A Web-Based Program of Research on Decision Making

Michael H. Birnbaum[1]

I started in Web-based research because I wanted to test a specialized sample of highly educated people to check results found with college students.

I previously published some interesting predictions of configural-weight models that had not yet been tested (Birnbaum, 1997). These configural models, fit to previous data, implied that people would violate stochastic dominance in choosing between specially constructed gambles. This was a striking prediction because many other current theories imply that people would conform to stochastic dominance, which is also a convincingly rational tenant of decision making. If one gamble always has as high or higher a probability of a better consequence than another, it is hard to argue that one should choose the dominated gamble.

In laboratory research, Juan Navarrete and I tested critical predictions of a configural weight model of decision making against those of the class of rank-dependent expected utility models that were at the time considered plausible descriptive models (Luce, 2000; Luce & Fishburn, 1991; 1995; Tversky & Kahneman, 1992; Quiggin, 1993). The results were quite startling, even to me (Birnbaum & Navarrete, 1998). We found that about 70% of undergraduates systematically violated stochastic dominance in specially constructed choices.

Because this finding violated both normative and descriptive theory, it seemed important to check if this result would also be observed in people besides psychology students. Although the result was predicted from a model fit to other student data, it was possible that these results were limited only to students. Perhaps students do not understand probability or misunderstand the task for some other reason. I wanted to test people holding doctorate degrees who had read at least one book or scientific article on decision making.

---

[1] California State University, Fullerton and Decision Research Center, Fullerton, USA.
E–mail: mbirnbaum@fullerton.edu URL: http://psych.fullerton.edu/mbirnbaum/home.htm
This research was supported by National Science Foundation Grants, SBR-9410572 and SES 99-86436. I thank Jonathan Baron, David Budescu, and an anonymous reviewer for helpful comments on the manuscript.

From my previous experience with studies of experts (Birnbaum & Hynan, 1986; Birnbaum & Stegner, 1981), I was aware of the difficulties and costs of recruiting and testing by postal mail. In such a survey, one must pay expenses of printing materials, addressing the envelopes, and postage for mailing the packets each way. It takes considerable time to address and send the materials and to wait for returns by mail. Return rates to mailed surveys are not particularly high, and efforts to increase return rates, such as reminder letters duplicate mailings or phone calls, are also time consuming and expensive. Once the paper questionnaires are returned, additional effort, time, and money must be spent to code, enter, and verify the data.

In contrast, Internet research does not require printing costs, mailing costs, or data coding and entry costs. I thought that it might be more convenient for my recruits to click a link in an E-mail message to an online study and to respond by pointing and clicking than to complete a paper questionnaire and mail it back. Although I had positive expectations for the method, I was not prepared for what happened when I sent an E-mail request for participation to the members of the *Society for Judgment and Decision Making* and *Society for Mathematical Psychology*.

Within a few days, there were completed data for more than a hundred participants, most holding doctorates, ready to analyze. Over the next few days, graduate students of decision making began to participate, then undergraduates in decision labs, followed by friends of these participants, and then friends of their friends. The Internet research proved so fruitful and so efficient compared with previous techniques, that I became interested in the methodological issues. I began a series of Web studies to check if results with this method would agree with those from the laboratory (Birnbaum, 2001).

Other investigators, also engaged in this type of research, were reaching similar conclusions. A consensus among the pioneers of Web-experimentation is that one can collect large samples of high quality data, completing in weeks what used to take months or a year to accomplish (see the survey in Musch & Reips, 2000 and the other chapters in Birnbaum, 2000b). There is a growing consensus that if the studies are properly done, Web and Lab research reach much the same conclusions (Krantz & Dalal, 2000).

Although data coding and data recording are automated by the HTML and CGI scripts of a Web study, which in theory should be free of errors, the programming of Web experiments can be tedious and vulnerable to mistakes. To make it easier to set up fool-proof studies, I constructed Web pages that create Web pages for running online studies of two popular types (Birnbaum, 2000c). Results obtained with judgment experiments constructed with these tools have been reported in a book on the general topic of Online Research (Birnbaum, 2001).

I have taught a laboratory course in which the undergraduates learned how to use the Internet to collect data. In my previous lab courses, most students do not have time to collect enough data to draw a conclusion, so their papers typically conclude that the results are too weak to draw any conclusions. With Web methods, however, my students are able in a couple of weeks to collect data from over a hundred participants in a very convenient fashion, so their papers have clear results and substantive discussion sections.

Because the participants, the medium, and experimental controls are different on the Web from those employed in the lab, it is important to establish whether results obtained in the new medium are different from those obtained with more traditional methods (Krantz & Dalal, 2000). To answer this question in my research domain, I began a series of related studies to re-examine phenomena observed in previous studies of decision making. Unless a new method reproduces classic results, any new finding might be suspected as an artifact.

The first two studies in this program (Internet A and B), which replicated and extended the Birnbaum and Navarrete (1998) results, have been published (Birnbaum, 1999c; 2000a). The main purpose of this chapter is to review the results of the next three studies in the series (C, D, and E), which I was privileged to present in my keynote address to the German Online Research Society Meeting in Nürnberg in October of 1999. These three studies replicate and extend several of the classic and modern findings in decision making in this new medium.

A second purpose of this chapter will be to discuss issues related to conducting research via the WWW. I have reached the conclusion that the marriage of Web-based research and the descriptive study of human decision-making is a good one, and this chapter will present reasoning and evidence in support of that conclusion. In addition, I will review major findings in the psychology of decision making, so that the chapter will provide an introduction to the descriptive study of risky decision making.

A Program of Decision Research on the Web

Risky decision making is one of the oldest problems in behavioral research. Bernoulli (1738/1954) posed the following question: Suppose a pauper was given a lottery ticket that would pay either 20,000 ducats or 0 ducats with equal probability. Would this person be willing to sell the ticket to a rich merchant for 9,000 ducats (less than its expected value of 10,000), and would both parties want to make this exchange? Bernoulli proposed the idea that the psychological value of money (utility) may not be a linear function of the objective, numerical measure. Bernoulli's theory, known in its modern form as Expected Utility (EU) theory, implied that it would be rational and beneficial for both to engage in this transaction.

Bernoulli's theory explained why people would buy and sell insurance, and it also explained the St. Petersburg paradox, the finding that people prefer a small amount of cash to certain gambles having infinite expected value. A number of important papers focused attention on these topics in the 1950s (Allais, 1953; Edwards, 1954; Savage, 1954; von Neumann and Morgenstern, 1947), and that interest has grown unabated to this day because human decision making contains puzzles that remain to be explained.

Despite early successes, EU theory has not been able to account for other seemingly paradoxical choices that people make (Allais, 1953; Edwards, 1954; Kahneman & Tversky, 1979). The Internet studies described below replicate many of the classic paradoxes and puzzles of behavioral decision making.

Methods of Internet Studies

In each of the studies (A through E), each person made twenty choices between gambles. The studies used different choices to test different properties. Recruits had two incentives: first, to help the study of decision making and second, to have the opportunity to win a cash prize. Participants in Internet A and B were told that each person would have a 1% chance to play one of their chosen gambles for real cash prizes.

Games were played and prizes awarded as promised. The expected value of participating in A and B, which took about five to ten minutes, was approximately equal to the expected value of a California State Lottery ticket (tickets cost $1 and have an expected value of $.50). Possible winnings ranged from $0 to $120. Of the first 1900 participants (Internet A and B), there were 19 winners, including 11 who won $90 or more. From each block of 100 participants in Internet A and B, one winner was selected randomly by the roll of two ten-sided dice. For each winner, a twenty-sided die was rolled to determine which of the twenty decisions would count. Then, two ten-sided dice were rolled to determine the prize of the participant's chosen gamble on that trial.

