Testing Independence Conditions in the Presence of Errors and Splitting Effects*

Michael H. Birnbaum[a], Ulrich Schmidt[b, c, ‡], and Miriam D. Schneider[b]

[a] Department of Psychology, California State University, Fullerton, USA
[b] Department of Economics, University of Kiel, Germany
[c] Kiel Institute for the World Economy, Kiel, Germany

Abstract
This paper presents an experimental test of several independence conditions implied by expected utility and alternative models. We perform a repeated choice experiment and fit an error model that allows us to discriminate between true violations of independence and those that can be attributed to errors. In order to investigate the role of event splitting effects, we present each choice problem not only in coalesced form (as in most previous studies) but also in split form. It turns out previously reported violations of independence and splitting effects remain significant even when controlling for errors. Splitting effects have a substantial influence on the tests of independence conditions. When choices are presented in canonical split form, in which probabilities on corresponding probability-consequence ranked branches are equal, violations of the independence conditions we tested become either reversed, insignificant or unsystematic.

Key words: Independence axiom, splitting effects, coalescing, errors, experiment

JEL classification: C91, D81

January 2010

---

* We thank Glenn W. Harrison, James C. Cox, Graham Loomes, Peter P. Wakker, Stefan Trautmann, and seminar participants in Tilburg, Orlando, Atlanta, Exeter, Barcelona, Utrecht, and Rotterdam for helpful comments.

‡ Corresponding author: Ulrich Schmidt, Department of Economics, University of Kiel, Olshausenstr. 40, 24098 Kiel, Germany, email: us@bwl.uni-kiel.de, Phone: +49 431 8801400, Fax: +49 431 8804621.

1 Introduction

There exist many studies reporting evidence that expected utility (EU) theory fails to provide an accurate description of peoples´ behavior in several choice problems under risk. One main problem is that choices violate the crucial independence axiom in a systematic way as shown by the famous paradoxes of Allais (1953). These violations have motivated the development of numerous alternative theories (e.g. rank-dependent utility, disappointment and regret models, prospect theory, etc.) which aim to provide a more realistic accommodation of actual choice behavior (see recent surveys by Sugden, 2004, Schmidt, 2004, Birnbaum, 2008, or Abdellaoui, 2009). Most of these new theories rely on independence conditions that are weakened variants of the independence axiom of EU. Experimental investigations of these weakened independence conditions revealed violation rates rather similar to those reported for the independence axiom of EU (Wakker, Erev, and Weber, 1994; Wu, 1994; Birnbaum and Chavez, 1997; Birnbaum, 2004b, 2005a, 2005b, 2008).

Other studies have shown that people are not perfectly consistent when choosing between risky lotteries (see e.g. Camerer, 1989; Starmer and Sugden, 1989; Harless and Camerer, 1994; Hey and Orme, 1994); that is, in repeated choice problems they may choose one option in the first round and choose the other option in the second one. Such preference reversals to the same problem suggest that choices involve a stochastic component. Econometric evaluations of the descriptive performance of EU and its alternatives also require modeling the random component of behavior or otherwise face the risk of reaching wrong conclusions. Nowadays one very intensively discussed questions in decision theory is how to model this stochastic component adequately (e.g., see Gul & Pesendorfer, 2006; Blavatskyy, 2007, 2008; Conte, Hey, and Moffat, 2007; Hey, Morone, and Schmidt, 2007; Wilcox, 2009; Harrison and Rutström, 2009; etc.)

An interesting question in this context is whether the empirical performance of EU improves if we model the stochastic component properly. For instance Hey (1995) concluded: “It may be the case that these further explorations may alter the conclusion to which I am increasingly being drawn: that one can explain experimental analyses of decision making under risk better (and simpler) as EU plus noise – rather than through some higher level functional –

as long as one specifies the noise appropriately (see also Buschena and Zilberman, 2000).[1] Some of the reported violations of EU might be at least partly caused by errors instead of being intrinsic violations. Several recent studies conclude that this may indeed be true, see Blavatskyy (2006) for violations of betweenness, Sopher and Gigliotti (1993), Regenwetter and Stober (2006), Birnbaum and Schmidt (2008, 2009) for violations of transitivity, and Schmidt and Hey (2004), Butler and Loomes (2007), and Berg, Dickhaut, and Rietz (2009) for preference reversals.

A study by Schmidt and Neugebauer (2007) provided some evidence in favor of the idea that random errors may account for the lion's share of evidence against EU. That study considered only choice problems where subjects chose the same option three times in row. Provided that the probability of errors is not too high, such repeated choices should reflect "true preferences" since it is rather improbable that a subject makes the same error three times in row. It turns out that in these choice problems the incidence of violations of independence decreases substantially. See also Butler and Loomes (2009).

In the present paper we will analyze whether reported violations of independence conditions implied by EU as well as violations of weaker conditions implied by certain nonexpected utility models may be attributable to errors. In the next section we show that EU with an error component could easily generate systematic patterns of violations such as have been observed in experimental research.

The goal of the present paper is to provide a more systematic analysis to estimate the incidence of "true" violations as opposed to those that might be attributed to "error". We perform a repeated choice experiment and fit an error model that is neutral with respect to violations of any independence condition. This model allows us to discriminate precisely which portion of violations can be attributed to errors and which part should be considered as "real" violations. Note that such an analysis is not possible with a model of EU plus error term (e.g., as used by Schmidt and Neugebauer, 2007) since that model presupposes that true preferences can be represented by EU and, thus, satisfy the independence axiom, coalescing, and transitivity. The model we use in the present paper, in contrast, assumes only that there is a true choice probability and an error rate that can be different for each choice; it does not assume transitivity, independence, or coalescing (implications of EU we aim to test).

[1] In this context also the studies of Harrison and Rutström (2008, 2009) seem to be relevant which criticize the low statistical power of many tests of EU and find evidence that the behavior of roughly 50% of subjects is better described by EU than by prospect theory.

A further systematic deviation from EU [and in fact also from most of alternatives to EU such as rank-dependent utility (RDU), rank- and sign-dependent utility (RSDU), and cumulative prospect theory (CPT)] is provided by splitting effects (also called violations of coalescing). A splitting effect (violation of coalescing) occurs if splitting an event with a given consequence into two separate events systematically influences choice behavior. For example, coalescing assumes that gamble G = ($50, 0.1; $50, 0.1; $0, 0.8) is equivalent to G' = ($50, 0.2; $0, 0.8). A *branch* of a gamble is a probability-consequence or event-consequence pair that is distinct in the presentation to the participants. In this case, G is a three-branch gamble, and G' is the two-branch coalesced form of G.

