Available online at www.sciencedirect.com
ScienceDirect
Organizational Behavior and Human Decision Processes 105 (2008) 122–133
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Dimension integration: Testing models without trade-offs ☆

Michael H. Birnbaum *, Adam R. LaCroix

Department of Psychology, California State University—Fullerton H-830M, P.O. Box 6846, Fullerton, CA 92834-6846, USA

Received 6 December 2006
Available online 26 October 2007
Accepted by Robyn Dawes

Abstract

This paper tests a behavioral property called dimension integration. The test evaluates models, such as lexicographic semi-orders and the priority heuristic, which assume that a person uses only one dimension at a time. It provides a way to compare such models against those that assume a person combines information from different dimensions. The test allows one to test the hypothesis that different people use different lexicographic semi-orders with different threshold parameters. In addition, by use of a "true and error" model, it is possible to "correct" for unreliability of choice in order to estimate the proportions of participants who show different response patterns that can be classified as integrative or not integrative. An experiment with 260 participants was conducted in which people made choices between two-branch gambles. The aggregate results violate the priority heuristic and six lexicographic semi-orders. The data also refute the theory that people use a mixture of these lexicographic semi-orders. In addition, few individuals appear to show response patterns consistent with non-integrative models. Instead, they show that most individuals show patterns consistent with the hypothesis that they combine information between dimensions.

Keywords: Choice; Decision making; Lexicographic semi-order; Priority heuristic; Risk; Utility

Theories intended to describe decision making under risk and uncertainty can be divided into those that postulate that people integrate information from different dimensions and those that assume that people only use one dimension at a time in making a choice. The family of integrative models includes expected utility theory (EU), cumulative prospect theory (CPT), transfer of attention exchange (TAX), gains decomposition utility (GDU), and many others (Birnbaum, 2005a, 2005b, 2005c; Luce, 2000; Luce & Marley, 2005; Marley & Luce, 2001, 2005; Tversky & Kahneman, 1992). Non-integrative models include lexicographic semi-orders, the priority heuristic, and single-dimension heuristics (Brandstaetter, Gigerenzer, & Hertwig, 2006; Tversky, 1969). Although the stochastic difference model (González-Vallejo, 2002) and other additive difference models have some similarities to the LS models and to the priority heuristic (they can violate transitivity), the stochastic difference model and additive difference models assume dimension integration.

Brandstaetter et al. (2006) reviewed a number of studies of decision making and argued that their priority heuristic provides a superior description of previously published decision making data than do integrative models. The data that were analyzed by Brandstaetter et al. (2006) were drawn mostly from studies that were designed to test between integrative models. None of the studies that they analyzed were designed to test the priority heuristic, so it would seem useful to test implications of the priority heuristic to see if it is indeed an accurate descriptive model.

☆ Support was received from National Science Foundation Grants BCS-0129453 and SES-0202448. We thank Jeffrey P. Bahra, Christopher Barrett, Roman Gutierrez, Phu Luu, and Cari Welsh, who assisted with collection of data and Eduard Brandstätter for comments on a previous draft and discussions of these issues.

* Corresponding author. Fax: +1 714 278 7134.
E-mail address: mbirnbaum@fullerton.edu (M.H. Birnbaum).

0749-5978/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2007.07.002

This paper employs a test of dimension integration that to our knowledge been used only once before (Birnbaum, submitted); this test allows one to determine whether people combine information and make trade-offs or if instead, they use only one piece of information at a time. Dimension integration provides a direct test between the family of non-integrative models and the integrative models.

The test of dimension integration allows us to test a more general family of theories as well as the priority heuristic. It allows for the possibilities that different people might use different lexicographic semi-orders, in which they examine dimensions in different orders, and that they might use different parameters. Tests of dimension integration also allow us to investigate whether people might use a mixture of lexicographic strategies with different parameters.

Brandstaetter et al. (2006) considered the possibility of such extensions, but they restricted their attention to a single order and fixed the parameters of their theory. The tests of integration in this paper allow one to test a much wider family of models than was evaluated by Brandstaetter et al. (2006). In addition, whereas Brandstaetter et al. (2006) proposed to describe only aggregate data, we can assess individual differences and estimate the percentage of individual participants who show different patterns of behavior that are compatible with or in violation of the family of non-integrative models.

Lexicographic semi-orders

Luce (1956) proposed a semi-order representation to describe preference behavior in which items that differ by small increments in utility are treated as indifferent. In such a representation, the indifference relation is not transitive. That is, *A* might be indifferent to *B*, *B* might be indifferent to *C*, and yet *A* might be preferred to *C*. For example, imagine a series of gold coins in which each adjacent pair of coins differs by the weight of one atom of gold. Because the weight of one atom is less than the resolution of most scales, people would evaluate any two adjacent coins in the series as equivalent. However, for some integer, *n*, the difference in weight between the first coin and the *n*th would be noticeable.

A lexicographic order is illustrated by the task of putting a list of words in alphabetic order. The first letter is checked and if that letter is different, the two words are ranked based on that letter alone (and subsequent letters have no effect). However, when the first letter is the same in two words, one checks the second letter; and only if the second letters are also identical is there a need to go on to check the third, and so on.

Tversky (1969) noted that preference can be intransitive in a lexicographic semi-order (LS). In a lexicographic semi-order, a person compares one dimension at a time and makes a decision based on that dimension only. Only when the difference in the first dimension is small does the person check the second dimension; the person examines the third dimension only when differences on the first two dimensions are not decisive.