Because the number of participants might grow exponentially and perhaps bankrupt me, instructions in subsequent studies (Internet C, D, and E) specified that three participants would be selected randomly to win the chance to play one of their chosen gambles in each 3-month period. That change in the incentive structure reduced the uncertainty of the cost to the experimenter and increased uncertainty for the participants, because they could not know in advance how many participants there might be. In 1999 and 2000, no further advertising or recruitment was required to maintain rates of participation from 70 to 440 people per month. Most winners were paid by check or postal money order, but in a few cases of small prizes won by foreign participants, prizes were sent as cash in the mail.

The complete instructions and materials of the five studies reviewed here (now retired) can be viewed at the following URL: http://psych.fullerton.edu/mbirnbaum/archive.htm

Demographics of the Samples

Table 1 shows the demographic characteristics of the five samples. Compared to the lab sample (124 students), people recruited via the Internet tended to be older, better educated, more likely male (though the majority of all samples is female), and more likely to report having read a scientific article or book on decision making.

Table 1. Demographic Characteristics of Lab and Internet A-E Samples

Characteristic	Lab	A	B	C	D	E
Age <= 22 years	91	20	22	32	31	28
Age > 40 years	0	20	24	19	15	19
College graduate	0	60	47	41	45	38
Doctorate	0	11	3	3	3	3
Read on DM	13	31	19	20	20	18
Female	73	56	61	63	59	63
Sample size *(n)*	124	1224	737	860	753	1438

Note: each entry is the percentage in each group. "Read on DM" indicates that the participant checked "yes" to indicate having read a scientific article or book on decision making.

The Internet A sample was recruited from subscribers to E-mail lists for the *Society for Judgment and Decision Making* and the *Society for Mathematical Psychology*. Among Internet A's 1224 participants were 135 with doctorates, of whom 95 held doctoral degrees *and* reported having read a scientific work on decision-making. This was the "expert" group that I set out to reach in my first Internet study (Birnbaum, 1999c). Participants in Internet B, C, D, and E were recruited via search engines, links in Web sites listing psychology experiments, and links in pages listing contests and games with free prizes. Demographics of the more recent Web samples are intermediate between those of Lab and Internet A on all characteristics except the percentage with age > 40 in Internet B.

Instructions specified that each participant must be over 18 and may only participate once. Because each person specified an E-mail address (needed in order to contact winners), it was possible to utilize E-mail addresses to check for multiple submissions. Multiple submissions can also be detected by capturing the remote IP (Internet Protocol) address by means of the CGI script that saves the data.

By means of the sort and filter utilities in Excel, it is easy to sort on IP, E-mail address, and other identifying variables to detect multiple submissions. In an analysis of 1000 Internet data records, there were 5% blank or incomplete records (those with 75% or more answers were retained). A blank record is sent when a person views the page and clicks the "submit" button without having made any responses. There were also 2% with duplicate E-mail addresses, of which in only one case were the two records sent on different days. In that case, a woman completed the survey twice, two months apart, and interestingly agreed on 19 of 20 of her decisions. In all other cases, such duplicate data records appeared within minutes of each other. One person sent 8 copies of the same data within 2 minutes. Checking by IP address revealed only 12 additional cases in which the same IP appeared (besides those with the same E-mail address), but these data records came on different dates, had different demographics, and different decisions. These results are similar to those of Reips (2000), who found that elimination of records with duplicate E-mail and IP addresses would only eliminate a small percentage of the data. When testing participants in the lab, of course, one will naturally find duplicate IP addresses, since different participants use the same computers.

Based on such analyses, I believe that multiple submissions have been infrequent in my judgment and decision-making studies. They typically occur within a few minutes of the first submission. A few people, after reading the "thank you" message, apparently use the "Back" key to view the materials again, perhaps adding a comment, before using the "submit" button to go forward, which sends another copy. In these cases, extra copies are deleted from the data file before analysis. I think one should have a clear policy set in advance of what to do in cases of multiple submissions when the participant changes his or her responses. In my decision research, I always take the *last* copy sent and delete any previous ones. In the case of decision research, I want the person's final decision. In other studies, such as between-subjects studies in which the participants' contexts must be isolated (e.g., Birnbaum, 1999a), I take only the *first* record from each person, and delete subsequent submissions.

Risk Aversion Results

Table 2 shows results for 4 choices that were common to all of the studies. Choices 3 and 4 test consequence monotonicity, the principle that improving one consequence, holding everything else the same should make a gamble better. Violations might be considered measures of the ″quality″ of the data, if we assume that random or careless responding would be the only reason to violate this principle. (Some might intentionally choose less favorable gambles because they think that by not being ″greedy,″ they might be more likely to be selected as a winner). The percentage of violations varies from 4% to 8% in the Internet samples compared to 7% in the Lab, averaged over these two choices.

Table 2. Choices Used to Assess Risk Attitudes and Consequence Monotonicity in Lab and Internet Studies

Choice	Choice		Study					
No.	First Gamble	Second Gamble	Lab	A	B	C	D	E
1	*A*: .50 to win $0 .50 to win $100	*B*: .50 to win $25 .50 to win $35	58	48	52	55	53	53
2	*C*: .50 to win $0 .50 to win $100	*D*: .50 to win $35 .50 to win $45	69	60	63	62	63	62
3	*E*: .50 to win $50 .30 to win $96 .20 to win $100	*F*: .50 to win $50 .30 to win $62 .20 to win $100	8	6	9	11	10	11
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	94	96	97	94	94	94

Note: Choice percentage indicates the percentage in each sample who chose the second gamble, shown on the right in the table.

Choice 2 in Table 2 illustrates a phenomenon that Bernoulli sought to explain:

C: .50 to win $0	*D*: .50 to win $35
.50 to win $100	.50 to win $45

Gamble *C* offers a 50-50 chance to win either $0 or $100. It has an expected value of $50. Gamble *D* offers a 50-50 chance to win either $35 or $45; its expected value is only $40. However, most people prefer *D* to *C*. The percentages in the last six columns of Table 2 show the percentages of people in each study who preferred the gamble printed on the right in Table 2. For this choice (*D* versus *C*), from 60 to 69% of participants chose *D* over *C*, even though *C* has the higher expected value. For the Lab sample (n = 124), a 95% confidence interval on a binomial with mean of 50% ranges from 41% to 59%; for the smallest Internet sample (n = 737), the 95% confidence interval is 46.3% to 53.7%. Therefore, a significant majority prefers *D* to *C* in all samples.

Presumably, many people choose *D* because it offers a "sure thing" to win at least $35. When people prefer sure things to gambles that have the same or lower expected values, their behavior is termed *risk averse*.

The degree of aversion was found correlated with experience reading about decision making (Birnbaum, 1999c). Of the 392 who said they had read such works, only 52% chose *D* over *C*. Of the 837 who said they had not, 64% chose *D*. Of the 95 in the "expert" group (those holding doctorates and who have read on decision making), only 49 (52%) preferred *D* over *C*.