There exists abundant evidence that splitting an event with a good consequence increases the attractiveness of a lottery in comparison to other lotteries (Starmer and Sugden, 1993; Birnbaum and Navarrete, 1998 Humphrey 1995, 2001).[2] According to configural weighting models, splitting the branch leading to the lowest consequence can also lower the evaluation of a gamble, even when that worst consequence is positive (Birnbaum, 2007; 2008). While Birnbaum and Navarrete employed splitting effects in order to generate substantial violations of first-order stochastic dominance, the papers of Humphrey note that splitting effects may have contributed to previously reported violations of transitivity (see also Birnbaum & Schmidt, 2008). It may well be the case that splitting effects may also contribute to violations of independence conditions. We will revisit this question while controlling for errors at the same time.

This paper is organized as follows. The next section presents our error model and discusses the issue of testing properties such as independence in the presence of errors. Section 3 presents our experimental design and method while section 4 reports the results. Discussion and concluding observations appear in the last section.

2 Errors and Violations of Independence

In this section we show that random errors can generate systematic violations of independence and present our error model. Consider a simple variant of the common ratio effect taken from Birnbaum (2001).

[2] For similar evidence of splitting effects in other contexts than choice under uncertainty see e.g. Weber, Eisenführ and von Winterfeldt (1988) and Bateman et al. (1997).

Choice 1: Which do you choose?	
R: .99 to win \$0	*S:* .98 to win \$0
.01 to win \$46	.02 to win \$23
Choice 2: Which do you choose?	
R′: .50 to win \$0	*S′:* \$23 for sure
.50 to win \$46	

Figure 1: A Common ratio effect

According to EU theory, a person should prefer R over S if and only if that person prefers R' over S' because $u(23) > (<) 0.5u(46) + 0.5u(0)$ implies $0.02u(23) + 0.98u(0) > (<) 0.01u(46) + 0.99\ u(0)$. There are four possible response patterns in this experiment, *RR′*, *RS′*, *SR′*, and *SS′*, where e.g. *RS′* represents preference for R in the first choice and S' in the second choice. The response patterns *RR′* and *SS′* are consistent with EU while the other two patterns violate the independence axiom of EU. Suppose we obtain data as follows from 100 participants:

	R′	*S′*
R	51	23
S	11	15

Table 1: A response pattern

In this case 23 people switched from R to S', whereas only 11 reversed preferences in the opposite pattern. The conventional statistical test (test of correlated proportions) is significant, $z = 2.06$ which is usually taken as evidence that EU theory is not correct. Can this result have occurred by random errors? Note that in principle systematic deviations from independence can also be explained by EU plus a random error term (for similar discussions, see Birnbaum, 2004b and Loomes, 2005). In this case a subject chooses R over S if $EU(R) - EU(S) + \varepsilon > 0$ where ε is a normally distributed random variable with $E(\varepsilon) = 0$. Suppose $EU(R) > EU(S)$ and note that $EU(R') - EU(S') = 50(EU(R) - EU(S))$. This shows that errors of this particular form may much more easily influence the choice between R and S than the choice between R' and S' implying that we observe more frequent erroneous *SR′* than *RS′* patterns.

Since this error model, however, assumes that true preferences can be represented by EU, it does not allow to test whether true preference are in fact satisfying independence. Therefore, we employ a different, and more general error model that allows us to test whether a given pattern of violations is real, described next.

Suppose that each person has a "true" preference pattern, which may be one of the four possible response combinations. Let $p_{RR'}$, $p_{RS'}$, $p_{SR'}$, and $p_{SS'}$ represent the "true" probabilities of the four preference patterns. These probabilities may be interpreted as the relative frequency of subjects for which true preferences correspond to the given pattern. However, due to errors subjects´ choices may deviate from true preferences. Let e represent the probability of an error in reporting one's true preference for the choice between R and S. Analogously, e' is the probability of an error for the choice between R' and S'. Is it possible that, given the data in Table 1, all subjects adhere to EU? In other words, are the data in Table 1 compatible with $p_{RS'} = p_{SR'} = 0$? The answer is "yes," despite the significant inequality in the two types of violations.

In our model, the probability that a person shows the observed preference pattern RS' is given as follows:

(1) $P(RS') = p_{RR'}(1 - e)e' + p_{RS'}(1 - e)(1 - e') + p_{SR'}ee' + p_{SS'}e(1 - e')$

In this expression, $P(RS')$ is the probability of observing this preference pattern, e and e' are the error rates in the choice between R and S and between R' and S', respectively. This probability is the sum of four terms, each representing the probability of having one of the "true" patterns ($p_{RR'}$, $p_{RS'}$, $p_{SR'}$, and $p_{SS'}$), and having the appropriate pattern of errors and correct responses to produce an observed data pattern. For example, the person who truly has the RR' pattern could produce the RS' pattern by correctly reporting the first choice and making an "error" on the second choice. There are three other equations like (1), each showing the probability of an observed data pattern given the model.

Given only the data of Table 1, this model is under-determined. There are four response frequencies to fit. These have three degrees of freedom, because they sum to the number of participants. The four "true" probabilities must sum to 1, and there are two "error" probabilities. Thus, we have three degrees of freedom in the data and five parameters to estimate, so many solutions are possible. Two solutions that fit the data perfectly are shown in Table 2:

Parameter	Model 1: EU holds	Model 2: EU does not hold
$p_{RR'}$	*0.80*	*0.67*
$p_{RS'}$	*0.00*	*0.17*
$p_{SR'}$	*0.00*	*0.00*
$p_{SS'}$	*0.20*	*0.16*
e	*0.10*	*0.15*
e'	*0.30*	*0.15*

Table 2: Fitting the data

This table shows that we could “save” EU in this case by allowing that people might have unequal errors in the two choices. But if the error rates are assumed to be equal, we would reject EU. So given this second error model and only the data in Table 1 it is not possible to decide conclusively one way or the other. In order to reach a firmer conclusion we need a way to evaluate this model without assuming that error rates are necessarily equal or that EU is correct. Put another way, we need to enrich the structure of the data so that we can determine the model. This can be done by adding replications of each choice problem in the experimental design.

Consider the case of one choice problem presented twice, for example, Choice 1 above. There are four response patterns possible, *RR*, *RS*, *SR*, and *SS*. The probability that a person will show the *RS* pattern is given as follows:

(2) $P(RS) = p(1 - e)(e) + (1 - p)e(1 - e) = e(1 - e).$

where p is the true probability of preferring R and e is the error rate on this choice. The probability of the opposite reversal, *SR*, is also predicted to be $e(1 - e)$. These expressions show that with replications, we can estimate the error rates for each choice without assuming anything about the property to be tested.