When comparing two-branch gambles of the form, $G = (x,\ p;\ y,\ 1-p)$, which represents a gamble with a probability of p to win cash prize x and otherwise to win y, and $F = (x',\ q;\ y',\ 1-q)$, where $x > y \geqslant 0$ and $x' > y' \geqslant 0$, there are three dimensions that could be examined: the lowest consequence (L), the probability to win (P), and the highest consequence (H). Suppose a person compares first the lowest consequences in the gambles, then the probabilities, and finally the highest consequences in a gamble. That strategy, defined more precisely below, will be denoted the low–probability–high lexicographic semi-order (*LPH* LS), as follows:

$$\begin{aligned}
&\text{If}(y - y' \geqslant \Delta_L)\{\text{choose } G\}\\
&\text{else if}(y' - y \geqslant \Delta_L)\{\text{choose } F\}\\
&\text{else if}(p - q \geqslant \Delta_P)\{\text{choose } G\}\\
&\text{else if}(q - p \geqslant \Delta_P)\{\text{choose } F\}\\
&\text{else if}(x - x' > 0)\{\text{choose } G\}\\
&\text{else if}(x' - x > 0)\{\text{choose } F\}\\
&\text{else}\{\text{choose randomly}\}
\end{aligned} \tag{1}$$

There are two parameters, Δ_L and Δ_P, which represent the difference thresholds for the lowest prize and probability, respectively. When gambles involve three or more branches, new parameters can be introduced for the thresholds on those additional dimensions.

This *LPH* LS is the same as the priority heuristic (PH) of Brandstaetter et al. (2006), except that the priority heuristic assumes that $\Delta_L = 0.1 \cdot \max(x, x')$, and $\Delta_P = 0.1$. The priority heuristic also assumes that the value of Δ_L is rounded to the nearest prominent number (integer powers of 10 plus one-half and double their values; i.e., 1, 2, 5, 10, 20, 50, 100, etc.) If a study involved only choices in which the highest prize in either gamble was \$100, then the priority heuristic model would be a special case of this *LPH* LS model where $\Delta_L = \$10$ and $\Delta_P = 0.1$. (In addition, the priority heuristic assumes that no matter how many branches there are in a gamble, people use at most four dimensions: the lowest consequence, probability of the lowest consequence, highest consequence, and probability of the highest consequence.)

With two-branch gambles, there are five other LS models, each with two parameters: *LHP*, *PLH*, *PHL*, *HPL*, and *HLP*, which differ only in the order in which the dimensions are considered. For example, the *PHL* LS model assumes that people first compare probabilities, which if they differ by less than Δ_P cause the person to check the highest outcomes, which are decisive only if the difference is greater than or equal Δ_H; otherwise, the person bases the decision on the lowest consequences.

All six of these LS models imply that no two dimensions should show dimension integration. There are three possible pairs of two dimensions: Lowest prize and highest prize, lowest prize and probability, and highest prize and probability. All six models imply no dimension integration in any of these three types of pair-wise tests. Birnbaum (submitted, Study 2) found evidence of dimension integration (described more precisely below), which violates the lexicographic semi-orders and the priority heuristic. However, in one of his tests, it appeared that a substantial number of individuals might be consistent with various lexicographic semi-order models.

This study will examine that case more deeply. In Birnbaum's (submitted) study of integration of the lowest and highest consequences, there were 92 people (out of 242) whose data showed integration of dimensions, but there were 80 who showed a response pattern compatible with three of the lexicographic semi-orders, 25 who were consistent with predictions of the priority heuristic, and another 13 who were consistent with other lexicographic models. This study improves on previous work in its selection of levels and in its use of replicated tests, which allows us to distinguish if such response patterns are due to "error" or "true" intention.

Test of dimension integration

Consider the series of four choices in Table 1, which tests integration of the lowest and highest consequences of a gamble. In this test with 50–50 gambles, the second alternative ("safe", *S*) is always the same. According to the priority heuristic model of Brandstaetter et al. (2006), the majority should prefer the second, "safe" gamble in all four choices because its lowest consequence is always at least \$20 higher than the lowest consequence in the first, "risky" gamble, and this always exceeds 10% of the highest consequence. However, in Choice 4, only 27% of 260 people preferred the "safe" gamble, and this is significantly less than 50%, contrary to the priority heuristic.

According to integrative models such as the TAX model, consequences within a gamble are aggregated. With parameters used to describe other data, the TAX model implies that a \$10 difference in the highest consequence fails to outweigh a \$20 difference in the lowest consequence in Choice 2; and a \$45 difference in the highest consequence in Choice 3 does not overcome a \$50 difference in the lowest consequence. However, their combination is predicted to tip the balance and produce a preference for the risky gamble in Choice 4. This pattern is also compatible with many other integrative models, including expected utility.

Table 1 shows predicted preferences for six LS models made from two orders of considering the two dimensions that are manipulated (lower and higher prizes) with three assumptions about the difference threshold parameters. Predicted preference for the safe option is indicated by "*S*" in Table 1; predicted preference for the risky gamble is indicated by "*R*." For example, a person who considered the lowest consequence first would always choose *S* if $\Delta_L \leqslant \$20$, since the "safe" option always has a lowest outcome that is at least \$20 higher than the lowest consequence in the "risky" gamble. This model is labeled *LH1* in Table 1, and its predicted pattern is denoted *SSSS*, because the person should choose *S* in all four choices. However, if $\$20 < \Delta_L \leqslant \$50$ (*LH2*), then this person would choose the "safe" option in Choices 1 and 3, but choose the "risky" gamble in the other two choices (*SRSR*). If $\Delta_L > \$50$, the person would always choose the "risky"

Table 1
Test of dimension integration

Choice			%*S*	Choice models							
No.	Risky (*R*)	Safe (*S*)		*LH1*	*HL3*	*LH3*	*HL1*	*HL2*	*LH2*	*TAX*	LS mixture
1	50 to win \$60 50 to win \$0	50 to win \$50 50 to win \$50	88	*S*	*S*	R	*R*	*S*	*S*	*S*	$a+b+c$
2	50 to win \$60 50 to win \$30	50 to win \$50 50 to win \$50	72	*S*	*S*	*R*	*R*	*S*	*R*	*S*	$a+c$
3	50 to win \$95 50 to win \$0	50 to win \$50 50 to win \$50	72	*S*	*S*	*R*	*R*	*R*	*S*	*S*	$a+b$
4	50 to win \$95 50 to win \$30	50 to win \$50 50 to win \$50	27	*S*	*S*	*R*	*R*	*R*	*R*	*R*	a

Six variants of lexicographic semi-order model make different predictions.