Evidence of risk aversion was also observed with the following choice (in Lab and Internet A):

E: $96 for sure	*F*: .01 to win $0
	.99 to win $100

In this case, 32% of the lab sample and only 26% of the Internet A sample chose Gamble *F* with the higher expected value (EV = $99) over *E*, $96 for sure. Only 30 of the 95 "experts" chose the gamble with higher EV in this case (32%), about the same as the percentage found in the Lab sample. This is not a choice where the experts would have had difficulty computing and conforming to EV, had they so intended. Preference for Gamble *D* over *C* and for *E* over *F* are examples of risk-averse behavior that can be explained by EU theory, as shown in the next section.

Expected Utility Theory

Expected utility (EU) can be written as follows:

$$EU(G) = \sum_{i=1}^{n} p_i u(x_i) \tag{1}$$

where there are *n* discrete, mutually exclusive and exhaustive outcomes of Gamble *G* = $(x_1, p_1, ...x_i, p_i; ...; x_n, p_n)$, with consequences and probabilities of x_i and p_i, respectively, and $u(x)$ denotes the utility of monetary consequence *x*. Bernoulli noted that if $u(x)$ is a negatively accelerated function, such as the logarithmic or square root function, then a person who chooses according to expected utility would exhibit risk aversion.

For example, suppose $u(x) = \sqrt{x}$. From Equation 1, the EU of Gamble *C* is 5 and the EU of Gamble *D* is 6.31. Therefore, EU theory with a square root utility function implies that people will prefer Gamble *D* to Gamble *C*. Even though *C* has the higher expected value, *D* has the higher Expected Utility.

Sometimes it is convenient to compare predicted values of gambles by their certainty equivalent (CE) values. The CE of a gamble is the amount of cash that would be equal in utility to the gamble. Suppose an amount of cash, *y*, is indifferent to Gamble *G*; then *y* is the CE of Gamble *G* [y = CE(*G*)]; we assume this means that $u(y) = EU(G)$; therefore, $y = u^{-1}[EU(G)]$. According to this (EU) model, CE(*C*) = \$25 < CE(*D*) = \$39.84.

Theories of decision making attempt to answer either of the following questions: "What *should* an idealized person do when confronted with a choice?" or "What *do* real people do when confronted with a choice?" Early literature suggested that the same theory might address both questions; however, phenomena were soon discovered that led most theorists to seek different theories for normative (what an idealized person should do) and descriptive (what the typical person does do) purposes.

In its original form as a theory that might be both normative and descriptive, EU theory treated utility as a function of final wealth states. However, it was soon realized that people's choices are not consistent with this treatment, as illustrated in the next section. Instead, it was found that choices involving the same "bottom lines" would result in different decisions, depending on how the alternatives were *framed,* or described, in terms making them seem like "losses" or "gains."

Framing Effects

Internet C was designed to replicate the type of framing effects observed in previous studies (see the review by Levin, Schneider, & Gaeth, 1998). Would Internet participants, choosing with chances of real cash consequences, make the same decisions when the same choices are described differently? For example, consider Choices 16 and 19 in Table 3:
16. Which do you choose?

e: .50 to win $0 .50 to win $100	*f:* $40 for sure

Consistent with the usual finding of risk aversion, 71% of the 860 participants in Internet C chose the sure $40 over the gamble with an EV of $50 on Choice 16.

Choice 19 of Table 3 is actually the same as Choice 16, except it is framed in terms of "loses:"
19. You get $100 and you must choose:

k: .50 to lose $100 .50 to lose $0	*l :* lose $60 for sure

Clearly, if you get $100 and you choose to lose $60 for sure, then you keep $40 for sure. In Gamble *k*, one receives either $100 or $0 with the same probabilities as in *e* of Choice 16. Although Choice 19 is the same as Choice 16 in terms of the final wealth states (bottom-lines), only 24.1% chose *l* over *k*. Of the 849 participants who responded to both questions, 429 (50.5%) chose *f* over *e and k* over *l* against only 27 (3.2%) who had the opposite reversal of preferences, $z = 18.8$. Thus, a simple change in wording causes the majority choice to switch from one side to the other.

When facing a choice between a sure gain and a fifty-fifty gamble to win, the majority appears risk averse. However, when choosing between a sure "loss" and a fifty-fifty gamble risking a "loss," the majority becomes risk seeking, now preferring the gamble to the sure "loss." Quotation marks are used here to denote so-called "losses" that are really gains, since in both situations the participants cannot lose anything and can only win or break even. Apparently, people try to avoid the sure "loss" of $60 (which guarantees them $40 to keep) even though they liked the sure "gain" of $40.

The other data in Table 3 show that similar framing effects are found for other pairs of choices constructed in the same manner. The only case in Table 3 where the majority is not reversed is for Choices 13 and 17, where the shift is still significant and in the same direction. Choice 17 resembles

Table 3. Choices Used to Assess Framing Effects and Risk Aversion (Internet C)

Choice #	Choice First Gamble	Second Gamble	% Choice
7	*M*: .95 to win $0 .05 to win $100	*N*: $5 for sure	63.0
20	*You get $100 and you must choose:* *m*: .95 to lose $100 .05 to lose $0	*n*: lose $95 for sure	20.0
13	*Y*: .05 to win $0 .95 to win $100	*Z*: $95 for sure	84.9
17	*You get $100 and you must choose:* *g*: .05 to lose $100 .95 to lose $0	*h*: lose $5 for sure	60.2
16	e: .50 to win $0 .50 to win $100	*f*: $40 for sure	71.0
19	*You get $100 and you must choose:* *k*: .50 to lose $100 .50 to lose $0	*l*: lose $60 for sure	24.1
10	*S*: .50 to win $0 .50 to win $100	*T*: $50 for sure	82.1
18	You get $100 and you must choose: *i*: .50 to lose $100 .50 to lose $0	*j*: lose $50 for sure	40.5

Note: % Choice indicates the percentage (of $n = 860$) who chose the second gamble; values less than 46.6 or greater than 53.4 are significantly different from 50%.

the wording of insurance, where a sure "loss" is paid to avoid the risk of a larger one.

Early theorists realized that such reversals of preference due to framing might be explained by EU theory if utility was defined on *changes* in wealth rather than on the bottom lines. Presumably, choices are made on the basis of *changes from the status quo* specified in the problem. The idea that utility should be defined on changes rather than wealth was proposed by

Markowitz (1952) and incorporated in descriptive, psychological theory by Edwards (1954) and Kahneman and Tversky (1979), among others.

Although EU theory could explain why people might show risk aversion, why people might buy and sell gambles, and why people might buy and sell insurance, it required some maneuvering to account for framing effects and other effects showing that the same person might both gamble *and* buy insurance. But theoreticians could not find a way to reconcile EU with the paradoxes of Allais (1953; 1979), which imply contradictions between EU theory and decisions that people make.

The Paradoxes of Allais

In the original Allais paradoxes, choices were posed in terms of large, hypothetical prizes involving millions of Francs. This led some investigators to wonder if these violations of EU theory were only found with extreme, hypothetical amounts and perhaps not with decisions for real, everyday consequences. Indeed, several of the highly educated participants in Internet A sent me E-mail messages stating that with the small consequences used in my studies, they anticipated that my results would conform to EV. One even suggested that a person would have to be "insane" not to use EV (a special case of EU) when consequences were of the magnitude used in my Internet studies. Interestingly, not even one person of the 1224 in Internet A actually conformed exactly to EV (Birnbaum, 1999c).

Studies D and E tested if Allais paradoxes are found with Internet participants who have small chances at real and modest cash prizes.