By adding replications to both choices in the test of EU, we have now four choices with 16 (4 × 4) possible response patterns, which have 15 degrees of freedom. But we still have only 5 parameters to estimate from the data, two error terms and four probabilities of the four “true” response patterns. (Because the four probabilities sum to 1, only three degrees of freedom are used in this estimation). The general model (which allows all four true probabilities to be non-zero) is now over-determined, with 10 degrees of freedom to test the general error model. EU theory is then a special case of this general model in which two of the true probabilities are fixed to zero. In sum, without replications, two theories are perfectly compatible with these data, one of which assumes EU is true. However, with replications we can estimate the error terms and test the applicability of EU model.

This study will include experiments in which there are four replications. With two choices and four replications, there are 256 possible response patterns ($4^4 = 256$). Because many of these patterns will be observed with zero frequency, we use the *G*-statistic, which is a special case of likelihood ratio test, to measure the badness of fit:

(3) $G = 2\sum f_i ln(f_i/q_i)$,

where f_i is the observed frequency and q_i is the predicted frequency of a particular response pattern. The parameters are then selected to minimize this statistic, which theoretically has a chi-square distribution. An advantage of this statistic over the standard Chi-Square is that cells with zero frequency (i.e., $f_i = 0$) have no effect on the statistic, whereas the standard Chi-Square is known to be biased by cells with low frequency (see Özdemir and Eyduran, 2005; McDonald, 2009). The difference in *G* between a fit of the model that allows all four patterns to have non-zero probabilities and the special case in which $p_{RS'} = p_{SR'} = 0$ is chi-square distributed with 2 degrees of freedom. This test allows us to conclude whether observed violations of EU are real, or whether they might be attributed to errors in the response of subjects. In contrast to the standard test of correlated proportions, our test is also able to qualify symmetric deviations from EU as real violations.

3 Experimental Design

The experiment was conducted at the University of Kiel with 54 participants, mostly economics and business administration students (all undergraduates). Altogether there were six sessions each consisting of nine subjects and lasting about 90 minutes. Subjects received a 5 Euro show-up fee and had to respond to 176 pairwise choice questions which were arranged in four booklets of 44 choices each. After a subject finished all four booklets one of her choices was randomly chosen and played out for real. The average payment was 19.14 Euro for 90 minutes, i.e. 12.76 Euro per hour, which clearly exceeds the usual wage of students (about 8 Euro per hour).

Choices were presented as in Figure 2 and subjects had to circle their preferred alternative. Prizes were always ordered from lowest to highest. Explanation and playing out of lotteries involved a container containing numbered tickets from one to 100. Suppose a subject could for instance play out lottery A in Figure 2. Then she would win 20 Euro if the ticket drawn numbered from 1 to 50, 30 Euro for tickets numbered from 51 to 80, and 40 Euro for a ticket between 81 and 100. All this was explained in printed instructions that were given to the participants and read aloud. Following instructions, subjects had to answer four transparent dominance questions as a test of understanding, which were checked by the experimenter before the participant was allowed to proceed.

A:	*50% to win 20 Euro*	*B:*	*33% to win 10 Euro*
	30% to win 30 Euro		*34% to win 15 Euro*
	20% to win 40 Euro		*33% to win 60 Euro*

Figure 2: Presentation of lotteries

Choices in the booklets were presented in pseudo-random order. The ordering of the choices was different in each booklet with the restriction that successive choice problems not test the same property. Only after finishing each booklet did a subject receive the next one. Moreover, for half of the subjects each booklet contained only coalesced or only split choice problems whereas for the other half split and coalesced choice problems were intermixed in each booklet. Our design included 11 tests of independence conditions, nine of which were investigated in both coalesced and canonical split forms. All 20 tests were replicated four times with counterbalanced left-right positioning. Additionally, in order to check the attentiveness of subjects, each booklet included two transparent stochastic dominance problems, one based on outcome monotonicity and one on event monotonicity.

The lottery pairs for each test are presented in Table 3. Each lottery pair consists of a safe lottery S (in which you can win prize s_i with probability p_i) and a risky lottery R for which possible prizes and probabilities are denoted by r_i and q_i respectively. We took the lotteries from previous studies that reported high violation rates but adjusted outcomes in order to get an average expected value of about 12 Euro. Table 3 shows only the coalesced forms of the lottery pairs. Some of these choice problems were also presented using the canonical split form of those pairs. In canonical split form, both lotteries are split so that there are equal probabilities on corresponding ranked branches and the number of branches is the same in both gambles and minimal. The canonical split forms of these choices can be found in the appendix.

The first six tests in Table 3 include four common consequence effects (CCE1-4) and two common ratio effects (CRE1 and 2). Such tests have been widely used as tests of EU; the paradoxes of Allais are special variants of a CCE and a CRE. This type of CCE can be formally described by $S = (z, p_1; s_2, p_2; s_3, p_3)$, $R = (z, q_1; r_2, q_2; r_3, q_3)$, $S' = (z, p_1 - \alpha; z', \alpha; s_2, p_2; s_3, p_3)$, and $R' = (z, q_1 - \alpha; z', \alpha; r_2, q_2; r_3, q_3)$ and all lotteries are presented in coalesced form. S' and R' are constructed from S and R by shifting probability mass (α) from the common consequence z to a different common consequence z', and converting to coalesced form. An EU maximizer will prefer S over R if and only if she prefers S' over R'. In principle, there is no restriction on the ordering of the consequences in this notation (for example, z and z' might be the lowest and highest consequences or vice versa). In practice, in all of our tests, z = \$0 is the lowest consequence and z' is either the middle or highest consequence of the three. In all of the cases studied here, there are no more than three distinct consequences in each test and coalescing is employed.

