Testing independence conditions in the presence of errors and splitting effects

Michael H. Birnbaum,

AQ1

1,*

Phone 657-278-2102
Email mbirnbaum@Exchange.FULLERTON.EDU

Ulrich Schmidt, 2,*,3

Phone +49 431 8801400
Email us@bwl.uni-kiel.de

Miriam D. Schneider, 2

1 Department of Psychology, California State University (CSUF), Fullerton, CA, 92834-6846 USA

2 Department of Economics, University of Kiel, Olshausenstr, 40, 24098 Kiel, Germany

3 Kiel Institute for the World Economy, Kiel, Germany

Abstract

This paper presents experimental tests of several independence conditions implied by expected utility and alternative models. We perform repeated choice experiments 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 many previous studies) but also in split forms. It turns out previously reported violations of independence and splitting effects

remain significant even when controlling for errors. However, splitting effects have a substantial influence on 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 properties tested could be reduced to insignificance or even reversed.

Keywords

Coalescing
Errors
Expected utility
Independence axiom
Prospect theory
Risky decision making
Splitting effects

JEL classifications

C91
D81

Electronic supplementary material

The online version of this article (doi: 10.1007/s11166-017-9251-5) contains supplementary material, which is available to authorized users.

1. Introduction

Many studies have concluded that expected utility (EU) theory fails to provide an accurate description of people's behavior in several choice problems under risk. Choices apparently violate the crucial independence axiom as shown by the famous paradoxes of Allais (1953). Such violations led to 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 account of actual choice behavior (see surveys by Abdellaoui 2009; Birnbaum 2008; Luce

2000; Starmer 2000; Sugden 2004; Schmidt 2004). Most of these new theories rely on independence assumptions that are weaker than assumptions of EU. Experimental tests of these weakened independence conditions revealed systematic violations, which rule out not only EU but also rank-dependent utility models and cumulative prospect theory (Birnbaum 2008; Birnbaum and Navarrete 1998; Wakker et al. 1994; Wu 1994).

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. Nowadays one very intensively discussed issue in decision theory is how to model this stochastic component adequately (e.g., Birnbaum and Bahra 2012; Gul and Pesendorfer 2006; Blavatskyy 2007, 2008, 2011, 2012; Conte et al. 2011; Hey et al. 200~~7~~9; Loomes, 2005; Wilcox 2008, 2011; Harrison and Rutström 2009).

AQ2

Is it possible that one can explain systematic violations of EU by proper modeling of the error component of choice? 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." Certain of the reported violations of EU might be at least partly caused by errors instead of being intrinsic violations (Blavatskyy 2006; Sopher and Gigliotti 1993; Schmidt and Hey 2004; Butler and Loomes 2007, 20~~09~~11; Berg et al. 2010).

AQ3

Schmidt and Neugebauer (2007) considered only cases where subjects chose the same option in a choice problem three times in row. Provided that the probability of errors is not too high, such repeated choices more likely reflect "true preferences" since it is rather improbable that a subject makes the same error three

times. It turns out that in these cases the incidence of violations of independence decreases substantially.

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 repeated choice experiments and fit an error model that is neutral with respect to violations of any independence condition. This model allows us to distinguish 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 assumes that true preferences are governed 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 (which can be different for each choice problem); it does not assume transitivity, independence, or coalescing (implications of EU that we aim to test).

A further systematic deviation from EU (and in fact also from many alternatives to EU such as rank-dependent utility (RDU), rank- and sign-dependent utility (RSDU), and cumulative prospect theory (CPT)) is provided by violations of coalescing (splitting effects). 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. A splitting effect is said to occur if, for example, G is preferred to a gamble, F, and yet F is preferred to the gamble G', apart from random error.

There exists abundant evidence that splitting a branch (e.g., an event) with a good consequence increases the attractiveness of that lottery in comparison to other lotteries (Starmer and Sugden 1993; Birnbaum and Navarrete 1998; Humphrey 1995, 2001).[1] 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 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. It may well be the case that splitting effects are the main cause of violations of independence conditions, as in the Allais paradoxes (Birnbaum 2004).

Birnbaum (2004) noted that the Allais common consequence problem can be decomposed into three simpler properties: coalescing (no splitting effects), transitivity of preference, and restricted branch independence. If a person satisfies these three properties, she would not show the paradoxical violation of EU. Restricted branch independence is a weaker form of independence that holds when the number of branches is the same in both gambles being compared and the probabilities of corresponding branches are equal. With these restrictions, the value of a common consequence is assumed to have no effect. For example, with three branch gambles, restricted branch independence requires that $S = (x, p; y, q; z, 1 - p - q)$ preferred to $R = (x', p; y', q; z, 1 - p - q)$ if and only if $S' = (x, p; y, q; z', 1 - p - q)$ preferred to $R' = (x', p; y', q; z', 1 - p - q)$. Note that the common consequence is z in the first choice and z' in the second choice. Presenting Allais common consequence problems in a proper split form can convert them to tests of restricted branch independence.

For example, consider the choice between S^* = (\$40, 0.2; \$2, 0.8) and R^* = (\$98, 0.1; \$2, 0.9). The *canonical split form* of this choice is the form in which probabilities on corresponding ranked branches are equal and the number of branches is minimal: S = (\$40, 0.1; \$40, 0.1; \$2, 0.8) versus R = (\$98, 0.1; \$2, 0.1; \$2, 0.8). Coalescing and transitivity imply that a person should choose S over R iff she chooses S^* over R^*. According to restricted branch independence, a person chooses S over R iff she chooses S' = (\$40, 0.1; \$40, 0.1; \$98, 0.8) over R' = (\$98, 0.1; \$2, 0.1; \$98, 0.8). By coalescing and transitivity, this holds iff (\$40, 0.2; \$98, 0.8) is chosen over (\$98, 0.9; \$2, 0.1). CPT satisfies coalescing (Birnbaum and Navarrete 1998), so we can test not only EU but also CPT by testing such pairs of choices in both coalesced and canonical split form.

This paper is organized as follows. The next section presents error models and

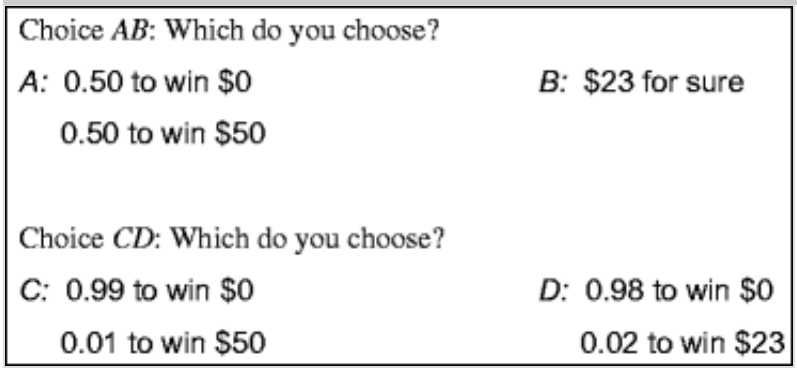
discusses the issue of testing properties such as independence in the presence of errors. Section 3 presents the experimental design and method of Experiment 1, while Section 4 reports the results. Sections 5, 6, and 7 present Experiment 2, which confirms and extends our results. Discussion and concluding observations appear in Section 8.