The Allais paradoxes can be viewed as violations of independence properties that are implied by EU theory. The independence properties listed in Table 4 have been tested in a number of investigations, reviewed by Camerer (1992), Luce (2000), Quiggin (1993), Starmer and Sugden (1989), and Wu and Gonzalez (1998), among others. As will be shown later, these properties can be usefully decomposed into simpler and perhaps more basic principles.

Consider Choices 8 and 9 of Table 5, which tested ratio independence in Internet D:

O: .99 to win $0 .01 to win $100	*P:* .98 to win $0 .02 to win $50
Q: .50 to win $0 .50 to win $100	*R:* $50 for sure

Table 4. Implications of Expected Utility Theory Tested in the Allais Paradoxes

Property Name	Expression
Ratio Independence	$G = (x, p; z, 1 - p) \succ F = (y, q; z, 1 - q) \Leftrightarrow$ $G' = (x, ap; z, 1 - ap) \succ F' = (y, aq; z, 1 - aq)$
Common Consequence Independence (Type 1)	$R = (y, p; z, 1 - p) \succ S = (x, q; z, 1 - q) \Leftrightarrow$ $R' = (y, p; x, r; z, 1 - p - r) \succ S' = (x, q + r; z, 1 - q - r)$
Common Consequence Independence (Type 2)	$R = (y, p; x, r; z, 1 - p - r) \succ S = (x, q; z, 1 - q) \Leftrightarrow$ $R' = (y, p + s; x, r - s; z, 1 - p - r) \succ S' = (y, s; x, q - s; z, 1 - q)$
Common Consequence Independence (Type 3)	$R = (y, p; z, 1 - p) \succ S = (x, q; z, 1 - q) \Leftrightarrow$ $R' = (y, p+ r; z, 1 - p - r) \succ S' = (y, r; x, q; z, 1 - q - r)$

Notation: G = (x, p; z, 1 – p) represents a gamble to win x with probability p and z otherwise. A = (y, p; x, q; z, 1 – p + q) represents a gamble to win y with probability p, x with probability q, and z with probability 1 – p + q. $A \succ B$ means A is preferred to B; $0 \le z < x < y$.

When choosing between *O* and *P*, most participants (64.3%) chose *O*. According to EU, $O \succ P$ (*O* is preferred to *P*) if and only if EU(*O*) > EU(*P*), which is equivalent to

$$.01u(100) + .99u(0) > .02u(50) + .98u(0) \tag{2}$$

or $$.01u(100) - .02u(50) > -.01u(0)$$

However, in Choice 9 of Table 5, most people (89.5%) chose *R* over *Q*: According to EU, $Q \prec R$ if and only if

$$.5u(100) + .5u(0) < 1u(50), \tag{3}$$

which implies $.5u(100) - 1u(50) < -.5u(0)$; dividing both sides of the inequality by the common factor 50, this implication contradicts the implication of Equation 2. There were 426 people (57% of the 743 who completed both choices) who chose *both* $O \succ P$ *and* $Q \prec R$, compared with only 20 (2.7%) who showed the opposite reversal of preferences. This

Table 5. Tests of Ratio Independence (Internet D)

Choice #	*a*	First Gamble	Second Gamble	% Choice
		Choice		
8		*O*: .99 to win $0	*P*: .98 to win $0	34.9
	1	.01 to win $100	.02 to win $50	
5		*I*: .90 to win $0	*J*: .80 to win $0	53.1
	10	.10 to win $100	.20 to win $50	
9		*Q*: .50 to win $0	*R*: $50 for sure	89.5
	50	.50 to win $100		

Note: Choice percentages show the percentage of choices (of *n* = 753) for the second gamble, shown on the right. Each choice is created from the first by multiplying both probabilities to win the higher prizes by common factors of *a* = 1, 10, or 50.

Table 6. Tests of Ratio Independence (Internet E)

Choice #	*a*	Choice		%Choice
		First Gamble	Second Gamble	
17		g: .98 to win $0	*h*: .99 to win $0	58.3
	1	.02 to win $48	.01 to win $96	
5		*I*: .80 to win $0	*J*: .90 to win $0	44.2
	10	.20 to win $48	.10 to win $96	
20		m: .20 to win $0	*n*: .60 to win $0	27.5
	40	.80 to win $48	.40 to win $96	
11		*U*: .02 to win $0	*V*: .51 to win $0	19.1
	49	.98 to win $48	.49 to win $96	

Note: Choice percentages show the percentage of choices (*n* = 1438) favoring the second gamble. Each choice is constructed from the first by multiplying both probabilities to win by *a* = 1, 10, 40, or 49.

paradox is termed the "constant ratio" paradox because the probabilities to win have been multiplied by the same common factor, yet the preferences change.

Table 6 shows that the constant ratio paradox occurs not only in choices involving "sure things" as in Choices 8 and 9 of Internet D (Table 5), but also as the probabilities to win are merely increased, as shown in Choices 17, 20, and 11 of Internet E. The difference in choice percentages is statistically significant between each pair of successive rows in both Tables 5 and 6 (as tested by binomial tests of correlated proportions). Thus, these data replicate and extend previous findings that the common ratio paradox cannot be attributed strictly to a *certainty* effect, a term coined to describe the original forms used by Allais (Kahneman & Tversky, 1979).

Allais (1953) also devised a "constant consequence" paradox, designed to test Savage's "sure thing" axiom (Savage, 1954). Wu and Gonzalez (1996; 1998) reviewed the literature for constant consequence paradoxes and presented new data for three distinct versions listed in Table 4. The original classic version of Allais (1953) was a Type 1. The variation labeled Type 1 in Table 4 is illustrated by Choices 7, 18, and 12 of Internet D (see Table 7). In this manipulation, a common branch (a probability-consequence pair that is the same in two gambles) with the lowest consequence has its consequence increased until it matches (and is combined with) the medium consequence in the pair. For example, consider Choice 7 in Table 7:

M: .90 to win \$0   *N*: .80 to win \$0
.10 to win \$100   .20 to win \$65

Table 7. Tests of Constant Consequence Independence Type 1 (Internet D)

Choice No.	Branch	Choice: First Gamble	Choice: Second Gamble	% Choice
7		*M*: .90 to win \$0 .10 to win \$100	*N*: .80 to win \$0 .20 to win \$65	62.2
18	.3 win \$65	*i*: .60 to win \$0 .30 to win \$65 .10 to win \$100	*j*: .50 to win \$0 .50 to win \$65	55.5
12	.8 win \$65	*W*: .10 to win \$0 .80 to win \$65 .10 to win \$100	*X*: \$65 for sure	71.0

Note: Choice percentage indicates the percentage (of n = 753) who chose the second gamble, shown on the right in the table.