In Table 3, the first row of a choice problem always characterizes the lotteries S and R and the second one the lotteries S' and R'. For example, in Choice 5 of CCE1 we have $z = s_1 = 0, p_1 = 0.8, p_2 = 0.2, s_2 = 19, p_3 = 0$ for S; and we have $q_1 = 0.90, q_2 = 0.10, r_2 = 44, q_3 = 0$ for R; in Choice 13, S' and R' are constructed by setting $\alpha = 0.4$ and $z' = r_2 = 44$. In this case, S' has added a new branch leading to \$44, but the branches leading to 44 are coalesced in R'. The four CCE problems of Table 3 are taken from Starmer (1992) who observed high violation rates with them. In all cases, S' has either added a branch leading to the highest consequence or lost the branch leading to the lowest consequence. The typical pattern of violations in CCE1-4 is that people prefer R over S but S' over R'. A test of CRE can be formally described by $S = (z, p_1; s_2, p_2)$, and $R = (z, q_1; r_2, q_2)$, $S' = (z, 1 - \beta(1 - p_1); s_2, \beta p_2)$, $R' = (z, 1 - \beta(1 - q_1); r_2, \beta q_2)$, i.e. S' and R' are constructed from S and R by multiplying all probabilities by β and assigning the remaining probability $1 - \beta$ to the common consequence z. EU implies again that people choose either the risky or the safe lottery in both choice problems. In CRE1 (from Birnbaum, 2001) and CRE2 (from Starmer and Sugden, 1989), however, substantial violations of EU have been observed with many people choosing R and S'.

		Safe Gamble			*Risky Gamble*		
Property	No.	p_1	p_2	p_3	q_1	q_2	q_3
		s_1	s_2	s_3	r_1	r_2	r_3
CCE1	5	*0.80*	*0.20*		*0.90*	*0.10*	
		0	19		0	44	
	13	*0.40*	*0.20*	*0.40*	*0.50*	*0.50*	
		0	19	44	0	44	
CCE2	1	*0.89*	*0.11*		*0.90*	*0.10*	
		0	16		0	32	
	2	*1.00*			*0.01*	*0.89*	*0.10*
		16			0	16	32
CCE3	5	*0.80*	*0.20*		*0.90*	*0.10*	
		0	19		0	44	
	6	*1.00*			*0.10*	*0.80*	*0.10*
		19			0	19	44
CCE4	9	*0.70*	*0.30*		*0.80*	*0.10*	*0.10*
		0	21		0	21	42
	10	*0.70*	*0.20*	*0.10*	*0.80*	*0.20*	
		0	21	42	0	42	
CRE1	15	*0.98*	*0.02*		*0.99*	*0.01*	
		0	23		0	46	
	16	*1.00*			*0.50*	*0.50*	
		23			0	46	
CRE2	20	*0.80*	*0.20*		*0.86*	*0.14*	
		0	28		0	44	
	19	*0.40*	*0.60*		*0.58*	*0.42*	
		0	28		0	44	
UTI	29	*0.73*	*0.02*	*0.25*	*0.74*	*0.01*	*0.25*
		0	15	60	0	33	60
	30	*0.73*	*0.02*	*0.25*	*0.74*	*0.26*	
		0	15	33	0	33	
LTI	33	*0.75*	*0.23*	*0.02*	*0.75*	*0.24*	*0.01*
		1	34	36	1	33	60
	34	*0.75*	*0.23*	*0.02*	*0.99*	*0.01*	
		33	34	36	33	60	
UCI	37	*0.20*	*0.20*	*0.60*	*0.20*	*0.20*	*0.60*
		9	10	24	3	21	24
	38	*0.40*	*0.60*		*0.20*	*0.80*	
		9	21		3	21	
LDI	23	*0.60*	*0.20*	*0.20*	*0.60*	*0.20*	*0.20*
		1	18	19	1	2	32
	24	*0.10*	*0.45*	*0.45*	*0.10*	*0.45*	*0.45*
		1	18	19	1	2	32
UDI	25	*0.20*	*0.20*	*0.60*	*0.20*	*0.20*	*0.60*
		6	7	20	1	19	20
	26	*0.45*	*0.45*	*0.10*	*0.45*	*0.45*	*0.10*
		6	7	20	1	19	20

Table note: The first lottery pair of a choice problem always characterizes the lotteries *S* and *R* and the second one the lotteries *S′* and *R′*.

Table 3: The lottery pairs

The remaining five independence properties in Table 3 are weakened variants of the independence axiom of EU which are assumed or implied by RDU (Quiggin, 1981, 1982; Luce, 1991, 2000; Luce and Fishburn, 1991; Luce and Marley, 2005), CPT (Starmer and Sugden, 1989; Tversky and Kahneman, 1992; Wakker and Tversky, 1993), and configural weight models (Birnbaum and McIntosh, 1996). A central property in this context is tail independence (TI), which was tested by Wu (1994). TI is a special case of ordinal independence (Green and Jullien, 1988). If two lotteries share a common tail (i.e. identical probabilities of receiving any outcome better than x_{i+1}), then the preference between these lotteries must not change if this tail is replaced by a different common tail. Upper Tail Independence (UTI) requires that $S = (s_1, p_1; s_2, p_2; \alpha, p_3) \prec R = (r_1, p_1; \gamma, p_2; \alpha, p_3)$ iff $S' = (s_1, p_1; s_2, p_2; \gamma, p_3) \prec R' = (r_1, p_1; \gamma, p_2 + p_3)$, where $r_1 < s_1 < s_2 < \gamma < \alpha$. TI, however, also demands that preferences must not change if lower common tails are exchanged which will be called lower tail independence (LTI). TI is implied by many models including all variants of RDU as well as CPT. Therefore, rejecting TI would provide serious evidence against all these models. In his experiments, Wu (1994) observed violation rates of UTI as high as 50%. Similar evidence has been reported by Birnbaum (2001). The lotteries we use for the test of UTI are taken from Wu (1994). LTI has, as far as we know, not been tested before. Our construction of lotteries in the test of LTI is similar to that used in the test of UTI.

Another property implied by CPT and the common versions of RDU is upper cumulative independence (UCI), which demands that decision weights depend only on cumulative probabilities. Formally, UCI demands that If $S = (s_1, p_1; s_2, p_2; \alpha, p_3) \prec R = (r_1, p_1; \gamma, p_2; \alpha, p_3)$ then $S' = (s_1, p_1 + p_2; \gamma, p_3) \prec R' = (r_1, p_1; \gamma, p_2 + p_3)$, where $r_1 < s_1 < s_2 < \gamma < \alpha$. Substantial violations of UCI have been reported by Birnbaum and Navarrete (1998) Birnbaum, Patton, and Lott (1999) and Birnbaum (2008). Our lottery pairs are taken from the 1999 paper which observed violation rates of 40.1% for these pairs, for which the typical violating pattern is *RS′*.