2. Errors and violations of independence

This section shows that random errors can generate data that appear to provide systematic violations of EU. It shows how the true and error model can be used to separate estimation of error from the model to be tested. Consider a simple variant of the common ratio effect (CRE) as shown in Fig. 1. ~~(Fig. 1).~~

Fig. 1

A Test of Common Ratio Independence



AQ4

According to EU theory, a person should prefer *A* over *B* if and only if that person prefers *C* over *D* because EU(*A*) > EU(*B*) if and only if EU(*C*) > EU(*D*). Proof: EU(A) = 0.5*u(*50*)* + 0.5*u*(0) > *u*(20) iff 0.01*u(*50*)* + 0.99*u(0)* > 0.02*u(*20*)* + 0.98*u*(0). There are four possible response patterns in this experiment, *AC*, *AD*, *BC*, and *BD*, where e.g. *AC* represents preference for *A* in the first choice and *C* in the second choice. The response patterns *AC* and *BD* are consistent with EU while the other two patterns violate the independence axiom of EU. Suppose we obtained data as ~~follows~~in Table 1 -f rom 100 participants: Insert Table 1 after this sentence, if possible.

In this case 33 people switched from B to C, whereas only 9 reversed preferences in the opposite pattern. The conventional statistical test (test of correlated proportions) is significant, $z = 3.7$, which is usually taken as evidence that EU theory can be rejected (e.g., Conlisk 1989).

Can such results occur by random errors? As has been previously noted, yes. Suppose for example a subject chooses A over B iff $EU(A) - EU(B) + \varepsilon > 0$ and chooses C and D iff $EU(C) - EU(D) + \varepsilon' > 0$ where ε and ε' are normally distributed, independent random variables with $E(\varepsilon) = E(\varepsilon') = 0$ and variances σ and σ'. Note that $EU(A) – EU(B) > EU(C) – EU(D)$ for any utility function, u. Depending on the error variances, the probability to make an erroneous response in the choice between C and D could be greater than that in the choice between A and B (for example, if $\sigma = \sigma'$), in which case we would observe more frequent erroneous BC than AD preference patterns. Thus, the standard test of inequality of different types of violations is not a diagnostic test of EU plus random error.

Since this EU-based error theory, however, assumes EU, it does not allow us to test EU; it merely provides an excuse for the violations. Therefore, in order to test EU empirically (rather than by assumption) we need more general error models that allow us to empirically estimate error rates and to distinguish true preferences from response errors. Such models have been developed (Birnbaum 2013; Birnbaum and Bahra 2012).

Suppose that each person has a true preference pattern, which may be one of the four possible response combinations. Let $p_{AC'}$, $p_{AD'}$, $p_{BC'}$, and $p_{BD'}$ denote the probabilities of the four true preference patterns, which represent the relative frequencies of subjects with these patterns. Let e represent the probability of an error in reporting one's true preference for the choice between A and B. Similarly, e' is the probability of an error for the choice between C and D. Is it possible that, given the data in Table 1, all subjects conform to EU? In other words, are the data in Table 1 compatible with $p_{BC'} = p_{AD} = 0$? The answer is "yes," despite the statistically significant inequality in the two types of violations.

Table 1

Hypothetical Frequencies of Response Patterns

	C	***D***
A	9	9
B	33	49

In the true and error (TE) model, the probability of observing the preference pattern *BC* is given as follows:

$$P(BC) = p_{AC}(e)\left(1 - e'\right) + p_{AD}(e)\left(e'\right) + p_{BC}\,(1 - e)\left(1 - e'\right) + p_{BD}\,(1 - e)\left(e'\right) \quad 1$$

AQ5

In this expression, *P(BC)* is the probability of observing this preference pattern, and *e* and *e′* are the error rates in the choice problems *A* versus *B* and *C* versus *D,* respectively. The error probabilities are assumed to be less than ½ and to be mutually independent. *P(BC)* is the sum of four terms, each representing the probability of having one of the true patterns (p_{AC}, p_{AD}, p_{BC}, and p_{BD}), and having the appropriate pattern of errors and correct responses to produce the *BC* data pattern. For example, a person who truly has the *AC* pattern could produce the *BC* pattern by making an error on the first choice and correctly reporting her preference 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, which have three degrees of freedom (df) 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 reproduce the data in Table 1 perfectly are shown in Table 2:

Table 2

Two models of Table 1 that fit equally well

Parameter	Model 1: EU holds	Model 2: EU does not hold
$p_{AC'}$	0.10	0.05
p_{AD}	0.00	0.00
$p_{BC'}$	0.00	0.34
$p_{BD'}$	0.90	0.61
e	0.10	0.14
e'	0.40	0.14

This table shows that we might try to "save" EU in this case by assuming that people have unequal error rates in the two choice problems. But if the error rates are actually equal, we should reject EU. Thus, given only the data in Table 1 it is not possible to rationally decide one way or the other. As noted in Wilcox (2008), none of the early error models he reviewed allowed any independent way to answer this question without making an assumption that affects the conclusion. We need a way to evaluate EU 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 error rates empirically. This enrichment is possible via replications of each choice problem (Birnbaum and Bahra 2012).

Consider the case of one choice problem presented twice: for example, Choice *AB* above. With two replicates, there are four response patterns possible, *AA*, *AB*, *BA*, and *BB*. The probability that a person will show the *AB* preference reversal is given as follows:

$$P(AB) = p\,(1-e)\,(e) + (1-p)\,e\,(1-e) = e\,(1-e)\,. \qquad 2$$

where p is the true probability of preferring A and e is the error rate on the choice between A and B. The probability of the opposite reversal, *BA*, is also $e(1-e)$. These expressions show that if we present each choice problem at least twice to the same people, we can estimate the error rates for each choice problem without having to assume that error rates are equal or that EU is true. An important property of the true and error models is that with the proper experimental designs,

they not only allow estimation of parameters to address empirical questions, but they are empirically testable.