Choice 18 is constructed by changing the consequence on the common branch of .3 to win $0 to a .3 probability to win $65 to both gambles as follows:

i: .60 to win $0	*j*: .50 to win $0
.30 to win $65	.50 to win $65
.10 to win $100	

Although 62% preferred *N* over *M* in Choice 7, only 55% preferred *j* over *i* in Choice 18 of Table 7. By "adding" another .5 to win $65 to both gambles, the gamble on the right becomes a "sure thing," and now 71% prefer the sure cash, *X,* to *W*. Each successive difference in choice percentages is statistically significant. Expected utility theory implies that there should be no effect of changing the consequence on a common branch (the common "consequence") in both gambles. The U-shaped change in choice probability, decreasing from Choices 7 to 18 and then increasing from 18 to 12 as a function of the probability shifted to the

Table 8. Tests of Common Consequence Independence Type 2 (Internet D)

Choice No.	branch	Choice: First Gamble	Choice: Second Gamble	% Choice
11		*U*: .20 to win $0 .70 to win $50 .10 to win $100	*V*: .10 to win $0 .90 to win $50	43.3
13	-.3 win $50 +.3 win $100	*Y*: .20 to win $0 .40 to win $50 .40 to win $100	*Z*: .10 to win $0 .60 to win $50 .30 to win $100	73.7
16	-.7 win $50 +.7 win $100	*e*: .20 to win $0 .80 to win $100	*f*: .10 to win $0 .20 to win $50 .70 to win $100	70.5
19	-.3 win $50 +.3 win $108	*k*: .20 to win $0 .40 to win $50 .10 to win $100 .30 to win $108	*l*: .10 to win $0 .60 to win $50 .30 to win $108	57.5

Note: Choice percentage indicates the percentage (of n = 753) in each sample who chose the second gamble, shown on the right in the table. Choices 13 and 19 test upper tail independence.

middle consequence, is consistent with the findings of Wu and Gonzalez (1996; 1998).

In Choices 11 and 13 of Table 8, probability is shifted from a middle consequence to a higher consequence, labeled Type 2 in Table 4. The difference in choice percentages between Choices 11 and 13 is statistically significant, $z = 12.1$, also in the direction previously observed in the literature (Wu & Gonzalez, 1998).

Choices 6 and 14 of Table 9 create the Type 3 (Table 4) common consequence condition, created by shifting the common branch from the lowest to the highest consequence (.8 to win \$0 is changed to .8 to win \$100 in both gambles). Of the 739 participants who expressed a choice on both trials, 313 switched from *K* to *b* against only 62 who switched from *L* to *a*, $z = 6.3$, producing a change from significantly more than half choosing *K* to significantly fewer than half choosing *a*.

In the common ratio paradox and in all three versions of the common consequence paradox (properties in Table 4), results obtained via Internet show the same patterns of violations of EU as those found in laboratory studies (Wu & Gonzalez, 1998). Contrary to speculations that results with real chances at small prizes would conform to EU if not EV theory, the results appear consistent with the idea that the same phenomena are being tested with small real prizes as with hypothetical, large prizes in the lab.

The Allais paradoxes caused many investigators to look for descriptive

Table 9. Tests of Common Consequence Independence Type 3 (Internet D)

Choice No.		Choice First Gamble	Second Gamble	% Choice
6		*K*: .90 to win \$0 .10 to win \$100	*L*: .80 to win \$0 .20 to win \$40	38.0
14	+.8 win \$100	*a*: .10 to win \$0 .90 to win \$100	*b*: .20 to win \$40 .80 to win \$100	72.1
10	Split .2 \$40 +.8 win \$110	*S*: .10 to win \$0 .10 to win \$100 .80 to win \$110	*T*: .10 to win \$40 .10 to win \$42 .80 to win \$110	53.5

Note: Choice percentage indicates the percentage (of $n = 753$) in each sample who chose the second gamble, shown on the right in the table. Choices 14 and 10 test upper cumulative independence.

models of behavior that could explain these choices. Investigators began to reach the conclusion that weights of events are affected not only by the events' probabilities, but also by the relative values of their consequences (Edwards, 1954; 1962; Birnbaum & Stegner, 1979; Kahneman & Tversky, 1979). For example, the weight of a .01 probability to win $30 depends on whether $30 is the worst possible consequence of the gamble or the best. There were several types of configural models considered (Birnbaum, 1999b), including one that has come to be called the "rank-dependent" approach.

Rank-Dependent Expected Utility Theory

Among the configural models, those of Rank Dependent Expected Utility (Quiggin, 1993), Cumulative Prospect theory (Tversky & Kahneman, 1992) and Rank- and Sign-Dependent Utility theory (Luce & Fishburn, 1991; 1995; Luce, 2000) use the same numerical representation for gambles having nonnegative consequences.

$$RDU(G) = \sum_{i=1}^{n} [W(\sum_{j=i}^{n} p_j) - W(\sum_{j=i+1}^{n} p_j)]u(x_i) \tag{4}$$

where *RDU(G)* is the rank dependent expected utility of Gamble *G*; the *n* mutually exclusive and exhaustive consequences of *G* have been ranked such that $0 < x_1 < x_2 < x_3 < \ldots < x_n$; *W(P)* is a strictly increasing monotonic weighting function that assigns decumulative weight to decumulative probability; $W(0) = 0$ and $W(1) = 1$.

This theory could account for risk aversion and the Allais paradoxes if the $W(P)$ function had the form of an inverse-S curve that was steeper near 0 and 1 than near 1/2 (Wu & Gonzalez, 1996; 1998). A function that has been fit to the $W(P)$ function is as follows:

$$W(P) = \frac{cP^{\gamma}}{cP^{\gamma} + (1-P)^{\gamma}} \tag{5}$$

where c (about .724) is an index of risk aversion for 50-50 gambles (in addition to that attributable to the $u(x)$ function), and γ (about .6) characterizes the inverse-S shape (with $\gamma < 1$). If $c = \gamma = 1$, $W(P) = P$, so the model reduces to EU.

My colleagues and I found that violations of branch independence were opposite of those implied by the inverse-S shaped weighting function that is required by RDU model to explain the Allais paradoxes (Birnbaum & McIntosh, 1996; Birnbaum & Chavez, 1997; Birnbaum, 1999b; 1999c).

Birnbaum and McIntosh (1996) noted that there are two possible interpretations of this result — (1) perhaps the weighting functions observed in different studies with different participants and different choices are different, or (2) perhaps the RDU model is wrong. We noted that an alternative configural weight model could describe both the Allais paradoxes and the observed pattern of violations of branch independence without contradiction. I set out to see if I could concoct new paradoxes — choices that might provide a direct contradiction in RDU.

The RDU model requires that choices satisfy the properties of transitivity, monotonicity, comonotonic branch independence, and coalescing (Table 10). *Transitivity* implies that if you prefer A to B and B to

Table 10. Testable Properties of Decision Making Theories

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

Notation: 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 where consequences maintain the same ranks. In tests of (noncomonotonic) branch independence and cumulative independence, $0 < z < x' < x < y < y' < z'$

C, then you prefer A to C. *Monotonicity* means that if one consequence is improved in a gamble, holding everything else the same, the gamble is improved. *Coalescing* assumes that if two branches in a gamble have the same consequence, one can combine them by adding their probabilities without changing the utility of the gamble. *Branch independence* is a weaker form of Savage's sure thing axiom (Savage, 1954) that does not presume coalescing. If two gambles share a common branch (the same probability event consequence combination), then the common consequence can be changed without affecting the preference between the gambles. RDU violates branch independence, but it satisfies it when the consequences are also comonotonic (i.e., when the common consequences, z and z' in Table 10 maintain the same ranks).

RDU must also satisfy properties that can be deduced from these premises, such as stochastic dominance, lower and upper cumulative independence, and event splitting (Birnbaum, 1997). The next section will present violations of RDU that are even more striking, perhaps, than the Allais paradoxes. These new tests were predicted to fail, based on predictions from another descriptive model of decision making.

Configural Weight Models

In a chapter on configural weight models, I noted that the models with which I was working had certain similarities to the rank-dependent models, but that they could be distinguished by experiment (Birnbaum, 1997). The configural weight models and the RDU model are similar in that they ascribe risk aversion or risk seeking to the assignment of additional weight to lower- or higher-valued items, rather than to the shape of the utility function. These classes of models differ in that the RDU models must satisfy coalescing, stochastic dominance, and lower and upper cumulative independence, whereas my configural weight models implied specific violations of these properties.

