; Birnbaum et al., 1999) call for a major revision in theories of decision making. Results can be approximated by the configural-weight TAX model, which implies violations of stochastic dominance, event-splitting effects, and violations of cumulative independence.

Acknowledgments—Support was received from National Science Foundation Grant SBR-9410572. Thanks are due Jamie N. Patton, Melissa Lott, Juan Navarrete, and Dana Storm for assistance in collecting the laboratory sample.

REFERENCES

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

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). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B.A. Mellers, & D.A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27–52). Norwell, MA: Kluwer Academic.

Birnbaum, M.H., & Beeghley, D. (1997). Violations of branch independence in judgments of the value of gambles. *Psychological Science, 8,* 87–94.
Birnbaum, M.H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71,* 161–194.
Birnbaum, M.H., & McIntosh, W.R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67,* 91–110.
Birnbaum, M.H., & Navarrete, J.B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17,* 49–78.
Birnbaum, M.H., Patton, J.N., & Lott, M.K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes, 77,* 44–83.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380–417.
Humphrey, S.J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11,* 263–274.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.
Karmarkar, U.S. (1979). Subjectively weighted utility and the Allais paradox. *Organizational Behavior and Human Performance, 24,* 67–72.
Krantz, J.H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and world-wide web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers, 29,* 264–269.
Krantz, J.H., & Dalal, R. (in press). Validity of web-based psychological research. In M.H. Birnbaum (Ed.), *Psychological experiments on the Internet.* San Diego: Academic Press.
Luce, R.D. (1990). Rational versus plausible accounting equivalences in preference judgments. *Psychological Science, 1,* 225–234.
Luce, R.D. (1998). Coalescing, event commutativity, and theories of utility. *Journal of Risk and Uncertainty, 16,* 87–113.
Luce, R.D., & Fishburn, P.C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty, 4,* 29–59.
Luce, R.D., & Fishburn, P.C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty, 11,* 5–16.
Machina, M.J. (1982). Expected utility analysis without the independence axiom. *Econometrica, 50,* 277–323.
Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model.* Boston: Kluwer.
Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6,* 235–254.
Stevenson, M.K., Busemeyer, J.R., & Naylor, J.C. (1991). Judgment and decision-making theory. In M. Dunnette & L.M. Hough (Eds.), *New handbook of industrial-organizational psychology* (pp. 283–374). Palo Alto, CA: Consulting Psychologists Press.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–323.
Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica, 63,* 1255–1280.
Wakker, P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty, 7,* 147–176.
Weber, E.U. (1994). From subjective probabilities to decision weights: The effects of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 114,* 228–242.
Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty, 9,* 39–60.
Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty, 16,* 115–139.

(RECEIVED 8/25/98; REVISION ACCEPTED 2/2/99)

APPENDIX: COMPARISON OF MODELS

The configurally weighted utility (CWU) of a gamble can be written as follows:

$$CWU(G) = \sum_{i=1}^{n} w(x_i, G)u(x_i) , \qquad (1)$$

where $G = (x_1, p_1; x_2, p_2; \ldots ; x_i, p_i; \ldots x_n, p_n)$ is a gamble with n distinct positive outcomes, ranked such that $0 < x_1 < x_2 < \ldots < x_i < \ldots < x_n$; $\sum_{i=1}^{n} p_i = 1$; $u(x_i)$ is the utility of the outcome; and $w(x_i, G)$ is its weight. All models discussed here—RSDU, RDU, CPT, RAM, TAX, EU (expected utility), and EV—are special cases of Equation 1, with different assumptions about the weights. Nonlinear utility models in which weights depend on the utilities of the consequences are also special cases of Equation 1.

For positive outcomes, RSDU reduces to RDU, which assumes that weights are as follows:

$$w(x_i, G) = W\Big(\sum_{j=i}^{n} p_j\Big) - W\Big(\sum_{j=i+1}^{n} p_j\Big) \qquad (2)$$

where $W(P)$ is a strictly monotonic function that assigns decumulative weight to decumulative probability, $P_i = \sum_{j=i}^{n} p_j$, where $W(0) = 0$ and $W(1) = 1$. Equation 2 implies stochastic dominance (Quiggin, 1993), coalescing (Luce, 1998), and cumulative independence (Birnbaum, 1997).

The model of CPT (Tversky & Wakker, 1995) further assumes that $W(P)$ in Equation 2 is given by the following:

$$W(P) = \frac{cP^{\gamma}}{cP^{\gamma} + (1-P)^{\gamma}},$$

where c and γ have been estimated to be less than 1, giving $W(P)$ an inverse-S shape.