With two replications of two choice problems (e.g., in a test of the common consequence effect), there are $2 \times 2 \times 2 \times 2 = 16$ possible response patterns, which have 15 degrees of freedom. But there are only 5 parameters to estimate from the 16 frequencies of these response patterns: two error terms and three probabilities of true response patterns. Because the four probabilities of true response patterns sum to 1, the fourth probability is determined. The general model (which allows all four true preference patterns to have nonzero probability) is now over-determined, leaving $15 - 5 = 10$ degrees of freedom to test the general error model. EU theory is a special case of this general model in which two of the true probabilities are fixed to zero. For example, in Table 1, EU assumes $p_{BC'} = p_{AD} = 0$.

Parameters can be estimated to minimize the deviations of fit of the predicted to observed frequencies of response patterns, which can be measured by the standard Chi-Square formula:

$$\chi^2 = \sum (f_i - q_i)^2 / q_i, \qquad 3$$

where f_i is the observed frequency and q_i is the predicted frequency of a particular response pattern. Two statistical tests are of primary importance: first, one can test the TE model as a fit to the 16 response frequencies. Second, one can test EU as a special case of that model. The difference in χ^2 between a fit of the general model (that allows all four response patterns to have non-zero probabilities) and the χ^2 for the special case (EU, in which $p_{BC} = p_{SRAD'} = 0$) is theoretically Chi-square distributed with 2 degrees of freedom. This test allows us to determine whether observed violations of EU are real, or whether they might be attributed to response errors.

In Experiment 1, we used four replications of each choice problem. With two choice problems and four replications, there are 256 possible response patterns ($4^4 = 256$). When cell frequencies are small, we use the *G*-statistic, which is a likelihood ratio test that takes on similar values to the standard index and is

theoretically distributed as Chi-Square: $G = 2\sum f_i ln(f_i/q_i)$. The G index is regarded as more accurate when cell frequencies are small; cells with zero frequency (i.e., $f_i = 0$) have no effect on the statistic, whereas the standard Chi-Square is known to be inflated when expected frequencies are low (see Özdemir and Eyduran 2005; McDonald 2009). Additional details about true and error models (including more elaborate models and analyses) are presented in our online supplement in Appendices B and C.

3. Experimental method and Design of Experiment 1

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 with nine participants and lasting about 90 min. 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 min, i.e. 12.76 Euro per hour, which clearly exceeds the usual wage of students (about 8 Euro per hour). As noted in Cox et al. (2015), this system is incentive compatible with EU theory; that is, a person who perfectly satisfies EU should report her true preferences on all choice problems.

Choices were presented as in Fig. 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 holding numbered tickets from 1 to 100. Suppose a subject could for instance play out lottery A in Fig. 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.

Fig. 2

Format for presentation of a choice between lotteries

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*

Choice problems in the booklets were presented in pseudo-random order. The ordering 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. The lotteries were adapted from previous studies that reported high violation rates but we 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. The canonical split forms of these choices can be found in Appendix A in the online supplement to this article. For example, Choice 5 in Table 3 is the choice between S = (\$19, 0.2; \$0, 0.8) and R = (\$44, 0.1; \$0, 0.9). The canonical split form of this choice is the form in which probabilities on corresponding ranked branches are equal and the number of branches is minimal: S^* = (\$19, 0.1; \$19, 0.1; \$0, 0.8) versus R^* = (\$44, 0.1; \$0, 0.1; \$0, 0.8), listed as Choice 7 in Table A.1 of Appendix A. Objectively, S is the same as S^* and R is the

same as R^*.

Table 3

The lottery pairs

Property	No.	Safe Gamble			Risky Gamble		
		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 *SR*	5	0.80	0.20		0.90	0.10	
		0	19		0	44	
S′R′	13	0.40	0.20	0.40	0.50	0.50	
		0	19	44	0	44	
CCE2 *SR*	1	0.89	0.11		0.90	0.10	
		0	16		0	32	
S′R′	2	1.00			0.01	0.89	0.10
		16			0	16	32
CCE3 *SR*	5	0.80	0.20		0.90	0.10	
		0	19		0	44	
S′R′	6	1.00			0.10	0.80	0.10
		19			0	19	44
CCE4 *SR*	9	0.70	0.30		0.80	0.10	0.10
		0	21		0	21	42
S′R′	10	0.70	0.20	0.10	0.80	0.20	
		0	21	42	0	42	
CRE1 *SR*	15	0.98	0.02		0.99	0.01	
		0	23		0	46	
S′R′	16	1.00			0.50	0.50	

		23			0	46	
CRE2 *SR*	20	0.80	0.20		0.86	0.14	
		0	28		0	44	
S′R′	19	0.40	0.60		0.58	0.42	
		0	28		0	44	
UTI *SR*	29	0.73	0.02	0.25	0.74	0.01	0.25
		0	15	60	0	33	60
S′R′	30	0.73	0.02	0.25	0.74	0.26	
		0	15	33	0	33	
LTI *SR*	33	0.75	0.23	0.02	0.75	0.24	0.01
		1	34	36	1	33	60
S′R′	34	0.75	0.23	0.02	0.99	0.01	
		33	34	36	33	60	
UCI *SR*	37	0.20	0.20	0.60	0.20	0.20	0.60
		9	10	24	3	21	24
S′R′	38	0.40	0.60		0.20	0.80	
		9	21		3	21	
LDI *SR*	23	0.60	0.20	0.20	0.60	0.20	0.20
		1	18	19	1	2	32
S′R′	24	0.10	0.45	0.45	0.10	0.45	0.45
		1	18	19	1	2	32
UDI *SR*	25	0.20	0.20	0.60	0.20	0.20	0.60
		6	7	20	1	19	20
S′R′	26	0.45	0.45	0.10	0.45	0.45	0.10
		6	7	20	1	19	20

The first lottery pair of a choice problem always characterizes the lotteries S and R

and the second one the lotteries S' and R'

The first six tests in Table 3 include four problems that had previously shown common consequence effects (CCE1–4) and two problems with common ratio effects (CRE1 and 2). The paradoxes of Allais are special variants of such CCE and 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 $R' = (z, q_1 - \alpha; z', \alpha; r_2, q_2; r_3, q_3)$ where 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 of this design, there are no more than three distinct consequences in each test and lotteries are presented in coalesced form.

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 adapted from Starmer (1992). 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.

The remaining five independence properties in Table 3 are weaker than the independence axiom of EU. Some of these 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). These properties test between the class of RDU models (including CPT) and earlier configural weight models that violate those properties (Birnbaum and Stegner 1979; Birnbaum 2008). Tail independence (TI), studied by Wu (1994), 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 is 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. Wu (1994) and Birnbaum (2008) reported large violations of TI, contradicting RDU, CPT, and EU. Our tests of UTI were adapted from Wu (1994). LTI has, as far as we know, not been tested before. Our test of LTI was constructed in similar fashion to that of UTI.