Lower and upper cumulative independence were derived as new tests that would soon be as troublesome to the RDU models as the Allais paradoxes were to EU. These properties were derived to characterize the contradiction in results between tests of restricted branch independence and the RDU account of the Allais common consequence paradoxes.

The Configurally Weighted Utility (CWU) of a gamble can be written as follows:

$$CWU(G) = \sum_{i=1}^{n} w(x_i, G)u(x_i) \tag{7}$$

where G = (x_1, p_1; x_2, p_2; ...; x_i, p_i; ...; 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. RDU, EU, and EV are special cases of Equation 7. The RAM and TAX models are also special cases of Equation 7.

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

$$w(x_i, G) = \frac{S(p_i) + \rho \sum_{j=1}^{i-1} S(p_i) - \rho \sum_{j=i+1}^{n} S(p_j)}{\sum_{j=1}^{n} S(p_j)} \tag{8}$$

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, (ρ<0). This branch 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.

Birnbaum and Chavez (1997) noted that if δ = –1, and γ = .6, then this TAX model makes very similar predictions to the RDU model of Tversky and Kahneman (1992) and the RAM model of Birnbaum and McIntosh (1996) for choices between two-outcome gambles and for the Allais paradoxes. However, the TAX model (with these same parameters) correctly predicts violations of branch independence and distribution independence, unlike the Tversky and Kahneman (1992) model. This model (and parameters) can also account for observed violations of stochastic dominance, cumulative independence, and event-splitting effects as well as the classic Allais paradoxes (Birnbaum, 1999b).

It is common in psychological research to fit models to the data and then state that "predictions" are consistent with the results to which they were fit. Thus, predictions in psychological research are often made, not in advance, but rather after the data have been collected. It is noteworthy that, in this case, the predictions were indeed calculated before the experiment was designed (Birnbaum & Navarrete, 1998).

To compute predictions of the TAX, RAM, and two variations of the cumulative prospect model of Tversky and Kahneman (1992), online calculators have been provided at the following URLs:
http://psych.fullerton.edu/mbirnbaum/taxcalculator.htm
http://psych.fullerton.edu/mbirnbaum/cwtcalculator.htm

These calculators can be used to explore the predictions of these decision-making models. They are useful in the design of experiments designed to test implications of the models.

Violations of Stochastic Dominance

Consider the following choice:

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

According to the RAM or TAX models, with parameters estimated from previous data, the certainty equivalent of Gamble *I* is less than the certainty equivalent of Gamble *J*, even though *I* dominates *J*.

Lab studies found that about 70% of undergraduates tested violate stochastic dominance on this type of choice (Birnbaum & Navarrete, 1998; Birnbaum, Patton, & Lott, 1999). My Internet A and B studies (Birnbaum, 1999c; 2000a) compared online samples with a lab sample tested with the same procedures. The online samples had lower rates of violation for this choice (58% and 63.5% in Internet A and B) than did the Lab sample (73%). However, the incidence of violations was still high and significantly greater than 50% in all three samples for this choice. We can reject the hypothesis that people are confused and just guessing in favor of the hypothesis that most people systematically choose the dominated gamble.

Internet A and B also provided tests of coalescing. Consider the following comparison:

U :	.05 to win \$12	*V* :	.05 to win \$12
	.05 to win \$14		.05 to win \$12
	.05 to win \$96		.05 to win \$90
	.85 to win \$96		.85 to win \$96

The choice between *U* and *V* is the same as that between *I* and *J*, except that the highest two consequences in *U* have been coalesced in *I* and the two lowest consequences in *V* have been coalesced in *J*. However, most participants prefer *J* to *I* and *U* to *V*. In Internet B, 64% of the 737 participants chose *J* over *I*, and 86% chose *U* over *V*, including significantly more than half of the participants who reversed preferences this way.

The Internet A study was intended to recruit highly educated people who should be able to understand the instructions and the concepts of probability used to describe the gambles. Because the typical "subject pool" consists of students in the same university and level, lab participants are homogeneous with respect to education. By recruiting and testing people online, it is possible to examine the generality of results across educational levels and other such characteristics.

Table 11 shows the relationship between violations of stochastic dominance, gender, and education in Internet A. Note that the smallest subsample size is 54, for females with doctorates. Let *G*+ refer to the dominant gamble (such as *I*) and *G*– refer to the dominated gamble (such as *J*); *GS*+ is the split version of *G*+ and *GS*– is the split version of *G*–. In each subsample, violations of stochastic dominance were significantly more frequent in the coalesced form (*G*+ versus *G*–) than in the split form (*GS*+ versus *GS*–), averaged over two choices counterbalanced for position. At the same time, violations are less frequent among males than females, and they are

Table 11. Tests of Stochastic Dominance and Monotonicity by Gender and Education (Internet A)

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

Notes: Education < 16 indicates less than bachelor's degree; 16= Bachelor's degree; 17-19 = Postgraduate studies; 20 = doctorate. Cell entries indicate percentage of violations of stochastic dominance and consequence monotonicity in coalesced and split forms, respectively, based on Choices 5 and 11 of Internet A sample. Corresponding lab sample values are shown in parentheses.

less frequent among those with doctorates than those without college degrees. In Internet B, recruited from people not trained in decision making, the effect of education was far less pronounced, and the incidence of violations of stochastic dominance in the coalesced form remained above 50% at all levels of education.

Internet C also included the following choice, which departs from the usual recipe used to create violations of stochastic dominance:

I^*:	.75 to win $2	J^*:	.70 to win $2
	.02 to win $95		.02 to win $3
	.03 to win $97		.03 to win $4
	.20 to win $98		.25 to win $98

In this case, the largest portion of probability is on the lowest outcome, unlike the case in the previous tests. In this case, 58.5% violated stochastic dominance, significantly more than half of the participants. This new finding shows that violations of stochastic dominance are not limited to the recipe used in previous studies with three consequences in which the highest consequence is most likely.

The properties of coalescing, stochastic dominance, and cumulative independence were shown to follow from the RDU representation (Birnbaum & Navarrete, 1998). Luce (1998; 2000) reviewed which decision theories satisfied or violated coalescing and showed that coalescing leads to RDU in the context of mild background assumptions. Luce noted that many other decision theories also imply coalescing. These results thus leave few descriptive decision theories standing.

Choices 13 and 19 in Table 8 create a test of upper tail independence, a property tested by Wu (1994). Upper tail independence is a combination of comonotonic branch independence, coalescing, and transitivity. RDU must satisfy the condition, but the TAX and RAM models with parameters from previous studies imply a preference reversal between these choices. Although the majority choice does not change, the shift is in the direction predicted by the TAX model and consistent with the effects previously reported by Wu (1994), $z = 7.1$.