The final property we test is distribution independence (DI). Whereas certain configural weight models and extended original prospect theory imply that DI holds, it should be violated according to RDU and CPT, at least if the weighting function is not linear, as commonly implied by empirical research (Camerer and Ho, 1994; Wu and Gonzalez, 1996; Tversky and Fox, 1995; Gonzalez and Wu, 1999; Abdellaoui, 2000; Bleichrodt and Pinto, 2000; Kilka and

Weber, 2001; Abdellaoui, Vossmann, and Weber, 2005). For three-outcome lotteries, DI demands that S = $(s_1, \beta; s_2, \beta; \alpha, 1 - 2\beta) \prec R = (r_1, \beta; r_2, \beta; \alpha, 1 - 2\beta)$ if and only if $S' = (s_1, \delta; s_2, \delta; \alpha, 1 - 2\delta) \prec R' = (r_1, \delta; r_2, \delta; \alpha, 1 - 2\delta)$, where $r_1 < s_1 < s_2 < r_2$. When α is the highest outcome, the condition is called upper distribution independence (UDI); when α is the lowest consequence, it is called lower distribution independence (LDI). The lotteries used in our tests of UDI and LDI are taken from Birnbaum (2005b). The evidence reported in that paper and in Birnbaum and Chavez (1997) indicates that one should observe either no violations or violations contrary to CPT with inverse-S weighting function.

4 Results

We first examine the eight tests of transparent dominance per person. Out of *54* subjects, five violated dominance once and one subject twice (7 out of 8 x 54 = 432 is 1.6%). We conclude from this result that our subjects were sufficiently attentive and motivated. This conclusion is supported by the finding that all our estimated error rates are in line with estimations in comparable studies, see below.

Table 4 summarizes our tests of the single independence conditions, using the true and error model. For all conditions (listed in the first column), we report the estimated probabilities of the four possible response patterns in columns two to five. A subscript, *S* in the first column indicates that this independence condition was tested by presenting both choices in canonical split form. Columns six and seven show the estimated error rates for the choices between *S* and *R (e)* as well as between *S′* and *R′ (e′)*. The final column presents the chi-square statistics between the fit of a general model (a model that allows all four response patterns to have non-zero probabilities) and the fit of a model which satisfies the respective independence condition (i.e., $p_{RS'} = p_{SR'} = 0$). One asterisk (two asterisks) in this column indicate that we can reject the null of $p_{RS'} = p_{SR'} = 0$ in favor of the general model at the 5% (1%) significance level.

We next analyze the tests of independence. This axiom demands $p_{RS'} = p_{SR'} = 0$ whereas empirical research on CCEs and CREs has reported systematical violations, most frequently of the *RS′* pattern. As in previous research with coalesced lotteries, we can confirm this result: In all six tests (CCE1-4, CRE1-2), independence can be rejected. Moreover, most violations are given by responses of the pattern *RS′* (estimated probabilities range from *10%* to *64%*, mean *28%*) whereas the opposite pattern *SR′* occurs very rarely (apart from CCE3, estimated percentages never exceed *3%*). In CCE3, the estimated probability of pattern *SR′* slightly exceeds that of *RS′*. In summary, we find that typical violations of the independence properties of EU can be also observed when errors are taken into account; in other words, these violations can be regarded as true violations.

Property	*Choices*	$p_{SS'}$	$p_{SR'}$	$p_{RS'}$	$p_{RR'}$	*e*	*e′*	Test
CCE1	*5, 13*	*0.44*	*0.02*	***0.30***	*0.24*	*0.15*	*0.11*	***20.36*****
$CCE1_S$	*7, 14*	*0.52*	***0.20***	*0.00*	*0.28*	*0.13*	*0.16*	***12.77*****
CCE2	*1, 2*	*0.02*	*0.00*	*0.10*	*0.88*	*0.02*	*0.08*	***15.33*****
$CCE2_S$	*3, 4*	*0.09*	*0.03*	*0.05*	*0.84*	*0.07*	*0.07*	***7.61****
CCE3	*5, 6*	*0.25*	***0.21***	*0.16*	*0.39*	*0.16*	*0.12*	***18.69*****
$CCE3_S$	*7, 8*	*0.52*	***0.24***	*0.00*	*0.25*	*0.13*	*0.16*	***12.63*****
CCE4	*9, 10*	*0.67*	*0.01*	***0.29***	*0.02*	*0.14*	*0.09*	***21.96*****
$CCE4_S$	*11, 12*	*0.80*	*0.01*	*0.02*	*0.17*	*0.14*	*0.12*	*0.82*
CRE1	*15, 16*	*0.25*	*0.00*	***0.64***	*0.11*	*0.11*	*0.07*	***44.64*****
$CRE1_S$	*17, 18*	*0.44*	*0.00*	***0.46***	*0.10*	*0.15*	*0.05*	***27.21*****
CRE2	*20, 19*	*0.57*	*0.00*	***0.20***	*0.23*	*0.14*	*0.11*	***18.00*****
$CRE2_S$	*21, 22*	*0.84*	*0.02*	*0.01*	*0.12*	*0.17*	*0.12*	*0.45*
UTI	*29, 30*	*0.06*	*0.01*	***0.52***	*0.40*	*0.13*	*0.18*	***18.76*****
UTI_S	*31, 32*	*0.17*	*0.01*	*0.00*	*0.82*	*0.14*	*0.18*	*0.05*
LTI	*33, 34*	*0.04*	*0.00*	*0.14*	*0.82*	*0.05*	*0.15*	*3.96*
LTI_S	*35, 36*	*0.04*	*0.00*	*0.01*	*0.95*	*0.06*	*0.08*	*0.24*
UCI	*37, 38*	*0.14*	*0.08*	*0.13*	*0.66*	*0.13*	*0.22*	*3.88*
UCI_S	*37, 39*	*0.13*	*0.09*	*0.02*	*0.76*	*0.13*	*0.09*	*5.07*
LDI	*23, 24*	*0.94*	*0.00*	*0.00*	*0.06*	*0.02*	*0.05*	*0.00*
UDI	*25, 26*	*0.16*	*0.02*	*0.03*	*0.79*	*0.09*	*0.10*	*1.80*

Table note: * denotes a significance level of 5%, ** a significance level of 1%.

Table 4: Tests of independence conditions

A quite different picture arises when the same tests are performed with choices presented in canonical split form. From our six tests, two ($CCE4_S$ and $CRE2_S$) are not significant (i.e. EU cannot be rejected) and two ($CCE1_S$ and $CCE3_S$) are significant but precisely in the opposite direction as observed in coalesced lotteries (including previous results). From the two remaining tests, only one (CRE1) shows the same pattern in both split and coalesced forms as in previous research whereas for $CCE2_S$ only very low violation rates (i.e. 3% and 5% for both violating patterns) are estimated. We can, therefore, conclude that splitting effects have a substantial influence on tests of the independence axiom of EU, especially with CCE tests. These results are consistent with Birnbaum's (2004a; 2008) hypothesis that CCE are largely due to violations of coalescing rather than to violations of branch independence.