Birnbaum and Navarrete (1998) deduced the properties of lower and upper cumulative independence (LCI and UCI) as theorems that must be satisfied by CPT (and RSDU) but which will be violated according to configural weight models. 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 were reported by Birnbaum and Navarrete (1998) and in subsequent research summarized in Birnbaum (2008).

The final property is distribution independence (DI), proposed by Birnbaum as a test of the weighting function in CPT. Whereas certain configural weight models

imply that DI holds, DI should be violated according to RDU and CPT, if the weighting function is not linear, as commonly inferred from 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 et al. 2005). For three-outcome lotteries, DI demands that $S = (s_1, \beta; s_2, \beta; \alpha, 1 - 2\beta) \succ R = (r_1, \beta; r_2, \beta; \alpha, 1 - 2\beta)$ if and only if $S' = (s_1, \delta; s_2, \delta; \alpha, 1 - 2\delta) \succ R' = (r_1, \delta; r_2, \delta; \alpha, 1 - 2\delta)$, 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).

Consistency check: We examined the eight tests of transparent dominance per person. Out of 54 subjects, 49 (91%) had no violations in eight tests; five had one and one person had two violations; a total of 7 violations out of 8 × 54 = 432 tests is 1.6%). We conclude from these checks that our subjects were indeed motivated and attentive.

4. Results of experiment 1

Table 4 summarizes tests of independence properties, using the true and error model. For each property tested (first column), Table 4 shows the estimated probabilities of the four possible response patterns in columns two through five. Columns six and seven show the estimated error rates for the choice problems between S and R *(e)* as well as between S' and R' *(e')*. The final column presents 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 that 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. A subscript "S" in the first column indicates that in that test, the choice problems were presented in canonical split form. Bold entries in Table 4 indicate substantial violations of the property tested.

Table 4

Tests of independence conditions

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**** AQ6
$CCE1_S$	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
*denotes a significance level of 5%, ** a significance level of 1%								

According to the independence axiom of EU, $p_{RS'} = p_{SR'} = 0$. Our results show that even allowing for a different error rate in each choice problem, EU is systematically violated in the manner reported in previous research. In tests CCE1–4 and CRE1–2, independence can be rejected. Estimated probabilities of violations range from 10% to 64% (mean 28%); apart from CCE3, the vast majority of violations are of one type. In CCE3, there are substantial violations of both pattern *SR′* and of *RS′*. In summary, we find that typical violations of the independence properties of EU reported for raw responses are also observed when errors are taken into account; in other words, we conclude that the violations observed in previous research are replicated in our experiment and are not attributable to random error.

A quite different picture arises when the same choice problems are presented in canonical split form. In $CCE4_S$ and $CRE2_S$ violations are reduced and not significant (i.e., EU cannot be rejected), and in $CCE1_S$ the violations of EU are in the opposite direction from that observed in coalesced form (i.e., the *SR′* pattern is more frequent than the *RS′* pattern), and these reversed violations are statistically significant. One case (CRE1) shows the same pattern in both split and coalesced forms as in previous research. We can, therefore, conclude that splitting effects have a substantial influence on tests of the independence axiom of EU. These results are consistent with Birnbaum's (2004) hypothesis that CCE are largely due to violations of coalescing.

Next, consider the tests of the weaker independence conditions implied or assumed by RDU, CPT, and EU. For UTI we observe a substantial and systematic violation: the estimated probability of the violating pattern *RS′* amounts to 52%. These results agree with the high violation rates observed by Wu (1994) and other research summarized in Birnbaum (2008). We conclude that violations of UTI are not caused by errors but reflect true preferences. This is a serious challenge for CPT and the whole class of rank-dependent models (including EU) which all imply that UTI holds. It is, however, noteworthy that violations of UTI virtually disappear when we present choices in 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 UTI seem to be likely caused by violations of coalescing.

Our new test of LTI did not generate significant violations, nor did the test of UCI. Comparing our results for UCI to previous ones, the failure to observe significant violations may be due to large error rates in our tests. In fact, the estimated error rates in our coalesced test of UCI are the highest of all choice problems. 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 24% for the split one) fall in line with previous results.

Our tests of LDI and UDI agree with the conclusions of previous research. Systematic violations of DI are predicted by CPT but are not observed in our tests, replicating previous failures to confirm predictions of the inverse-S weighting function commonly proposed for rank-dependent models.

Because violations of coalescing rule out RDU and CPT models (as well as EU) and because the tests of independence depend on the form of presentation (split or coalesced), we also present direct tests of these splitting 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, aside from error. In contrast, Table 5 shows that many people make different choice responses, even when corrected 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 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 of coalescing, $p_{RS'} = p_{SR'} = 0$, can be rejected in favor of the general model allowing for splitting effects. ~~V~~Substantial violations of coalescing (indicated in bold in Table 5) ~~iolations of coalescing~~ contradict CPT, RDU, and EU models, among others.

Table 5

Tests of Splitting effects

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****

5. Experiment 2

Experiment 1 concluded that violations of common ratio and common consequence independence cannot be attributed to error and are strongly affected by the form in which choice problems are presented. When branches are split in canonical split form, violations of independence are either reduced markedly or in some cases reversed. Such results are largely consistent with configural weight models (Birnbaum 2008), which can violate coalescing. In Experiment 2, we replicate and extend these findings for Allais paradoxes in a new experiment with American participants with different lotteries, and we test a new manipulation, called *double splitting* in which both the upper and lower branches are split. This manipulation

allows us to test among variations of configural weighting models.

Birnbaum and Stegner (1979), Eq. 10) proposed a "revised" configural weight model in which configural weights are transferred from branch to branch in proportion to the absolute weight of the branch that loses weight. This model is now known as the transfer of attention exchange (TAX) model (e.g., Birnbaum and Navarrete 1998). Consider gambles defined as $G = (x_1, p_1; x_2, p_2; \ldots x_j, p_j; \ldots x_i, p_i; \ldots; x_n, p_n)$, where the outcomes are ordered such that $x_1 \geq x_2 \geq \ldots \geq x_j \geq \ldots \geq x_i \geq \ldots \geq x_n \geq 0$ and $\Sigma\, p_j = 1$. Note that $x_j \geq x_i$. The TAX model can be written as follows:

$$U(G) = \frac{\sum_{i=1}^{n} t(p_i)u(x_i) + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left[u(x_i) - u(x_j)\right] \omega\left(p_i, p_j, n\right)}{\sum_{i=1}^{n} t(p_i)} \quad (4)$$