The TAX model assumes that weights are transferred among branches (distinct probability-consequence pairs) according to the ranks of consequences and the weight each branch has to lose. When lower outcomes have greater configural weight, lower-valued branches "tax" weight from higher-valued ones; with $\rho < 0$, relative weights are as follows:

$$w(x_i, G) = \frac{S(p_i) + \rho \sum_{j=1}^{i-1} S(p_i) - \rho \sum_{j=i+1}^{n} S(p_j)}{\sum_{j=1}^{n} S(p_j)},$$

where $S(p_i)$ is a function of the probability to win x_i; the weight given up by this branch is $\rho \sum_{j=1}^{i-1} S(p_i)$, indicating that this branch gives up weight to all branches with lower consequences than x_i ($\rho < 0$). This branch, in turn, takes weight from branches with higher consequences. Birnbaum and Chavez (1997) approximated $\rho = \delta/(n + 1)$, $S(p) = p^{\gamma}$, and $u(x) = x$, for $0 < x < \$150$.

Subjectively weighted utility (SWU) theory is the nonconfigural, special case of Equation 1 in which $w(x_i, G) = w(p_i)$; $SWU(G) = \sum_{i=1}^{n} w(p_i)u(x_i)$. EU theory assumes $w(x_i, G) = p_i$; $EU(G) = \sum_{i=1}^{n} p_i u(x_i)$. EV is the special case of EU where $u(x_i) = x_i$; $EV(G) = \sum_{i=1}^{n} p_i x_i$. Although EV and EU have been rejected in previous studies (e.g., Kahneman & Tversky, 1979), they provide benchmarks for assessing the accuracy of more complex models.

TAX, CPT, EU, and EV models were fit to compare the relative accuracy of the models in describing individual data. Each person's data were fit (minimizing the negative log likelihood of the choices given each model using methods of Birnbaum & Chavez, 1997). After a model was fit to each person, utility differences were computed to see if the model correctly predicted each choice. The percentages of subjects for whom the models predicted 15 or more choices (75% accuracy or better) are given in Table A1.

The TAX model can account for violations of stochastic dominance, event-splitting effects, and violations of lower and upper cumulative independence. The model was fit assuming $u(x) = x$. In the Internet sample, median estimates

Table A1. *Percentage of cases for which 15 or more choices (of 20) are correctly predicted by four models*

	Model			
Sample	Transfer of attention exchange	Cumulative prospect theory	Expected utility	Expected value
Internet (n = 1,224)	65	46	23	7
Lab 1 (n = 124)	67	58	37	17
Lab 2 (n = 124)	60	43	28	9

Note. Lab 1 and Lab 2 refer to the first and second replicates in the lab data, respectively. Lab 1 choice correctly predicted 15 or more choices in Lab 2 in 80% of cases.

of γ and δ for the TAX model are .791 and –.333, respectively. The mean number of choices correctly predicted was 15.53 (78%), and the model had perfect scores of 20 for 66 people. For the lab sample, mean correct predictions in the first and second replicates were 14.75 and 15.36; median estimates were γ = .982 and .681 and δ = –.463 and –.574, respectively. Correlations between replicates were .31, .726, and .543, for γ, δ, and number correct, respectively, and all were significant. The TAX model was not quite as accurate on average in the lab (mean = 15.05, or 75%) as were other choices by the same subject (16.4, or 82%).

For the CPT model in the lab, means of correct predictions were 14.22 and 13.99; median estimates were γ = .759 and .749 and c = .481 and .438, respectively. Correlations between replicates were .63 and .56 for γ and c, respectively. For the Internet sample, medians were γ = .743 and c = .597, respectively. For the Internet data, CPT had an average of 14.91 correct (75%), significantly worse than the TAX model, $t(1223) = –8.05$. Within-person, the TAX model predicted more choices correctly than CPT for 614 people; 414 had more correct predictions by CPT, and 196 were tied. Thus, significantly more people were fit better by TAX than by CPT, z = 6.24.

For EU, utility was approximated by $u(x) = x^{ß}$. For the Internet sample and the first and second replicates of the lab sample, median estimates of ß were .611, .587, and .585, respectively (correlation between lab replicates = .89). On average, EU correctly predicted 13.55, 12.93, and 12.43 choices correctly (68%, 65%, and 62%) in the Internet sample and two lab replicates, respectively.

No one in either sample was perfectly consistent with EV. This result seemed a bit surprising because a number of people (including several with doctorates) sent comments stating they simply chose by EV. However, no one wrote that he or she actually computed EV, and apparently no one did. For the Internet sample and two lab replicates, EV correctly predicted a mean of 12.4 choices (62%), 10.44 choices (52%), and 10.15 choices (51%), respectively; no one in the lab sample had more than 17 choices consistent with EV.

In summary, TAX, CPT, EU, and EV models predicted 75% or more of individual choices in 65%, 47%, 25%, and 8% of the fits, respectively.