Notes: LH refers to lexicographic semi-order (LS) model in which lowest consequence (*L*) is considered before the highest (*H*); HL refers to LS model in which highest consequence is considered first. *R* = predicted preference for the "risky" gamble; *S* = predicted preference for the "safe" option. In *LH1*, *LH2*, and *LH3*, $\$0 < \Delta_L \leqslant \$20$, $\$20 < \Delta_L \leqslant \$50$, and $\$50 < \Delta_L$, respectively. In *HL1*, *HL2*, and *HL3*, $\$0 < \Delta_H \leqslant \$10$, $\$10 < \Delta_H \leqslant \$45$, and $\$45 < \Delta_H$, respectively. In the LS Mixture model (last column), *a*, *b*, and *c* are the probabilities that a person uses *LH1* or *HL3* (*SSSS*), *LH2 (SSRR)*, and *HL2 (SRSR)*, respectively; the mixture model is further described in text. TAX refers to the transfer of attention exchange model with parameters from previous data, which predicts *SSSR* pattern.

gamble, which always has the higher best consequence (*LH3*).

Similarly, if a person started by comparing the highest consequences and if $\Delta_H \leqslant \$10$, then that person would also always choose the "risky" gamble (*HL1*). If $\$10 < \Delta_H \leqslant \$45$, as in *HL2*, that person would choose the risky gamble on Choices 1 and 2 and the safe option on Choices 3 and 4 in Table 1 (*SSRR*). Finally, *HL3* assumes that a person starts with the highest consequence but $\Delta_H > \$45$; in this case, the person would always choose the "safe" option. Table 1 shows that of these six possible lexicographic semi-orders for this situation, none of them produces the integrative pattern of preferences predicted by integrative models such as TAX with its prior parameters, which is *SSSR*.

The choice percentages in Table 1 display results of an experiment described below with 260 participants. The majority choice percentages agree with the TAX model and do not agree with any of the LS models in Table 1. But is it possible that the results reflect a mixture of LS strategies?

The LS mixture model

Consider the possibilities that (a) different people might employ different versions of these LS models, or that (b) the same individual might alternate among different LS models. For example, a person might start with the lowest prize on one trial and then start with the highest prize on another trial. On one trial, a person might use one value for Δ_L and use a different value on another trial. But suppose that on any given choice, people use one of the six LS strategies listed in Table 1. Let $a =$ the probability of using either *LH1* or *HL3* (i.e., the probability of choosing "safe" in all four choices, *SSSS*); let $b =$ the probability of using model *LH2* (which generates the pattern *SRSR*) and let $c =$ the probability of using model *HL2* (*SSRR*). According to this LS *mixture model*, the probabilities of choosing the "safe" option in Choices 1, 2, 3, and 4 are $a + b + c$, $a + c$, $a + b$, and a, respectively. There are four empirical proportions and three unknowns. We can estimate the three parameters from Choices 1, 2, and 3, and use them to predict the actual choice proportion for Choice 4. That comparison provides a test of the LS mixture model.

In Table 1, the difference between the first two empirical choice proportions gives an estimate of $\hat{b} = 0.88 - 0.72 = 0.16$. Similarly, the difference between first and third gives, $\hat{c} = 0.16$; therefore, we subtract these two estimates from the choice proportion in Choice 1, yielding $\hat{a} = 0.88 - 0.16 - 0.16 = 0.56$, which is compared with the observed proportion in Choice 4, which should be the same, 0.56, apart from error. Instead, the observed proportion is 0.27, which is significantly less than 0.56. This result indicates that we cannot represent the choice proportions in Table 1 as a mixture of LS models.

Individual differences and error theory

Now suppose that each person has a different "true" pattern that might be one of the LS patterns, or the pattern predicted by TAX, or indeed any of the 16 possible response patterns in 4 choices ($2^4 = 16$). Suppose also that when presented with the same (or nearly identical) choices, the person has the same "true" pattern of preferences, but may have an "error" in evaluating his or her true preference on any given trial. Perhaps some choices are "easier" than others; in which case, the rate of "error" would be lower. This error model resembles that of Sopher and Gigliotti (1993), with improvements by Birnbaum (2004b) and Birnbaum and Bahra (2007).

We can estimate error rates in this model from reversals of preference between repeated presentations. For example, consider Choice 1 in Table 1. Let $p =$ the probability that a person truly prefers the "safe" gamble in this choice, and $e =$ the probability that a person makes an error. It is assumed that $0 \leqslant e \leqslant 1/2$. The probability of choosing the "safe" alternative on both repetitions is given as follows:

$$P(SS) = p(1 - e)^2 + (1 - p)e^2 \qquad (2)$$

In this case, the person who "truly" prefers *S* has correctly reported the preference twice and the person who truly prefers *R* has made two errors. The probability that a person reverses preference from *R* to *S* is given as follows:

$$P(RS) = pe(1 - e) + (1 - p)(1 - e)e = e(1 - e) \qquad (3)$$

The probability that a person shows the opposite reversal of preference, $P(SR)$, is also $e(1 - e)$, so the probability of either type of preference reversals is $2e(1 - e)$. If we observe 32% preference reversals, for example, we estimate that the error rate, $e = 0.2$, because $2e(1 - e) = 2 \cdot 0.2 \cdot 0.8 = 0.32$. Similarly, if we observed $P(S) = 0.8$ and $P(SS) = 0.64$, we would estimate that the true probability of preference is 1, because $P(SS) = 1 \cdot (0.8) \cdot (0.8) + 0 \cdot (0.2) \cdot (0.2) = 0.64$. Except in limiting cases where everyone has the same true preferences, we do not expect independence to hold because different people have different true preference patterns. These calculations are analogous to the "correction for attenuation" in test theory. We extend this model below to the replicated test of dimension integration with four choices, where each of the four choices is allowed to have a different "error" rate and each person is allowed to have a different "true" preference pattern.

For the test of dimension integration, the formulas must be expanded to account for patterns of four choices, each of which is replicated. In other words, the model represents the probabilities of showing response patterns on eight choices. This expansion, presented in the results section, allows one to estimate the "true" probability of each of the 16 possible choice patterns in this test. This allows us to estimate the proportion of people who show patterns compatible with or in violation of patterns predicted by different choice models (as in Table 1).