where $U(G)$ is the utility of the gamble, $t(p)$ is a weighting function of probability, $u(x)$ is the utility function of money, and $\omega(p_i, p_j, n)$ is the configural transfer of weight between outcomes x_i and x_j in a gamble with n branches. Note that if the configural transfer is positive, then the branch with a higher valued consequence transfers weight to branches with lower valued consequences; when it is negative, then the higher-valued branch gains weight and the lower valued branches give up this same weight. The weight transfers are assumed as follows:

$$\omega(p_i, p_j, n) = \begin{cases} \delta t(p_j)/(n+k), & \delta \geq 0 \\ \delta t(p_i)/(n+k), & \delta < 0 \end{cases} \quad (5)$$

where n is the number of distinct branches, and k was fixed to 1. If $\delta = 1$ and $k = 1$ then the weights of two, three, and four equally likely outcomes will be (2/3, 1/3), (3/6, 2/6, 1/6), and (0.4, 0.3, 0.2, 0.1), for lowest to highest valued branches, respectively. For small consequences, $0 < x < \$150$, it has been found that $u(x) = x$ provides a reasonable approximation to choice data from undergraduates. In addition the weighting function of probability has been approximated as $t(p) = p^{0.7}$. This model is known as the "prior" TAX model because these parameters from 1997 were used (with considerable success) to predict results of new experiments, such as those reviewed in Birnbaum (2004, 2008).

Viscusi's (1989) prospective reference theory is a special case of TAX in which $\delta = 0$; EU is also a special case of TAX when $t(p) = p$ and $\delta = 0$.

This model is the same as in Birnbaum and Navarrete (1998), except that the consequences are ranked here in decreasing value (e.g., $x_1 \geq x_2$, to match the notational convention used in CPT), so $\delta < 0$ in the earlier publication corresponds to $\delta > 0$ in Eq. 5.

According to the TAX model, splitting the branch of a gamble leading to the highest consequence in the gamble should make the gamble better and splitting the lowest branch of a gamble should make a gamble worse (Birnbaum 2008).

Except for two cases, the trends observed in Experiment 1 (Table 5) are consistent with the predictions of the prior TAX model; that is, splitting the branch with the best consequence of a gamble tends to make it more attractive and splitting the branch with the worst consequence of a gamble makes it worse. In the two cases not predicted from previous parameters (Choice Problems 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 the split form was chosen relative to splitting the middle branch of the safe gamble.

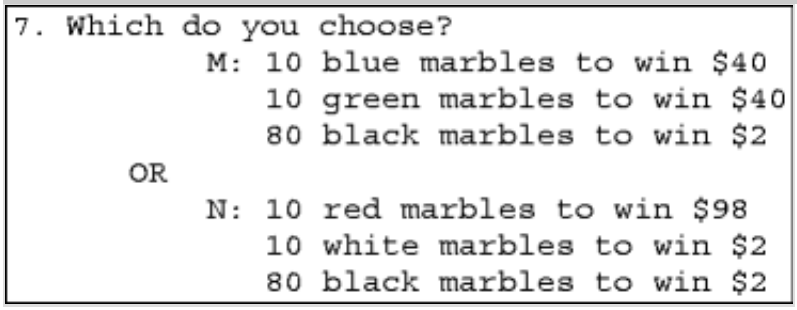
Double splitting: Consider, for example, A = ($98, 0.5; $2, 0.5) and B = ($98, 0.25; $98, 0.25; $2, 0.25; $2, 0.25). B is the same as A except for coalescing, and the manipulation from A to B is called "double splitting" because both upper and lower branches are split in the same way. To test the effect of this manipulation, each of these gambles can be compared with a third gamble, C. Will A be chosen over C more often than B over C, when corrected for error? According to the prior TAX model, the overall utility of a gamble tends to decrease as the number of branches increases, consistent with the idea that increasing complexity of an alternative lowers its evaluation (Sonsino et al. 2002). This manipulation allows us to compare three variants of the TAX model in which the term $(n + 1)$ in Eq. (5) is replaced by $(n + k)$, where $k > 0$, $k < 0$ (where $n + k > 0$), or $k = 0$; these cases imply that double splitting will tend to decrease, increase, or have no effect on the utility of a gamble, respectively.

6. Method of experiment 2

Participants made choices between gambles, viewed via computers in a lab. Each choice appeared as in Fig. 3.

Fig. 3

Format for the display of a choice problem in Experiment 2

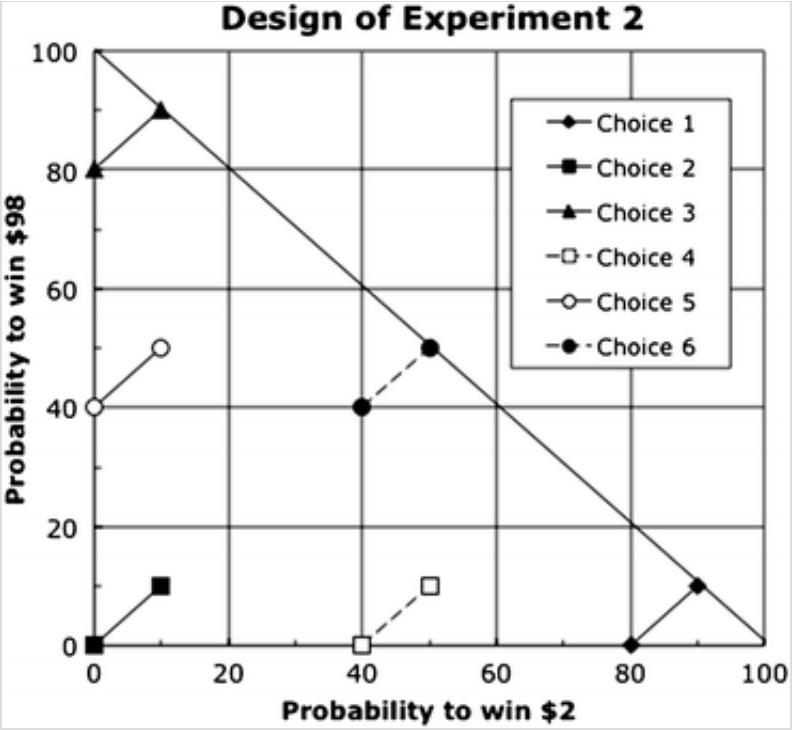


Instructions read (in part) as follows: “Your prize will be determined by the color of marble that is drawn randomly from an urn. The urns always have exactly 100 marbles, so the number of any given color tells you the percentage that wins a given prize.” Participants clicked a button beside the gamble they would rather play in each trial. They were informed that 3 lucky participants would be selected at random to play one of their chosen gambles for real money, so they should choose carefully. Prizes were awarded as promised.