Next, consider the tests of the weaker independence conditions implied or assumed by RDU and CPT. For UTI we observe a substantial and systematic violation: the estimated probability of the violating pattern RS' amounts to *52%*. This picture is entirely in line with the high violation rates observed by Wu (1994) and similar evidence reported by Birnbaum (2001). We can conclude that violations of TI are not caused by errors but reflect true preferences. This is a serious challenge for CPT and the whole class of rank-dependent models which all imply that TI holds. It is, however, astonishing that this clear evidence of violations of TI virtually disappears when we present choices in their canonical split form. The estimated frequency of the RS' pattern decreases from *52%* in the coalesced test to *0%* in the split test while the frequency of the opposite violation SR' amounts to only *1%* in both cases. Therefore, violations of TI in ours (consistent with previous studies) seem to be mainly caused by violations of coalescing.

Our new test of LTI did not generate significant violations. The same is true for UCI. Comparing our results to the high violation rates of UCI observed in previous papers (Birnbaum and Navarrete, 1998; Birnbaum, Patton, and Lott 1999) with lotteries identical to those in our coalesced test, the difference of results may be due to large error rates in these tests. In fact, the estimated error rates in our coalesced test of UCI are the highest of all our choices. The high error rates may explain why we estimated only relatively low true violation rates for UCI whereas our observed violation rates (on average 37% for the coalesced presentation and 23.6% for the split one) are more in line with previous results.

Our tests of DI yield results similar to previous tests by Birnbaum and Chavez (1997) and Birnbaum (2005b). Systematic violations of DI predicted by CPT are not observed in our tests, which fail to confirm predictions of the inverse-S weighting function commonly proposed for rank-dependent models.

Problems	p_{SS}	p_{SR}	p_{RS}	p_{RR}	e	e'	Test
1-3	*0.02*	*0.00*	*0.07*	*0.90*	*0.02*	*0.07*	***7.89****
5-7	*0.47*	*0.00*	**0.28**	*0.26*	*0.15*	*0.13*	***17.42*****
9-11	*0.70*	*0.00*	*0.06*	*0.24*	*0.14*	*0.14*	*1.78*
10-12	*0.82*	*0.14*	*0.00*	*0.04*	*0.09*	*0.12*	***7.83****
13-14	*0.52*	**0.22**	*0.00*	*0.26*	*0.11*	*0.16*	***18.52*****
15-17	*0.29*	*0.00*	*0.13*	*0.57*	*0.11*	*0.15*	***8.04****
19-21	*0.76*	*0.03*	*0.07*	*0.14*	*0.11*	*0.12*	*4.56*
20-22	*0.56*	*0.00*	**0.23**	*0.20*	*0.14*	*0.16*	***17.31*****
29-31	*0.06*	*0.00*	*0.13*	*0.80*	*0.13*	*0.13*	***10.66*****
30-32	*0.18*	**0.40**	*0.00*	*0.42*	*0.18*	*0.18*	***24.72*****
33-35	*0.02*	*0.02*	*0.01*	*0.95*	*0.05*	*0.07*	*3.51*
34-36	*0.05*	*0.13*	*0.01*	*0.81*	*0.15*	*0.08*	*3.76*
38-39	*0.13*	*0.15*	*0.01*	*0.71*	*0.22*	*0.08*	*3.85*
41-42	*0.35*	*0.04*	**0.35**	*0.27*	*0.20*	*0.14*	***15.19*****

Table 5: Tests of Splitting effects

Because violations of coalescing rule out RDU and CPT models and because the tests of independence depend on the form of presentation (split or coalesced), we also provide direct tests of these effects in Table 5. Table 5 compares choices in a given lottery pair in coalesced form with the same choice in canonical split form. If no splitting effects occur, each subject chooses the risky lottery in both problems or the safe lottery in both problems. In contrast, Table 5 shows that many people choose differently even when we control for errors, where e (e') is the estimated error rate of the choice problem stated first (second) in the first column of the table. The last column shows again chi-square tests comparing the fit of a special case model that satisfies coalescing ($p_{RS'} = p_{SR'} = 0$), and the fit of a general model that allows for splitting effects (allowing non-zero probabilities of all four possible response patterns). It turns out that in nine out of 14 tests the null hypothesis, $p_{RS'} = p_{SR'} = 0$, can be rejected in favor of the general model allowing for splitting effects. Except for two cases in the table, the results are consistent with the predictions of the configural weight models with their prior parameters (Birnbaum, 2008); that is, splitting the branch with the best consequence of a gamble tends to make it better and splitting branch with the worst consequence of a gamble makes it worse. In the two cases not predicted by configural weighting using previous parameters (#13-14 and #30-32) it was found that splitting both the upper and lower branches of the risky gamble increased the frequency with which it was chosen relative to splitting the middle branch of the safe gamble.

5 Conclusions

Three major conclusions can be drawn from our study: First, results require us to reject the hypothesis that violations of independence properties tested in the coalesced form can be explained as a mere consequence of choice errors. Even when we control for errors, EU can be significantly rejected in all six tests of common consequence and common ratio effects. In addition, when testing weaker independence conditions we found significant violations of upper tail independence but the other tests were not significant when controlling for errors. Second, violations of coalescing are also robust with respect to our control for errors. In nine of fourteen tests coalescing could be rejected, which refutes EU, RDU, and CPT. Third, splitting effects have a strong impact on tests of independence conditions; when both lotteries in each choice are presented in canonical split form, only one test (a common ratio effect) remains significant in the usual direction, all other tests were either insignificant, unsystematic, or significant in the opposite direction. Additionally, no significant violations of weaker independence conditions were observed when tested in canonical split form.

Given our evidence, the question arises whether one form of presenting lotteries in a choice (coalesced or canonical split form) may be regarded as ”better” in some sense than the other. Indeed, when Savage (1954) found that he violated his sure-thing principle in the common consequence effect, he reformed the Allais choices in canonical split form in order to convince himself to satisfy his own axiom. Violations of first-order stochastic dominance (Birnbaum, 1999; 2008) are nearly eliminated when choices are presented in canonical split form. From these considerations, one might be tempted to conclude that the canonical split form of a choice is the “right” way to present a choice.