Method

Participants made choices between gambles that were displayed via browsers on computers. They were told that each gamble consisted of a container holding exactly 100 tickets with different values printed on them, and a randomly drawn ticket would determine the gamble's prize. Each choice appeared as in the following example:

```
1.Which do you choose?
  A: 50 tickets to win $100
     50 tickets to win $0
OR
  B: 50 tickets to win $35
     50 tickets to win $25
```

Participants clicked a button beside the gamble they would rather play in each choice. Instructions are available from the following URL: http://psych.fullerton.edu/mbirnbaum/psych466/exps/gls_2-branch.htm.

Replicated lower by upper consequence

This study was included among a series of similar, self-contained studies of judgment and decision making. This study consisted of 23 choices between pairs of 50-50 gambles, which can be viewed at the following URL: http://psych.fullerton.edu/mbirnbaum/psych466/exps/ph_lh_adl.htm.

Table 2
Series A test of dimension integration ($n = 260$)

Choice No.	Choice		Replication pattern				Mixture model		Parameter estimates	
	Risky (R)	Safe (S)	*RR*	*RS*	*SR*	*SS*	%*S*	Theory	*p*	*e*
19	50 to win $51 50 to win $0	50 to win $50 50 to win $50	7	13	8	232	93	$a+b+c$	0.973	0.043
23	50 to win $51 50 to win $40	50 to win $50 50 to win $50	14	26	40	180	82	$a+c$	0.956	0.152
9	50 to win $80 50 to win $0	50 to win $50 50 to win $50	28	34	17	181	79	$a+b$	0.879	0.116
5	50 to win $80 50 to win $40	50 to win $50 50 to win $50	177	22	43	18	19	a	0.065	0.153

Notes: Choices 20, 13, 16, and 12 were the same as 19, 23, 9, and 5, respectively, except that the positions of "safe" and "risky" gambles were reversed, the "safe" gamble was always ($51, 0.5; $49, 0.5) instead of ($50, 0.5; $50, 0.5), $51 in the "risky" gamble was replaced by $52, $80 was replaced by $82; and both $0 and $40 were unchanged. Parameter estimates (last two columns) are best-fit values from the true and error model, fit to response patterns for replications.

Table 3
Series B test of dimension integration ($n = 260$)

Choice No.	Choice		Replication pattern				Mixture model		Parameter estimates	
	Risky (R)	Safe (S)	*RR*	*RS*	*SR*	*SS*	%*S*	Theory	*p*	*e*
11	50 to win $60 50 to win $0	50 to win $50 50 to win $50	14	18	9	219	89	$a+b+c$	0.94	0.06
22	50 to win $60 50 to win $30	50 to win $50 50 to win $50	33	39	26	162	75	$a+c$	0.85	0.15
7	50 to win $95 50 to win $0	50 to win $50 50 to win $50	44	30	30	156	72	$a+b$	0.79	0.13
4	50 to win $95 50 to win $30	50 to win $50 50 to win $50	147	42	35	36	29	a	0.17	0.18

Notes: Choices 15, 17, 2, and 18 were the same as 11, 22, 7, and 4, respectively, except that the positions of "safe" and "risky" gambles were reversed, the "safe" gamble was always ($52, 0.5; $48, 0.5) instead of ($50, 0.5; $50, 0.5), $60 in the "risky" gamble was replaced by $59, $95 was replaced by $89; $30 was replaced by $31, and $0 was unchanged.

There were two series of four choices each testing dimension integration (Series A and B), where each choice was replicated in a slightly altered version and with positions of "safe" and "risky" options counterbalanced. These are described in Tables 2 and 3.

In addition, there were six "filler" choices (Series C), designed to test a specific prediction of intransitivity with replication. These will be described in Discussion. Choices from all three series were intermixed and presented in random order.

Although the Internet was used as a network for display of the experimental materials and collection of data, participants were recruited from the usual "subject pool" in the psychology department and tested in labs via computers connected to the WWW. There were 260 college undergraduates, enrolled in lower division psychology, who completed all choices. Of these, 61% were female and 92% were 22 years of age or younger.

Results

LH1 lexicographic semi-order model of Table 1 and the priority heuristic of Brandstaetter et al. (2006) imply that the percentage choosing the "safe" gamble (labeled "%*S*" in Tables 2 and 3) should be greater than 50% in all four rows. Instead, Tables 2 and 3 show that the majority responses conform to the pattern predicted by integrative models such as TAX and EU. The first three percentages (93%, 82%, and 79%) in Table 2 are significantly greater than 50% and the fourth (19%) is significantly less than 50%. (For $n = 260$, percentages outside the interval from 44% to 56% fall outside a 95% confidence interval and are "significantly different" from 50%). The same result was observed four times (two replicates each of the tests in Series A and B; i.e., Tables 2 and 3) and these were significant in all four cases.

The choice percentages in Tables 2 and 3 violate the LS mixture model that allows people to switch from one lexicographic semi-order to another and to use different parameter values on different trials. In Table 2, the estimated parameters are $\hat{b} = 93 - 82 = 11\%$, $\hat{c} = 93 - 79 = 14\%$, so from the first three percentages, we have $\hat{a} = 93 - 11 - 14 = 68\%$, which should equal the choice percentage in the fourth row of Table 2. Instead the observed choice percentage in the fourth row of Table 2 is only 19%, significantly less than 50%. The figures for Table 3 are similar: based on the first three percentages, the estimates are $\hat{b} = 89 - 75 = 14\%$, $\hat{c} = 89 - 72 = 17\%$, so the first three percentages, imply $\hat{a} = 58\%$. The observed percentage in the fourth row is only 29%, significantly less than 50%.

When we "correct" the estimated choice proportions for unreliability, according to the "true and error" model (last two columns of Table 2), the estimated "true" percentages in the four rows of Table 2 are 97, 96, 88, and 07, respectively. The estimated "true" percentages in Table 3 (last two columns) are 94, 85, 79, and 17. In both cases, the corrected percentages are still closer to the predictions of TAX and farther from the predictions of the priority heuristic.