The main design analyzed the Allais paradoxes as the result of violations of restricted branch independence, violations of coalescing, and random error. Figure 4 shows the main design of Experiment 2. The triangle in Fig. 4 represents a probability simplex depicting lotteries of the form ($98, p; $40, $1 - p - q$; $2, q). The abscissa shows q, the probability to win $2 and the ordinate shows p, the probability to win $98. “Sure things” to win $40, $98, or $2 correspond to the lower left-, upper left-, and lower right-corners of this figure, respectively.

Fig. 4

Main design of Experiment 2 (see also Table 6)



Each line segment in Fig. 4 represents a choice between a “safe” gamble and a “risky” gamble. Each of these six choice problems was presented in both coalesced and in canonical split form, making 12 choice problems in the main design. In Sample 2, the same design was used but consequences $98, $40, and $2 were changed to $96, $48, and $4, respectively.

Table 6 lists the choice problems of Fig. 4. Each choice problem in Table 6 is created from the problem in the row above by either coalescing/splitting or by applying restricted branch independence. A person should make the same choice response in every choice problem of Table 6, according to EU, aside from random error. That is, a person should either choose the risky gamble in all 13 choice problems or choose the safe gamble in all 13 cases. According to CPT, people should make the same choice responses in problems that differ only in coalescing/splitting, but they can violate restricted branch independence.

Table 6

Choice problems used in main design of Experiment 2 (Sample 1) Note: Problem Number refers to Fig. 4

		%*R*	%*R*	True &	True &

Problem No.	Choice Problem		Sample 1	Sample 2	Error Model Sample 1		Error Model Sample 2	
	S	***R***	***n* = 104**	***n* = 107**	***p***	***e***	***p***	***e***
1	($40, 0.2; $2, 0.8)	($98, 0.1; $2, 0.9)	59	66	0.64	0.17	0.72	0.13
1a	($40, 0.1; $40, 0.1; $2, 0.8)	($98, 0.1; $2, 0.1; $2, 0.8)	42	40	0.40	0.11	0.37	0.11
2	$40 for sure	($98, 0.1; $40, 0.8; $2, 0.1)	57	63	0.60	0.13	0.65	0.08
2a	($40, 0.1; $40, 0.8; $40, 0.1)	($98, 0.1; $40, 0.8; $2, 0.1)	59	56	0.61	0.09	0.57	0.11
3	($98, 0.8; $40, 0.2)	($98, 0.9; $2, 0.1)	25	17	0.16	0.15	0.09	0.10
3a	($98, 0.8; $40, 0.1; $40, 0.1)	($98, 0.8; $98, 0.1; $2, 0.1)	60	47	0.69	0.23	0.46	0.14
4	($40, 0.6; $2, 0.4)	($98, 0.1; $40, 0.4; $2, 0.5)	73	74	0.80	0.13	0.88	0.19
4a	($40, 0.1; $40, 0.4; $40, 0.1; $2, 0.4)	($98, 0.1; $40, 0.4; $2, 0.1; $2, 0.4)	42	46	0.38	0.19	0.44	0.13
5	($98, 0.4; $40, 0.6)	($98, 0.5; $40, 0.4; $2, 0.1)	46	41	0.43	0.17	0.35	0.19
5a	($98, 0.4; $40, 0.1; $40, 0.4; $40, 0.1)	($98, 0.4; $98, 0.1; $40, 0.4; $2, 0.1)	46	40	0.42	0.21	0.38	0.09
6	($98, 0.4; $40, 0.2; $2, 0.4)	($98, 0.5; $2, 0.5)	21	24	0.09	0.15	0.13	0.15
6a	($98, 0.4; $40, 0.1;	($98, 0.4; $98, 0.1;	40	24	0.24	0.19	0.30	0.10

6a	$40, 0.1; $2, 0.4)	$2, 0.1; $2, 0.4)	40	34	0.34	0.19	0.30	0.10
6c	($98, 0.1; $98, 0.3; $40, 0.2; $2, 0.4)	($98, 0.5; $2, 0.3; $2, 0.1; $2, 0.1)	23	20	0.18	0.08	0.09	0.13

In choice problem 1 of Table 6 (first row), the common branch is 80 marbles to win $2 (Problem 1), 80 to win $40 (Problem 2), or 80 to win $98 (Problem 3). Problem 4 is generated from Problem 1 by changing 40 marbles to win $2 to 40 marbles to win $40; Problem 5 is generated from Problem 4 by changing the common branch of 40 marbles to win $2 to 40 marbles to win $98. Problem 6 is generated from Problem 1 by changing a common branch of 40 marbles to win $2 to 40 marbles to win $98.

Choice problems marked #1a, 2a, 3a, etc. are the same as #1, 2, 3, etc., respectively, except these are presented in canonical split form. For example, Problem 1a is the same as Problem 1, except that the branch of 20 marbles to win $40 in the safe gamble has been split to two branches of 10 marbles each to win $40, and the branch of 90 marbles to win $2 in the risky gamble has been split into a branch with 80 marbles to win $2 and 10 to win $2. Problem 6c is another split form of Problem 6, designed to improve the “safe” gamble relative to the “risky” one. Problems 1, 1a, 2, 3, and 3a (Sample 1) match choice problems used by Birnbaum (2004, Series A).

There were 13 additional choice problems, shown in Table 7, which were presented on alternating trials. These additional trials tested double splitting of two branch gambles into four and eight branch gambles as well as other tests of coalescing and idempotence. The same design was used in Sample 2, except $96, $44, and $4 were substituted for $98, $40, and $2. The order of choice problems was randomized with the restriction that successive choice problems not come from the same design. Complete materials can be viewed at URLs:

http://psych.fullerton.edu/mbirnbaum/decisions/allais5_.htm

http://psych.fullerton.edu/mbirnbaum/decisions/allais6_.htm

Table 7

Tests of double splitting (#7, 7a, 7b; 8, 8a, 8b), idempotence (#7a, 7c; 8a, 8c), and stochastic dominance (#10d)

Problem	Choice		%*R*	%*R*	True and Error Sample 1		True and Error Sample 2	
No.	***S***	***R***	***n* = 104**	***n* = 107**	***p***	***e***	***p***	***e***
7	$40 for sure	($98, 0.5; $2, 0.5)	36	37	0.34	0.06	0.35	0.05
7a	$40 for sure	($98, 0.25; $98, 0.25; $2, 0.25; $2, 0.25)	40	38	0.40	0.02	0.37	0.02
7b	$40 for sure	($98, 0.13; $98, 0.13; $98, 0.12; $98, 0.12; $2, 0.12; $2, 0.12; $2, 0.13; $2, 0.13)	40	38	0.39	0.05	0.38	0.02
7c	($40, 0.25; $40, 0.25; $40, 0.25; $40, 0.25)	($98, 0.25; $98, 0.25; $2, 0.25; $2, 0.25)	35	34	0.34	0.04	0.33	0.05
8	$48 for sure	($98, 0.5; $2, 0.5)	32	36	0.30	0.06	0.35	0.02
8a	$48 for sure	($98, 0.25; $98, 0.25; $2, 0.25; $2, 0.25)	40	37	0.40	0.03	0.37	0.02
		($98, 0.13; $98, 0.13; $98, 0.12;						