Choices 14 and 10 in Table 9 provide a test of upper cumulative independence (Birnbaum & Navarrete, 1998). Gambles *S* and *T* have been created from *a* and *b* by changing a common branch of .8 to win $100 to .8 to win $110 on both sides. In addition, the lower branch of .2 to win $40 has been split, with one of the .1 branches increased to $42 in *T*. Note that the change from Choice 14 to Choice 10 makes the right side better (.1 at$40 has been increased to .1 at $42 from *b* to *T*). Therefore, *T* should be

Table 12. Tests of Event Splitting and Monotonicity (Internet E)

Choice No. Choice Type	First Gamble	Second Gamble	% Choice
13	*Y*: .80 to win \$0 .20 to win \$48	*Z*: .80 to win \$0 .10 to win \$2 .10 to win \$96	54.5
19	*k*: .80 to win \$0 .10 to win \$47 .10 to win \$48	*l*: .80 to win \$0 .10 to win \$2 .10 to win \$96	33.4

Note: Choice percentage indicates the percentage (n = 1438) who chose the second gamble, shown on the right in the table.

relatively better than *S*. If a person prefers *b* to *a*, that person should prefer *T* to *S*, according to RDU. The TAX and RAM models violate coalescing; in particular, they imply that splitting the lower branch of *T* makes it worse, despite the increase in value, because more relative weight is given to these lower outcomes. Of the 314 people who changed preferences, 228 switched from *b* to *S* against only 86 who switched from *a* to *T*, z = 8.01, a significant split in the direction predicted by the configural weight models.

Table 12 shows a test of event splitting and monotonicity from Internet E. According to RDU, if a person prefers *Z* to *Y*, then one should prefer *l* to *k*. In Table 12, note that the choice is exactly the same, except that the .20 branch to win \$48 has been split into .1 to win \$47 and .1 to win \$48. According to RDU, splitting has no effect, but reducing \$48 to \$47 should make *k* worse than *Y*; however, according to the RAM or TAX models, splitting the higher outcome of *Y* change makes *k* better than *Y*. Of the 526 who changed preferences, 411 switched from *Z* to *k* compared with only 115 who switched in the opposite direction, z = 12.91. The TAX model, with the median parameters estimated from Internet A (γ = .791 and δ = –.333), predicts a reversal of preference in this case, consistent with the majority choice.

Table 13 shows results from Internet E of another test of Upper Tail Independence, similar to a test by Wu (1994). According to RDU, a person should prefer $T \succ S$ if and only if $f \succ e$. This property rests on coalescing and comonotonic branch independence. According to the RAM and TAX models, splitting the .48 branch to win \$92 from *T* to *f* makes *f* relatively better. Of the 610 who switched preferences, 503 switched from *S* to *f*,

Table 13. Test of Upper Tail Independence (Internet E)

Choice No.	First Gamble	Second Gamble	% Choice
10	*S*: .50 to win \$0 .07 to win \$68 .43 to win \$92	*T*: .52 to win \$0 .48 to win \$92	34.0
16	e: .50 to win \$0 .07 to win \$68 .43 to win \$97	*f*: .52 to win \$0 .05 to win \$92 .43 to win \$97	62.0

Note: Choice percentage indicates the percentage (n = 1438) who chose the second gamble, shown on the right in the table.

against only 107 who switched in the other direction. This result is consistent with the prediction of the TAX model, z = 16.03, and is not predicted by RDU, which does not predict systematic changes in preference.

Model Comparisons

The RAM and TAX models make similar predictions to those of CPT for the values of two-outcome gambles and the Allais paradoxes. Both TAX and CPT models correctly predict the reversal of preference between Choices 8 and 9 in Table 5, thus accounting for the original Allais paradox. However, neither TAX nor CPT models, with parameters fit to previous data, predict correctly the shifts between Choices 8 and 5 in Table 5 or between Choices 17 and 5 of Table 6. The inverse S weighting function of CPT implies that the difference between the gambles should be a U-shaped function of the common factor, *a*. Instead, choice percentages decrease monotonically as *a* increases. If this monotonic decrease is observed within individuals and with other values of *a*, then it may constitute another phenomenon that needs explanation.

Both TAX and CPT can account for the three types of common consequence paradoxes in Tables 7, 8, and 9. Both predict reversals between 7 and 18 and between 18 and 12 in Table 7. Both predict the reversal between Choices 11 and 13 in Table 8, and both predict the reversal between Choices 6 and 14 in Table 9.

The TAX model predicts violations of stochastic dominance, coalescing, upper cumulative independence, and upper tail independence. The TAX model (based on median parameters estimated in Internet A) predicts a reversal of preference between Choices 13 and 19 in Table 8, violating upper tail independence in Internet E. This model predicts a reversal between 14 and 10 in Table 9, violating upper cumulative independence in Internet D. Although the percentages in Choices 19 and 10 do not fall below 50% as predicted by TAX with prior parameters, these shifts in choice percentages in Tables 8 and 9 are significant and in the direction predicted. The TAX model correctly predicts violations of stochastic dominance and reversals of preference due to event-splitting (as in Table 11) observed in Internet A and B. It also correctly predicts the reversals of preference in Table 12 produced by event splitting (and contrary to monotonicity) and it correctly predicts the violation of upper tail independence in Table 13 in Internet E. Thus, the configural weight, TAX model accounts for effects observed that violate any form of RDU model.

Discussion and Conclusions

The findings in these Internet studies fit with other research done in the lab and via the Web. The Internet studies reveal that the Allais paradoxes have greater generality across gambles and demographics than previously established. Analysis by education, gender, and experience shows that although there are some trends related to demographic characteristics, the same phenomena appear in people other than college students.

In addition, the Internet studies show that properties that can be deduced from coalescing are consistently violated. Violations of coalescing appear to be the explanation for event-splitting effects, violations of stochastic dominance, violations of lower and upper cumulative independence, and violations of upper tail independence. These violations are to the RDU models what the Allais paradoxes were to EU. There is no specification of functions or parameters than can account for these phenomena within RDU.

To create a violation of RDU, start with a choice between gambles and then split the lower valued outcome in one gamble and the higher outcome in the other. According to the configural weight, RAM and TAX models with parameters estimated in previous research, splitting a lower-valued branch decreases the value of a gamble and splitting a higher-valued branch increases the value of a gamble. According to the RDU models, such splitting should have no effect on the choice.

The consistency of laboratory and Internet findings suggests by induction that the Internet can be trusted to reach conclusions that can be replicated. Because the links to my decision-making study are already in place and because new participants continue to visit the site, I have found it convenient to introduce a new study every three or four months to test new variations. It has become possible to collect more data with more convenience and less expense than ever before.

Because Internet samples are so heterogeneous, one might think that they have larger error, which would be a drawback in lab research with small sample sizes. But the Internet samples are also quite large, large enough that the data can be partitioned by demographic variables such as gender, education, and age to empirically check if the phenomenon under investigation varies with these variables (Birnbaum, 1999c). At the same time, such partitioning reduces random error within each sub-sample. Within each of these sub-samples, there are enough participants that one can have sufficient data to reach clear conclusions on each test. That means that the Internet data can be used to assess the generality of lab findings. So, what might at first seem a drawback can be seen as an advantage of Internet research.

For example, consider that 58.3% of the complete Internet E sample chose the riskier gamble on Choice 17 of Internet E. Because this percentage might seem a narrow majority, one might ask, does the majority choice differ in different subgroups? Suppose data are divided by gender, nationality, experience reading on decision making and education: are the same results found in each sub-sample? Of the 1438 participants in Internet E, there were 905 females, including 669 females born in the United States who had *not* read on decision making. These can be further broken down by level of education, with the finding that within each level of education, the majority made the same choice. The same choice was also in the majority for the 130 American females who *had* read on decision making at each level of education. The same was true of males at each level of education, whether they had or had not read on decision making. In this case, the modal choice was the same within each of these finely partitioned sub-groups. This empirical demonstration of consistency of conclusions across groups is reassuring to those of us who try to model the "typical" person's behavior.