True and error model: Individual differences

The frequencies of each response pattern for Tables 2 and 3 have been tabulated in Tables 4 and 5, respectively. The pattern *SSSR* in the next to last row of Table 4 indicates preference for the "safe" option in Choices 19, 23, and 9, and preference for the "risky" alternative in Choice 5, respectively. The entry of 149 in the column labeled "Replicate 1" shows that 149 people showed this response pattern on these four choices. Responses to Choices #20, 13, 16, and 12 (see note to Table 2) are treated as Replication 2 of #19, 23, 9, and 5, respectively. The entry in the second column of the next to last row shows that 132 people showed the *SSSR* pattern on these four trials. The 98 in the column labeled "Both" indicates that 98 people showed the *SSSR* pattern on both replicates (all eight of these choices) in Series A.

The PH of Brandstaetter et al. (2006) implies that people should show the pattern *SSSS*. The last row of Table 4 shows that 24 people showed this pattern on the first replication, 43 showed this pattern on the second replication, and only 9 people showed this same pattern on both sets of four choices. None of the lexicographic semi-orders predicts the modal pattern *SSSR*,

Table 4
Tests of dimension integration, Series A ($n = 260$).

Response pattern	Number who show each pattern			Estimated probability
	Replicate 1	Replicate 2	Both replicates	
RRRR	1	6	1	0.03
RRRS	1	6	1	0.01
RRSR	6	3	0	0.02
RRSS	0	0	0	0
RSRR	3	4	0	0.01
RSRS	1	1	0	0
RSSR	2	0	0	0
RSSS	1	0	0	0
SRRR	13	4	0	0
SRRS	0	1	0	0
SRSR	26	14	4	0.02
SRSS	7	6	0	0
SSRR	20	36	6	0.04
SSRS	6	4	0	0
SSSR	**149**	**132**	**98**	**0.80**
SSSS	24	43	9	0.06

Replicate 1 consisted of choices #19, 23, 9, and 5. Replicate 2 used reversed positions reversed (see Table 2), "both replicates" indicates the same pattern was repeated on both sets. Estimated probabilities are estimates in the true and error model, with all parameters free. Entries in bold show results for the pattern predicted by the TAX model with prior parameters.

Table 5
Tests of dimension integration using Series B (See Table 3, $n = 260$)

Response pattern	Number who show each pattern			Estimated probability
	Replicate 1	Replicate 2	Both replicates	
RRRR	9	11	5	0.06
RRRS	2	3	0	0
RRSR	2	3	0	0
RRSS	2	3	0	0.02
RSRR	1	5	1	0.01
RSRS	3	1	0	0.01
RSSR	2	1	0	0
RSSS	2	5	0	0.01
SRRR	10	12	4	0.04
SRRS	6	5	0	0
SRSR	21	31	6	0.04
SRSS	7	4	0	0
SSRR	40	30	13	0.11
SSRS	3	7	0	0
SSSR	**97**	**96**	**52**	**0.52**
SSSS	53	43	24	0.20

Estimated probabilities are estimates in the true and error model, with all parameters free. Entries in bold show results for the pattern predicted by the TAX model with prior parameters.

which is implied by integrative models, such as TAX as fit to previous data. Similar results are shown for Series B in Table 5, where 97 and 98 people show the pattern predicted by TAX, including 52 who showed it on Choices #11, 22, 7, and 4 as well as on Choices 15, 17, 2, and 18. Indeed, this pattern violating the lexicographic semi-orders and priority heuristic was the most frequent pattern for individuals in all four sets of choices (two replicates each of Series A and Series B).

To estimate the proportion of individuals that "truly" has each choice pattern we extend the "true and error" model to a study with four choices and two replications. The probability that a person who "truly" has the pattern predicted by the priority heuristic (*SSSS*) will show instead the pattern predicted by TAX (*SSSR*) on four choices is given as follows:

$$P(SSSR|SSSS) = (1 - e_1)(1 - e_2)(1 - e_3)e_4 \quad (4)$$

Assuming that the true pattern is *SSSS*, this person has correctly reported his or her preference on the first three choices and made an error on the fourth choice. The probability that a person will show this same pattern on two replications of the four choices, given her or his true pattern is *SSSS* is as follows:

$$P(SSSR \cap SSSR|SSSS) = (1 - e_1)^2(1 - e_2)^2(1 - e_3)^2 e_4^2 \quad (5)$$

Here a person has reported six preferences correctly and made two errors on the fourth choice. The probability that a person exhibits the preference pattern *SSSR* on one replication is the sum of 16 terms as follows:

$$P(SSSR) = \sum_{i=1}^{16} P(SSSR|H_i)p(H_i) \quad (6)$$

where $P(SSSR|H_i)$ is the probability of showing the *SSSR* pattern given the "true" pattern is H_i where $H_1 = SSSS$, $H_2 = SSSR$, $H_3 = SSRS, \ldots, H_{16} = RRRR$, and $p(H_i)$ are the true probabilities that people have the hypothesized patterns, H_i as their "true" patterns. There are sixteen equations for the sixteen possible response patterns in Expression 6. Four of these 16 preference patterns are compatible with LS models, *SSSS*, *SRSR*, *SSRR*, and *RRRR* (see Table 1).

With two replications of the four choices in Table 1, there are 256 possible response patterns with 256 equations. This model has been fit to the data, which allows us to estimate the rates of "errors" on the four choices, and the "true" probabilities of the 16 possible patterns. The six LS models in Table 1 permit only four "true" response patterns (*SSSS*, *SRSR*, *SSRR*, and *RRRR*). All other sequences should have zero probability, including the pattern predicted by the TAX model with its parameters estimated from previous data (*SSSR*).