8b	$48 for sure	$98, 0.12; $2, 0.12; $2, 0.12; $2, 0.13; $2, 0.13)	39	38	0.39	0.02	0.38	0.02
8c	($48, 0.25; $48, 0.25; $48, 0.25; $48, 0.25)	($98, 0.25; $98, 0.25; $2, 0.25; $2, 0.25)	35	33	0.34	0.06	0.31	0.04
9	$50 for sure	($98, 0.7; $2, 0.3)	59	57	0.59	0.02	0.58	0.02
9a	$50 for sure	($98, 0.7; $2, 0.1; $2, 0.1; $2, 0.1)	54	57	0.55	0.07	0.58	0.05
9b	$50 for sure	($98, 0.5; $98, 0.1; $98, 0.1; $2, 0.3)	67	64	0.69	0.05	0.65	0.08
10a	($98, 0.85; $98, 0.05; $2, 0.1)	($98, 0.9; $2, 0.05; $2, 0.05)	44	42	0.41	0.14	0.40	0.10
10d	($98, 0.85; $96, 0.05; $2, 0.1)	($98, 0.9; $4, 0.05; $2, 0.05)	40	39	0.34	0.18	0.34	0.15

Participants were 211 undergraduates enrolled in lower division psychology courses at California State University, Fullerton (USA). There were 104 and 107 who served in Samples 1 and 2, respectively. The sample was 57% and 62% female, and 87% and 86% were 20 years of age or younger in the two groups, respectively. Each person completed the entire study twice, separated by other tasks that required about 20 min of intervening time. Participants were told that three lucky participants would be selected by chance to play one of their chosen lotteries for real cash prizes.

7. Results of experiment 2

The percentages representing preference for the "risky" gamble in the main design are shown under "%*R*" in Table 6. According to EU, people should make the same responses in all rows of Table 6, apart from error. Instead, the observed choice percentages vary from a low of 17% (Choice Problem 3 in Sample 2) to a high of 74% (Problem 4 in Sample 2). These are extremely large violations of EU in raw choice percentages.

Choice Problems 1 and 3 also produce large violations of EU (from 59% and 66% preferring the risky gamble in Problem 1 to only 25% and 17% in Problem 3). This Allais Paradox is reversed in Problems 1a and 3a, in split form. This reversal of the direction of the Allais paradox under splitting contradicts RDU, CPT and EU models, but it replicates previous results with the same choice problems (Birnbaum 2004). It also confirms results from Experiment 1 (CCE1 and CCE1s). The next section analyzes the question of whether errors can account for these violations.

7.1. True and error analysis

The true and error model was fit to the replicated presentations of each choice problem in each sample of Experiment 2, and was fit to the data combined across samples as well. Details are presented in the online supplement in Appendix B, which shows that the true and error model gave a good fit to the data for replications and that the rival model of response independence (implied by random preference models) could be rejected in every statistical test.

Because the true and error model gives a good representation of the data, we can use its parameter estimates of the true choice probabilities to address the main issue of this article: Can we attribute violations of EU to random error?

The estimated error rates and true choice probabilities for the true and error model are presented in Tables 6 and 7 for the main design and double-splitting designs, respectively. According to EU, a person should choose either the "risky" gamble or the "safe" gamble in every choice of Table 6, but should not switch, except by

random error. The choice proportions, corrected for error, should therefore be constant in all rows of Table 6. Instead, the estimated true probabilities choosing the risky gamble vary from as low as 0.09 and 0.13 for Problem 6 in Samples 1 and 2 respectively, to as high as 0.80 and 0.88 for Problem 4. Similarly, they vary from 0.1~~5~~6 and 0.~~10~~09 in Problem 3 to 0.64 and 0.72 in Problem 1. These represent huge violations of EU, corrected for error.

According to RDU, CPT, and EU, choice responses should not differ if choice problems are presented in coalesced or split form. In reality, there are large effects of splitting in Tables 6 and 7. Choice Problems 1 versus 1a, 3 versus 3a, and 4 versus 4a produced splitting effects large enough to reverse modal choices. In all three of these cases, the branch leading to the highest consequence of one gamble is split and the lowest branch of the other gamble was split. Most people who switched preferences did so as predicted by the TAX model; that is, choice proportions increased for gambles whose higher branch was split relative to gambles whose lower branch was split.

When gambles are presented in canonical split form, it should be easy to see and cancel identical branches, if a person wanted to follow this editing principle of original prospect theory. If a person used cancellation in Experiment 2, her data would satisfy restricted branch independence. Instead, there are significant violations of restricted branch independence; note that Problems 1a, 4a, 5a, and 6a all yield estimated probabilities less than 0.5, but Problems 2a and 3a yielded estimated probabilities that exceed 0.5 in Sample 1. As shown in Table 6, violations of restricted branch independence are not diminished by the correction for error.

The observed violations of restricted branch independence are opposite from Allais paradoxes in coalesced form, and they are opposite predictions of CPT based on the inverse-S probability weighting function. These findings corroborate previous results (Birnbaum 2004) and extend them by showing that these failures of CPT for the Allais paradoxes cannot be attributed to random error.

The prior TAX model correctly predicted the modal choices in nine cases of Table 6 but failed to predict the modal response in Choice Problems 2, 2a, 6, and 6a. In

all four of these cases, people tend to prefer the gamble that gives three positive outcomes over a gamble giving only one or two possible consequences. CPT with prior parameters of Tversky and Kahneman (1992) failed to predict these same four choices (2, 2a, 6, and 6a), but it also failed to predict four other cases that were correctly predicted by prior TAX: #1a, 3a, 4a, and 6c, all involving choice problems in split form.

Appendix B in the online supplement analyzes a true and error model in which different people are allowed to have different amounts of noise in their data. The analysis shows that even allowing individual differences in error rates, the same conclusions are reached regarding EU and CPT.

Appendix C in our online supplement considers a more complex true and error model in which the probability of making an error depends not only on the choice problem but also on a person's true preference state. Can we save EU by doubling the number of error rate parameters to be estimated from the data? For example, perhaps those who truly prefer the risky gamble in Choice Problem 1 are more likely to make an error in Problem 3 than those who truly prefer the safe gamble. This error model would imply a shift of choices from risky to safe in Problems 1 and 3; but can it save the EU model? As shown in Appendix C, even this more complex error model requires us to reject the EU model.