This approach cannot save EU, of course, as shown by the significant violations remaining when choices are presented in canonical split form (Table 4). But could it save the RDU and CPT models? The present data show significant violations of UTI in the coalesced form but not in the split form. Furthermore, the weaker independence conditions implied by RDU and CPT are not significantly violated in our data in canonical split form (Table 4). The splitting results of Table 5 violate RDU and CPT, of course, but these results would be considered outside the new, limited domain of these models, which would be considered applicable only to choices in canonical split form. In this approach, these models would no longer be considered as theories of the Allais paradoxes and other results in the literature that have used coalesced lotteries. To account for violations of restricted branch independence in canonical split form (e.g., Birnbaum, 2008), these models would need to give up the inverse-S weighting function.

Another potential problem with this approach arises, however, from the fact that how a lottery is split in canonical form depends on the other lottery with which it is compared. If the utility of a lottery depends on how it is split, and if how it is split depends on the lottery with which it is compared (as is the case in canonical split form), it means that utility of one gamble varies depending on the other lottery. But the RDU and CPT models assume that the utility of a lottery is independent of how it is split and independent of the other lottery of a choice. Violations of these types of independence could result in violations of transitivity (Birnbaum and Schmidt, 2008), which would rule out these models even in the restricted domain of canonical split forms.

Because splitting effects rule out a number of popular models including CPT, and because the operational definition of branch splitting depends on the format for presentation of lotteries (i.e., the mode of representing and displaying probabilities and choices), the question naturally arises, is there a format in which violations of coalescing and other violations of CPT are minimized? So far, 15 different formats have been tested by Birnbaum (2004b, 2006), Birnbaum, Johnson, and Longbottom (2008), and Birnbaum and Martin (2003). Violations have been observed when lotteries are represented by pie charts, histograms, lists of equally likely consequences, aligned and unaligned matrices showing the connections between tickets and prizes, with lotteries presented side-by-side or one above the other, with and without event framing (using same or different colored marbles for common probability-consequence branches), with probabilities of consequences and with decumulative probabilities (to win prize x or more), with positive, negative, and mixed gambles, and with other variations of format and method. Although there are small effects of format, no format has yet been found in which violations of coalescing or stochastic dominance are reduced to levels that might allow retention of RDU or CPT.

Given the present results showing that splitting effects cannot be attributed to error, it seems time to set aside those models that cannot account for these phenomena and concentrate our theoretical and experimental efforts on differentiating models that can account for these phenomena. Models which imply splitting effects include (stripped) original prospect theory (apart from its editing rule of combination), subjectively weighted utility (Edwards, 1954; Karmarkar, 1979), prospective reference theory (Viscusi, 1989), gains-decomposition utility

(Luce, 2000; Marley and Luce, 2001), entropy modified linear weighted utility (Luce, Ng, Marley, and Aczel, 2008a, 2008b), and configural weight models (Birnbaum, 2005a, 2007, 2008).

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46:1497-1512.

Abdellaoui, M. (2009). Rank-dependent utility. In: Anand, P., Pattanaik, P. K., and Puppe, C. (eds.), *The Handbook of Rational and Social Choice*, 90-112. Oxford University Press, Oxford, New York.

Abdellaoui, M., Vossmann, F., and Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science*, 51: 1384-1399.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risqué, critique des postulates et axioms de l'ecole americaine. *Econometrica*, 21:503-546.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997). Does part-whole bias exist? An experimental investigation. *Economic Journal*, 107: 322-332. Berg, J. E., Dickhaut, J. W., & Rietz, T. (2009). Preference Reversals: The Impact of Truth-Revealing Monetary Incentives. Games and Economic Behavior, forthcoming.

Birnbaum, M. H. (2001). Decision and choice: Paradoxes of choice. In Smelser, N. J. and Baltes, P. B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences.* Elsevier Oxford.

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. (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. (2008). New paradoxes of risky decision making. *Psychological Review*, 115: 463-501.

Birnbaum, M. H. and 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., Johnson, K., and Longbottom, J. L. (2008). Tests of cumulative prospect theory with graphical displays of probability. *Judgment and Decision Making, 3*, 528-546.

Birnbaum, M. H. and Martin, T. (2003). Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In S. L. Schneider and

J. Shanteau (Eds.), *Emerging perspectives on decision research* (pp 84-107). New York: Cambridge University Press.

Birnbaum, M. H. and McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes*, 67:91-110.

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

Birnbaum, M. H., Patton, J. N., and Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes*, 77:44-83.

Birnbaum, M. H. and Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty*, 37:77-91.

Birnbaum, M. H. and Schmidt, U. (2009). Testing transitivity in choice under risk. *Theory and Decision*, forthcoming.

Blavatskyy, P. (2006). Violations of betweenness or random errors?. *Economics Letters,* 91: 34-38.

Blavatskyy, P. (2007). Stochastic expected utility theory. *Journal of Risk and Uncertainty*, 34:259-286.

Blavatskyy, P. (2008). Stochastic utility theorem. Journal of Mathematical Economics, 44:1049-1056.

Bleichrodt, H. and Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science*, 46: 1485-1496.

Buschena, D. E. and Zilberman, D. (2000). Generalized expected utility, heteroscedastic error, and path dependence in risky choice. Journal of Risk and Uncertainty, 20:67-88.

Butler, D. J. and Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *American Economic Review*, 97:277-297.

Butler, D. J. and Loomes, G. C. (2009). Imprecision as an Account of Violations of Independence. Working paper, University of East Anglia.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 2:61-104.

Camerer, C. F. and Ho, T. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8:167-196.

Conte, A., Hey, J. D., and Moffat, P. G. (2007). Mixture models of choice under risk. Discussion Paper No. 2007/06, University of York, Department of Economics.

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

Gonzalez, R. and Wu, G. (1999). On the form of the probability weighting function. *Cognitive Psychology*, 38: 129-166.

Green, J. R. and Jullien, B. (1988). Ordinal independence in nonlinear utility theory. Journal of Risk and Uncertainty, 1:355-387.

Gul, F. and Pesendorfer, W. (2006). Random expected utility. *Econometrica*, 74:121-146.

Harrison, G. W. and Rutström, E. (2008). Risk aversion in the laboratory. In: Cox, J. C. and Harrison, G. W. (eds.), *Risk Aversion in Experiments.* Emerald, Bingley (UK).

Harrison, G. W. and Rutström, E. (2009). Expected utility and prospect theory: One wedding and decent funeral. *Experimental Economics*, forthcoming.

Harrless, D. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62:1251-1289.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, 39:633-640.

Hey, J. D., Morone, A., and Schmidt, U. (2007). Noise and bias in eliciting preferences. Discussion Paper No. 2007/04, University of York, Department of Economics.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62:1291-1326.

Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty*, 11:263-274.

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

Karmarkar, U. S. (1979). Subjectively weighted utility and the Allais paradox. Organizational Behavior and Human Performance, 24: 67-72.