This "true and error" model was fit to the frequencies in Tables 4 and 5, which show the individual response patterns for Series A and B (Tables 2 and 3), respectively. In the most general version of the "true and error" model fit to the data, there are sixteen "true" probabilities and four "error" probabilities to estimate in each series. These are estimated from the 16 frequencies of each pattern on both replicates and the 16 average frequencies of each pattern on either the first or second replicate but not both. These 32 mutually exclusive frequencies sum to the total number of participants and have 31 degrees of freedom.

The columns labeled "estimated probability" in Tables 4 and 5 show the best-fit estimates for Series A and B, respectively, which were estimated to minimize the χ^2 between predicted and obtained frequencies. The fit of the general model in Tables 4 and 5 yielded $\chi^2(12) = 20.1$ and 15.4, respectively. Neither is significant (with $\alpha = 0.05$), suggesting that the general "true and error" model can be retained for both Series A and B.

The class of LS models was tested by fitting a special case of the general true and error model, with the restriction that the "true" probability of the *SSSR* pattern (which violates all six LS models) is 0, and all other parameters are free. These solutions yield $\chi^2(13) = 199.1$ and 73.6, with differences of $\chi^2(1) = 179.0$ and 58.2, which are significant and quite large (the critical value of $\chi^2(1) = 3.84$ for $p < .05$). Therefore, we can reject the hypothesis that priority heuristic model or any combination of the six LS models is descriptive of the data of either Series A or B.

Models with fewer parameters than used in the general model are also compatible with the data. For example, the assumption that only the *SSSS*, *RRRR*, and *SSSR* patterns have non-zero probabilities fits the data of Series A (Table 4) with $\chi^2(25) = 30.1$, an acceptable fit. For this solution, the best-fit estimate is that 86% of the participants had *SSSR* as their "true" pattern. For Series B, assuming that *SSSS*, *SSRR*, *SSSR* and *RRRR* were the only real patterns, the model yielded, $\chi^2(24) = 28.1$, with 53% estimated to have the *SSSR* pattern as their "true" pattern. In sum, data from both series indicate that the majority of people show evidence of integration, contrary to all LS models and contrary to priority heuristic. The most frequent response pattern by individuals is the pattern predicted by TAX with its prior parameters. This pattern is compatible with other integrative models as well, including EU.

Discussion

These results give clear answers to five empirical questions: first, the majority data are not consistent with the priority heuristic, which implies that the majority should have chosen the "safe" gamble in all four rows of Tables 2 and 3.

Second, the majority data are not consistent with any of six possible LS models (Table 1). This means that we can reject all six LS models with any order of considering the dimensions and with any threshold parameters.

Third, we can reject the hypothesis that the data are produced by people shifting randomly among a mixture of these six different LS models from trial to trial.

Fourth, we can reject the hypothesis that most people do not integrate information in favor of the hypothesis that the majority of individuals in this study did integrate the information.

Fifth, if some people are using the LS models or the priority heuristic, there are not very many of them. For example, in Table 4, the "true and error" model indicates that 6% of the participants had *SSSS* as their "true" pattern. This pattern is consistent with the priority heuristic. (It is consistent as well with other models, including integrative models). If we supposed that all of these people used the priority heuristic, we would estimate that 6% of participants used this strategy. Summing over all patterns compatible with LS models, perhaps as much as 15% of the sample used a lexicographic semi-order in this test.

Can we revise the priority heuristic model to give a better account of these data? Brandstaetter et al. (2006) suggested that the priority heuristic model might be extended to include the hypothesis that the first reason considered is expected value (EV). According to this revised model, if one alternative yields an EV twice as great as the other or more, people choose the gamble with the higher EV. Only if the EVs differ by less than a factor of 2 do they employ the priority heuristic as described here.

The computation of EV involves integration of probabilities and prize values, which would allow this EV model to account for evidence of dimension integration for any pair of dimensions. However, the choices used in Series A and B differ by less than a factor of 2 on the crucial trials where the priority heuristic goes wrong. In Series A, Choice 5 (Table 2) the expected values are \$60 and \$50 and yet people violate the priority heuristic. In Choice 4 of Series B (Table 3), EV = \$62.5 and \$50. These differ by only 20% and 25% in EV, respectively (the EVs in the counterbalanced replicate versions are similar). Thus, we can reject this more complex extension of priority heuristic that allows EV as the dimension with highest priority, as long as the threshold for ratios of EV is assumed to be greater than 1.25.

A second way to revise the priority heuristic to account for evidence of trade-offs, as in Tables 2 and 3 would be to extend the approach of González-Vallejo (2002) and incorporate that into the priority heuristic. In her approach, the difference along a given dimension is compared to the maximum value of that dimension within a choice. She used an additive difference model, which is integrative across all pairs of dimensions. As is the case in the priority heuristic, her additive difference model is not transitive. Brandstaetter et al. (2006) compare the difference in lowest outcomes to the largest outcome in either gamble (which they treat as an aspiration level defined on a choice), but otherwise do not adopt her integrative model.

A third way to modify the priority heuristic would be to assume that the parameters change for each new set of choice problems. Note that the ratio of the difference between the two lowest consequences to the maximum consequence in either gamble is 0.98, 0.20, 0.62, 0.12 in Series A, and 0.83, 0.33, 0.53, and 0.21 in Series B. We cannot set a single parameter, δ_L, such that $\Delta_L = \delta_L \max\ (x,x')$ to account for the reversals. However, if we allow that different δ_L should be permitted in Tables A and B, we could account for the data if we assumed that $0.125 < \delta_L \leqslant 0.20$ for Table A and $0.21 \leqslant \delta_L < 0.33$ in Table B. This approach seems unattractive because it requires one parameter to fall in two mutually exclusive ranges.

If we allow different parameters and incorporate the rounding assumption of the priority heuristic, we could take $\Delta_L = R[.44 \max\ (x,x')]$, where $R[.]$ represents the rounding to nearest prominent numbers (integer powers of 10 plus double and half their values; i.e., 1, 2, 5, 10, 20, 50, 100, etc.). This would yield Δ_L = \$20, \$20, \$50, \$50 for successive choices in both Tables 2 and 3, which would agree with both tables. Because Brandstaetter et al. (2006) are skeptical of estimation of any parameters from the data (they argue that their parameter 0.1 is based on the cultural base ten number system), it seems doubtful that they would consider any of these modifications to their theory to be very attractive.