The Method section of a published paper includes only those features thought important to replicate a work; because space in the journals is at a premium, it happens that details of experimental procedure are necessarily abbreviated or omitted. That means that different investigators may attempt to "replicate" with procedures that may differ in dozens of features from one lab to the next, causing differences in results, when they occur, to be difficult to pin down. The use of Internet research makes it easy to communicate the exact procedures used in a study to fellow scientists, who can view the complete instructions and materials via the WWW.

The methods of Internet research lend themselves to easy replication and variation. One can thus test effects of instructions, stimulus displays, or other procedures deemed important.

Once a study is online where colleagues can see and criticize it, one can get helpful advice and criticism more quickly than with other methods. In the past, it was often several months or even years after the completion of a study that one would first present one's research for review. In contrast, Online Research is available for review and criticism much more quickly — I have even received advice and opinion while collecting data. When reporting my results at conferences and colloquia, I have received suggestions of method that colleagues conjecture might alter or even reverse the results.

For example, Sandra Schneider (personal communication, November, 1999) suggested that people might not violate stochastic dominance if probabilities in the gambles were presented graphically, rather than only numerically. Perhaps with a graphic display, people might be able to "see" the dominance relation between *G*+ and *G*–, as they apparently can between *GS*+ and *GS*–. Cristof Tatka (personal communication, July, 2000) suggested that if the positions of the displayed consequences were reversed, effects attributed to rank of the consequences might be reversed, since rank and position had been confounded in previous tests.

Within a few weeks of each suggestion, it was possible to have the appropriate experiments online. Each study was set up as a between-subjects experiment, with different methods of representing the gambles. Participants were randomly assigned to different conditions by clicking on their birth months, and the association of birth month to conditions was counterbalanced over the run of experiment.

Choices that were common to Internet A and B were presented to new visitors to the site, who received different displays of the choices. Consequences were displayed in the original order or reversed order; probabilities were presented as text, or by means of pie charts. The pie chart was described as a spinner mechanism, where any final position of the spinner was equally likely.

Once these studies were online, it was possible to have clean answers to each question within weeks. Violations of stochastic dominance in the text version (n = 172) produced 59.6% violations in the coalesced version against 9.8% violations in the split form, averaged over two choices. These results are quite close to the corresponding figures of 58.8% and 7.9% from Internet B, averaged over the same two choices. When the consequences are presented in reverse order, from highest to lowest (n = 169), the corresponding figures are 62.4% and 10.7%. With two versions of the pie chart format (n = 353 and 305), the percentages were 65.2% and 64.4% in the coalesced form, against 8.5% and 8.2% in the split form, respectively. These studies illustrate how easy it is to test ideas about the source of the violations that might otherwise remain under consideration for years. In this case, they also show that these particular procedural manipulations do not suffice to significantly reduce the violations.

One of the ideas concerning procedure that has proved quite resistant to extinction (despite many empirical tests) is the notion that changes in financial incentives might make violations of EU go away. In response to the original Allais paradoxes, some argued that the consequences were too large to produce behavior that would conform to EU theory. In reaction to my experiments, it was suggested that perhaps my consequences were too small. One academic even theorized that my participants calculated the values of all of the gambles, calculated the differences, and decided that value of making the calculations was so small that they then decided intentionally not to use their calculations, which somehow led them to systematically violate stochastic dominance. This line of argument, however, does not account for the difference in behavior between the coalesced and split forms of the same choices. The financial incentives were the same in both forms of the choice, but the behavior is quite different.

Camerer and Hogarth (1999) reviewed the literature on financial incentives, concluding that financial incentives alone have not been shown to have drastic effects on choices between gambles. From many comments I received from participants in my Internet studies and in studies with students, I believe that the participants are motivated and even enthusiastic about the potential for cash prizes in my studies. I suspect that the problem lies not with incentives or motivation, but with understanding.

The small, but systematic correlation between education and violations of stochastic dominance observed in Internet A is consistent with the idea that people with more education can see the dominance relationships better than others. Of course, a correlation with education might be due to some other unmeasured factor (e.g., intelligence or wealth) that is correlated with education. One could establish the causal effects of a particular training program by an experiment in which people are randomly assigned to conditions with different forms of training. It seems plausible that with enough experience or the right training, people might learn to avoid violations of stochastic dominance. It would be interesting to see if such training would generalize to produce consistency with other principles of decision making.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine [Rational human behavior when facing risk: A criticism of the postulates and axioms of the American School]. *Econometrica, 21*, 503-546.

Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School. In M. Allais & O. Hagen (Eds.), *Expected utility hypothesis and the Allais paradox* (pp. 27-145). Dordrecht, The Netherlands: Reidel.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropoliannae, 5*, 175-192.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22*, 23-36. (Translation of 1738 paper by L. Sommer).

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

Birnbaum, M. H. (1999a). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*(3), 243-249.

Birnbaum, M. H. (1999b). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B. A. Mellers, & D. A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27-52). Norwell, MA: Kluwer Academic Publishers.

Birnbaum, M. H. (1999c). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399-407.

Birnbaum, M. H. (2000a). Decision making in the lab and on the Web. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 3-34). San Diego, CA: Academic Press.

Birnbaum, M. H. (Ed.). (2000b). *Psychological experiments on the Internet*. San Diego: Academic Press.

Birnbaum, M. H. (2000c). SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet. *Behavior Research Methods, Instruments, and Computers, 32*(2), 339-346.

Birnbaum, M. H. (2001). *Introduction to Behavioral Research on the Internet*. Upper Saddle River, NJ: Prentice Hall.

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., & Hynan, L. G. (1986). Judgments of salary bias and test bias from statistical evidence. *Organizational Behavior and Human Decision Processes, 37*, 266-278.

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.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48-74.

Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of heredity and environment. *Journal of Experimental Social Psychology, 17*, 159-182.

Camerer, C. F. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 207-251). Boston: Kluwer Academic Publishers.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production theory. *Journal of Risk and Uncertainty, 19,* 7-42.

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

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.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Eds.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: Academic Press.

Luce, R. D. (1998). Coalescing, event commutativity, and theories of utility. *Journal of Risk and Uncertainty, 16*, 87-113.

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

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.

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

Musch, J., & Reips, U.-D. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-87). San Diego, CA: Academic Press.

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

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

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). San Diego, CA: Academic Press.

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

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

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

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

As far as automated decision-making is concerned, the Commission concludes in recital 25 of its decision on the adequacy of the Privacy Shield (â€œthe adequacy decisionâ€) that in areas where companies most likely resort to the automated processing of personal data to take decisions affecting the individual (e.g. credit lending, mortgage offers, employment), U.S. law, notably the Equal.Â The method of research included primary and secondary sources, as well as interviews with experts representing various stakeholder groups and potential domains of ADM application. Additional interviews have been conducted as part of the legal analysis to ensure that both industry and regulator views are included for each area analysed. I started in Web-based research because I wanted to test a specialized sample of highly educated people to check results found with college students. I previously published some interesting predictions of configural-weight models that had not yet been tested (Birnbaum, 1997). These configural models, fit to previous data, implied that people would violate stochastic dominance in choosing between specially constructed gambles.Â @inproceedings{Birnbaum2001AWP, title={A Web-Based Program of Research on Decision Making}, author={Michael H. Birnbaum}, year={2001} }. Michael H. Birnbaum. I started in Web-based research because I wanted to test a specialized sample of highly educated people to check results found with college students.