Kilka, M. and Weber, M. (2001). What determines the shape of the probability weighting function under uncertainty. *Management Science*, 47: 1712-1726.

Loomes, G. (2005). Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data. *Experimental Economics,* 8:301-23.

Luce, R. D. (1991). Rank- and sign-dependent linear utility models for binary gambles. *Journal of Economic Theory*, 53:75-100.

Luce, R. D. (1992). *Utility of Gains and Losses: Measurement-Theoretic and Experimental Approaches*. Lawrence Erlbaum Asccociation, Mahwah, NJ.

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

Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (2008a). Utility of gambling I: Entropy-modified linear weighted utility. *Economic Theory, 36*: 1-33.

Luce, R. D., Ng, C. T., Marley, A. A. J., & Aczél, J. (2008b). Utility of gambling II: Risk, Paradoxes, and Data. *Economic Theory*, *36*: 165-187.

Marley, A. A. J., and R. D. Luce (2001), Ranked-Weighted Utilities and Qualitative Convolution, Journal of Rsik and Uncertainty 23, 135-163.

McDonald, J. (2009). *Handbook of Biological Statistics*. Baltimore, Maryland: Sparky House Publishing. (2nd edition).

Özdemir, T., and Eyduran, E. (2005). Comparison of Chi-Square and Likelihood Ratio Chi-Square Tests: Power of Test. *Journal of Applied Sciences Research* 1: 242-244, 2005.

Quiggin, J. (1981). Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics*, 25:160-169.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:323-343.

Regenwetter, M. and Stober, C. (2006). Testing transitivity by testing the triangle inequality instead of weak stochastic transitivity. Edwards Bayesian Research Conference, Fullerton, CA, January 2006.

Schmidt, U. (2004). Alternatives to expected utility: Formal theories. In: Barberà, S., Hammond, P.J., and Seidl, C. (eds), *Handbook of Utility Theory*, volume II: Extensions, 757-838. Kluwer Academic Publishers, Dordrecht.

Schmidt, U. and Hey, J. D. (2004). Are preference reversals errors? An experimental investigation. *Journal of Risk and Uncertainty*, 29:207-218.

Schmidt, U. and Neugebauer, T. (2007). Testing Expected Utility in the Presence of Errors. *Economic Journal*, 117:470-485.

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

Starmer, C. (1992). Testing new theories of choice under uncertainty using the common consequence effect. *The Review of Economic Studies*, 59:813-830.

Starmer, C. and Sugden, R. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypothesis. *Annals of Operational Research*, 19:79-102.

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

Sugden, R. (2004). Alternatives to expected utility: Foundations. In: Barberà, S., Hammond, P.J., and Seidl, C. (eds) *Handbook of Utility Theory*, volume II: Extensions, 685-755. Kluwer Academic Publishers, Dordrecht.

Tversky, A. and Fox, C. R. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, 110:585-603.

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

Wakker, P., Erev, I., and Weber, E. (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty*, 9:195-230.

Wakker, P. and Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, 7:147-175.

Weber, M., Eisenführ, F., and von Winterfeldt, D. (1988). The effect of splitting attributes in multiattribute utility measurement. *Management Science*, 34: 431-445.

Wilcox, N. T. (2009). ‘Stochastically more risk averse:’ A contextual theory of stochastic discrete choice under risk, *Journal of Econometrics*, forthcoming.

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

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

Appendix

Problem	No.	p_1 s_1	p_2 s_2	p_3 s_3	p_4 s_4	q_1 r_1	q_2 r_2	q_3 r_3	q_4 r_4
CCE1$_S$	7	*0.80*	*0.10*	*0.10*		*0.80*	*0.10*	*0.10*	
		0	19	19		0	0	44	
	14	*0.40*	*0.10*	*0.10*	*0.40*	*0.40*	*0.10*	*0.10*	*0.40*
		0	19	19	44	0	0	44	44
CCE2$_S$	3	*0.89*	*0.01*	*0.10*		*0.89*	*0.01*	*0.10*	
		0	16	16		0	0	32	
	4	*0.01*	*0.89*	*0.10*		*0.01*	*0.89*	*0.10*	
		16	16	16		0	16	32	
CCE3$_S$	7	*0.80*	*0.10*	*0.10*		*0.80*	*0.10*	*0.10*	
		0	19	19		0	0	44	
	8	*0.10*	*0.80*	*0.10*		*0.10*	*0.80*	*0.10*	
		19	19	19		0	19	44	
CCE4$_S$	11	*0.70*	*0.10*	*0.10*	*0.10*	*0.70*	*0.10*	*0.10*	*0.10*
		0	21	21	21	0	0	21	42
	12	*0.70*	*0.10*	*0.10*	*0.10*	*0.70*	*0.10*	*0.10*	*0.10*
		0	21	21	42	0	0	42	42
CRE1$_S$	17	*0.98*	*0.01*	*0.01*		*0.98*	*0.01*	*0.01*	
		0	23	23		0	0	46	
	18	*0.50*	*0.50*			*0.50*	*0.50*		
		23	23			0	46		
CRE2$_S$	21	*0.80*	*0.06*	*0.14*		*0.80*	*0.06*	*0.14*	
		0	28	28		0	0	45	
	22	*0.40*	*0.18*	*0.42*		*0.40*	*0.18*	*0.42*	
		0	28	28		0	0	45	
UTI$_S$	31	*0.73*	*0.01*	*0.01*	*0.25*	*0.73*	*0.01*	*0.01*	*0.25*
		0	15	15	60	0	0	33	60
	32	*0.73*	*0.01*	*0.01*	*0.25*	*0.73*	*0.01*	*0.01*	*0.25*
		0	15	15	33	0	0	33	33
LTI$_S$	35	*0.75*	*0.23*	*0.01*	*0.01*	*0.75*	*0.23*	*0.01*	*0.01*
		1	34	36	36	1	33	33	60
	36	*0.75*	*0.23*	*0.01*	*0.01*	*0.75*	*0.23*	*0.01*	*0.01*
		33	34	36	36	33	33	33	60
UCI$_S$	37	*0.20*	*0.20*	*0.60*		*0.20*	*0.20*	*0.60*	
		9	10	24		3	21	24	
	39	*0.20*	*0.20*	*0.60*		*0.20*	*0.20*	*0.60*	
		9	9	21		3	21	21	

Table note: The first lottery pair of a choice problem always characterizes the lotteries *S* and *R* and the second one the lotteries *S′* and *R′*.

Table A1: The lottery pairs in canonical split form