What can one make of the seemingly "good" fit of priority heuristic to previously published data according to Brandstatter, et al? Birnbaum (in press) presented four objections concerning their contests of fit. First, Brandstaetter et al. (2006) did not analyze a number of previous studies where the priority heuristic fails to predict the results. The priority heuristic does not account for the observed pattern of violations of restricted branch independence (Birnbaum & Navarrete, 1998); it makes wrong predictions for more than half the modal choices in that study. The priority heuristic cannot account for violations of distribution independence (Birnbaum, 2005b, 2005c; Birnbaum & Chavez, 1997). It fails to predict violations of stochastic dominance in cases where 70% of participants violate it (Birnbaum, 1999, 2004a, 2005a, 2005b, 2005c; Birnbaum & Navarrete, 1998) and it fails to predict satisfactions of stochastic dominance in cases where 90% or more satisfy it. It does not account for systematic violations of upper and lower cumulative independence (Birnbaum, 1997, 1999, 2004b, 2006; Birnbaum & Navarrete, 1998).

Second, their contests of fit did not allow parameter estimation to the models that use parameters. Parametric models do not assume that everyone has the same parameters nor do they assume that every experiment will induce the same parameters. For example, both CPT and TAX can perfectly fit the Kahneman and Tversky (1979) data if they are allowed to estimate a parameter representing the exponent of the utility function from those data. Because those data can be fit perfectly by TAX, CPT, and PH, those data are simply not diagnostic among these models. The conclusion of Brandstaetter et al. (2006) that the data fit better for PH than CPT or TAX is strictly based on use of non-optimal parameters estimated from other data and extrapolated to those data. When parameters are estimated for all models compared, the conclusions reverse: the best-fit TAX and CPT models outperform the best-fit version of PH.

Third, global indices of fit can be systematically misleading when comparing the success of models when we do not allow a model to estimate its scales and parameters from the data (Birnbaum, 1973, 1974). Apparently "good" fit indices often coexist with systematic errors of prediction. A closer look at the data that were treated in Brandstätter et al. shows that the priority heuristic makes systematic errors in predicting the data of Tversky and Kahneman (1992) and Mellers, Chang, Birnbaum, and Ordóñez (1992).

The method of analysis in Brandstaetter et al. (2006) contains an additional problem: they used one index of fit to optimize certain models, and then compared models on another index. The parametric models are usually fit with least-squares or maximum likelihood, whereas heuristic models may be devised to maximize percent correct. A least-squares solution does not necessarily produce the highest percentage of correct predictions. If we compare models based on percentage correct, we should use that same criterion to optimize fit in both models to be compared.

When one analyzes only success in predicting modal choices, one can miss quite a lot of useful information. For example, by counting individual choices in Table 2, we might say that 75% of the modal choices were correctly predicted by the priority heuristic (it predicts *S*, *S*, *S*, and *S*), and 100% of modal choices were correctly predicted by TAX. However, when we examine response patterns of individuals, as in Table 4, we see that only 6% of the participants showed the combined response pattern predicted by the priority heuristic (*SSSS*), whereas 80% show the pattern predicted by the TAX model (*SSSR*). Although these are aspects of the same data, they contain different information and convey quite different impressions of the relative merits of the models.

A better way to compare models (than by computing global indices of fit to selected data) is to investigate their implications, and test predictions when those implications are different. Birnbaum (submitted) noted that the family of LS models implies a property he called priority dominance, implies no dimension integration (as tested here), no dimension interaction, and violates transitivity. On the other hand, transitive, integrative models (such as TAX, CPT, and GDU) violate priority dominance, show both dimension integration and interaction, and satisfy transitivity. Birnbaum reported four tests of dimension integration involving all pairs of dimensions in two-branch gambles, including a test of probability and highest consequence, probability and lowest consequence, and lower and upper consequences. All tests showed systematic evidence that people are integrating each pair of dimensions. He also reported tests of dimension interaction showing evidence of a multiplicative relation between probability and prize.

The priority heuristic and LS models imply that most people should be systematically intransitive in certain situations where the EV are nearly equal, as in Tversky's (1969) study. But Tversky (1969) never claimed that most people are intransitive, only that some people can be pre-selected who violate transitivity. Tversky's selected data have been reanalyzed, with the result that not all analyses agree that anyone was significantly intransitive in his study (For contrasting analyses and arguments, see papers by Iverson & Falmagne, 1985; Iverson, Myung, & Karabatsos, submitted for publication; Regenwetter, Stober, Dana, & Kim, 2006).

Birnbaum and Gutierrez (2007) conducted a study in which people were asked to choose between the same gambles used by Tversky, except using procedures similar to those used in most of the studies summarized by Brandstaetter et al. (2006). Whereas Tversky (1969) used pie charts to represent probability and did not present probability information numerically, Birnbaum and Gutierrez presented both probabilities and prizes

Table 6
Test of transitivity (Series C, $n = 260$)

Choice No.	Choice		Response pattern				%*S*	Parameter estimates	
	First (F)	Second (S)	*FF*	*FS*	*SF*	*SS*		*p*	*e*
8	A: 50 to win \$100 50 to win \$20	B: 50 to win \$60 50 to win \$27	190	23	25	22	18	0.09	0.10
3	B: 50 to win \$60 50 to win \$27	C: 50 to win \$45 50 to win \$34	140	44	39	37	30	0.17	0.20
21	C: 50 to win \$45 50 to win \$34	A: 50 to win \$100 50 to win \$20	35	33	20	172	76	0.84	0.12

Notes: Choices 8, 3, and 21 were replicated with choices 6, 14, and 10, respectively, except that the positions of "first" and "second" gambles were reversed. According to any of three lexicographic semi-orders: LPH, LHP, PLH LS with $\$7 < \Delta_L \leqslant \$14$, people should prefer the first gamble in all three rows. This prediction is contradicted by results in the last row. According to PH, the majority should prefer *A* over *B*, *C* over *B*, and *C* over *A*, contrary to data in last two rows. According to TAX with prior parameters, *A B C*, which is consistent with the modal choices.

numerically. They found that modal choices were perfectly consistent with transitivity. Using the true and error model, they estimated that fewer than 5% of individual participants were likely intransitive. Even when probability was displayed with pie charts without numerical probabilities, the estimated percentage of those who appeared to be systematically intransitive was about 6%. These results failed to confirm the predicted pattern of intransitive behavior that people should exhibit according to the priority heuristic.

The present study included a replicated test of transitivity (Table 6). Three of the LS models (cases in which the lowest payoff has priority over the highest payoff) predict violations in this case, if $\$7 < \Delta_L \leqslant \$14$. According to these models, people should prefer *A* to *B*, *B* to *C*, and *C* over *A*. Table 6 shows that in both replicates (counterbalanced for position), most people preferred *A* over *C*, contrary to this prediction.

Table 7 shows the number of people who showed each response pattern in this test of transitivity. When these frequencies are fit to a "true and error" model that allows all eight response patterns (including all transitive and intransitive patterns), the estimated "true" percentages of intransitive cycles were both 0%, and the estimated percentage of people who appear to conform to ordering predicted by the TAX model with prior parameters was 80%.

Table 7
Test of transitivity in Series C (see Table 6)

Response pattern	Observed frequency			Estimated probability
	Replicate 1	Replicate 2	Both replicates	
ABC	13	21	1	0
ABA	147	134	106	0.80
ACC	18	20	8	0.07
ACA	37	38	10	0.06
BBC	9	15	0	0
BBA	10	14	0	0
BCC	15	12	7	0.07
BCA	11	6	1	0

Estimated errors for the three choices are 0.10, 0.13, and 0.11, respectively. The fit of the true and error model yielded, $\chi^2(5) = 5.57$, which is not significant, indicating an acceptable fit. According to this solution, 80% of the participants have the pattern predicted by TAX as their "true" pattern (*ABA*), and no one was intransitive.

The priority heuristic coincides with these LS models if it were assumed that the aspiration level, Δ_L is \$10 in all three choices of Table 6. However, if $\Delta_L = \$5$ in Choice 3 instead, then priority heuristic is wrong on two of the three modal choices in Table 6 since it would then predict that the majority should have chosen *C* over *B* in Choice #3. In fact, only 17% showed this preference on that choice and 7% are estimated to show this transitive order predicted by the priority heuristic. Birnbaum (submitted) summarizes other tests for intransitivity predicted by LS and priority heuristic; none of them show evidence that more than six per hundred are intransitive. Thus, empirical studies do not confirm that the majority of participants show systematic violations of transitivity as predicted by the priority heuristic or the lexicographic semi-orders.

Two possible specifications of variability of response were evaluated in this study. The results cannot be described in terms of a mixture of lexicographic semi-orders in which people randomly use different orders and different threshold parameters from trial to trial. The general "true and error" model (which assumes individual differences in true preferences and random "errors" in response) was evaluated and found compatible with the data. This model showed that the data cannot be described in terms of different people having different true orders that are generated by different lexicographic semi-orders with different parameters; instead, the majority show evidence of the *SSSR* pattern that is not predicted by those models.

Other models have been postulated to describe variability of choice behavior (Busemeyer & Townsend, 1993; Link, 1992; Luce, 1959, 1994; Thurstone, 1927). These models of choice contradict lexicographic semi-orders because they imply transitivity, so it seems inappropriate to assume them when evaluating such intransitive models (Birnbaum & Gutierrez, 2007). Nevertheless, if we were to apply these models to the present data, we would reach the same conclusions.

This study used hypothetical monetary incentives rather than real ones. Previous research with the Allais paradoxes has found that violations of coalescing that appear to produce the Allais paradoxes occur in hypothetical choices among prizes in the millions of dollars as well as in studies with real chances to win modest prizes (less than $100) such as used in this study (Birnbaum, 2007). Similar results have also been obtained for violations of stochastic dominance with real and hypothetical consequences (Birnbaum, 2007; Birnbaum & Martin, 2003). Those who theorize that financial incentives should have some effect usually argue that people should be more "rational" when making real monetary decisions than they would be if the decisions have only hypothetical consequences (Camerer & Hogarth, 1999). If so, then one would theorize that these results underestimate the case against the lexicographic semi-orders, which violate the principle of transitivity, which is widely regarded as a "rational" principle.

In summary, this study contributes to the growing case against lexicographic semi-orders and the priority heuristic as descriptive models of risky decision making. It shows that most people appear to integrate information, contrary to this family of non-integrative models.

References

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

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

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

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

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*(2), 87–106.

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

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

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

Birnbaum, M. H. (2005c). New paradoxes of risky decision making. submitted for publication, Available from URL http://psych.fullerton.edu/mbirnbaum/birnbaum.htm.

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

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

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

Birnbaum, M. H. (submitted) Testing heuristic models of decision making: Priority dominance, dimension integration, and dimension interaction.

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

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

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

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

Brandstaetter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Choices without tradeoffs. *Psychological Review, 113*, 409–432.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A dynamic cognition approach to decision making. *Psychological Review, 100*, 432–459.

Camerer, C. F., & 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.

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

Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10*, 131–153.

Iverson, G. J., Myung, J. I., Karabatsos, G. (submitted for publication). Intransitivity of preference: Revisited, Available from Geoffrey J. Iverson, Institute for Mathematical Behavioral Sciences, UCI, Irvine, CA 92697.

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

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

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, 24*, 178–191.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

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

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

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

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

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

Mellers, B. A., Chang, S., Birnbaum, M. H., & Ordóñez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 347–361.

Regenwetter, M., Stober, C.D., Dana, J., Kim, A. (2006). Tversky's intransitivity of preference revisited. Paper presented to Society for Mathematical Psychology, Vancouver, July, 2006.

Sopher, B., & Gigliotti, G. (1993). Intransitive cycles: Rational Choice or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision, 35*, 311–336.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286 (Reprinted 1994, 101, 266–270).

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*, 31–48.

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