7.2. Double splitting and stochastic dominance

Table 7 shows that double splitting (splitting both the upper and lower branches of a gamble) did not affect overall choice proportions by very much. The largest effect occurred in Choice Problems 7 and 7a; in all other tests, data are consistent with the hypothesis that fewer than 5% of participants were affected by this manipulation. The prior TAX model with $k = 1$ in Eq. (4) implies that double splitting should have reduced the utility of the lotteries. Instead, it appears that TAX would be improved with $k = 0$, instead of $k = 1$. A caveat is that with so many choice problems using splitting manipulations in this study, some participants may have learned to recognize the equivalence of the splitting manipulation; therefore, a different magnitude effect might be observed in situations where fewer choices of this type would be presented. However, if these results generalize, it suggests

that the proportion of weight transferred in the TAX model might better be represented as δ/n rather than $\delta/(n + 1)$ in Eq. (5).

Choice Problem 10d (Table 7) is a test of first order stochastic dominance of a type that has produced high rates of observed violations in previous studies (e.g., Birnbaum and Navarrete 1998). In this new study, the observed percentages of violations are 60 and 61% in Samples 1 and 2, respectively. Problem 10a is a test of coalescing/splitting related to Problem 10d, which shows similar rates of violation. The prior TAX model correctly predicts these two modal choices, which violate both EU and CPT. Corrected for error, estimated true rates of violation are higher for both problems.

8. Conclusions

Three major conclusions can be drawn from our study: First, we 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 Experiment 1. In addition, we found significant violations of upper tail independence controlling for errors. Second, violations of coalescing are systematic. In nine of fourteen tests of Experiment 1 coalescing could be rejected, which refutes EU, RDU, and CPT. Third, when both lotteries in each choice are presented in canonical split form, only one test (a common ratio effect) remained significant in the same direction; all other tests were either insignificant, unsystematic, or significant in the opposite direction. In Experiment 2, we replicated the main findings of Experiment 1 and extended them, showing that the Allais paradoxes, the effect of splitting, and the violations of restricted branch independence cannot be attributed to error.

These results agree with Birnbaum's (2004, 2008) conclusions that the Allais paradox is largely due to violations of coalescing and that violations of restricted branch independence can work in the opposite direction from the Allais paradox. Schmidt and Seidl (2014) reached similar conclusions regarding the common ratio effect and coalescing. But neither of those studies controlled for unequal error

rates in different choice problems, so the present findings (corrected for error) strengthen the case beyond those from these earlier papers.

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 axiom in the common consequence effect, he reformed the Allais choices in (what we now call) canonical split form in order to convince himself to satisfy his own axiom. Violations of first-order stochastic dominance are nearly eliminated when choice problems are presented in canonical split form (see review in Birnbaum 2008). 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.

Presenting choices in canonical split form cannot save EU, however, as shown by the significant violations even when choices are presented in canonical split form (Tables 4 and 6). Birnbaum and Bahra (2012) found that when sufficient data are collected from each person to fit an individual true and error model to each person, violations of restricted branch independence remain significant. But would restricting the theoretical domain to choices in canonical split form save RDU and CPT models? These models can violate restricted branch independence. Experiment 1 shows 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 that violate RDU and CPT as do violations of stochastic dominance in coalesced form in Table 7 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, restricting the domain to canonical split form, CPT and RDU models could no longer be considered as theories of the Allais paradoxes and other results in the literature that have used coalesced lotteries. Further, to account for violations of RBI in canonical split form (e.g., Birnbaum 2008), these models would need to give up the inverse-S weighting function, since the observed violations of RBI are opposite in direction from those predicted by that any inverse-S weighting function.

Another potential problem with this approach (restricting attention to choices presented in canonical split form) arises, however, from the fact that how a lottery is split in canonical form depends on the other lottery with which it is compared. That means that lottery *A* appears in one form when paired with *B* and in another form when paired with *C*. 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 the utility of one gamble varies depending on the other lottery with which it is compared. 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 apparent violations of transitivity, which would then rule out these models even within 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 choices between lotteries (i.e., the experimental methods of representing and displaying probabilities and choices), the question naturally arises, is there a procedure or format in which violations of coalescing and other violations of CPT are minimized? So far, 15 different formats have been tested, all showing strong evidence against CPT (Birnbaum, 2008; see our online supplement for additional references).

Violations have been observed when lotteries are represented via 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 regular probabilities to win *x*, and with decumulative probabilities (probabilities to win prize *x* or more), with positive, negative, and mixed gambles, with dependent and independent gambles, and with other variations of format and procedure. 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, nor has a format been found in which systematic violations of restricted branch independence match the predictions of the inverse-S

decumulative weighting function.

In experiments with many trials, rates of violation of EU sometimes decrease with practice. We further analyzed the data of Experiment 1 and found that violations of coalescing diminished significantly with practice but persisted throughout the experiment (Appendix A; Birnbaum and Schmidt 2015). Whether or not people can learn to satisfy EU in a long enough experiment, perhaps with education and training, remains an open question (Humphrey 2006).

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. Models that 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 et al. 2008a, b), and configural weight theory (Birnbaum 2008).

Acknowledgements

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. Thanks are due to Jeffrey P. Bahra for assistance in data collection for Experiment 2.

9. Electronic supplementary material

ESM 1

(PDF 861 kb)

References

AQ7

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

Abdellaoui, M. (2009). Rank-dependent utility. In P. Anand, P. K. Pattanaik, & C. Puppe (Eds.), *The Handbook of rational and social choice* (pp. 90–112). Oxford: Oxford University Press.

Abdellaoui, M., Vossmann, F., & 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., & Sugden, R. (1997). Does part-whole bias exist? An experimental investigation. *Economic Journal, 107*, 322–332.

Berg, J. E., Dickhaut, J. W., & Rietz, T. (2010). Preference reversals: The impact of truth-revealing monetary incentives. *Games and Economic Behavior, 68*, 443–468.

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

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463–501.

Birnbaum, M. H. (2013). True-and-error models violate independence and yet they are testable. *Judgment and Decision making, 8*, 717–737.

Birnbaum, M. H., & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and*

Decision making, 7, 402–426.

Birnbaum, M. H., & 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., & Schmidt, U. (2015). The impact of learning by thought on violations of independence and coalescing. *Decision Analysis*. doi: 10.1287/deca.2015.0316 .

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

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.

Blavatskyy, P. (2011). A model of probabilistic choice satisfying first-order stochastic dominance. *Management Science, 57*, 542–548.

Blavatskyy, P. (2012). Probabilistic choice and stochastic dominance. *Economic Theory, 50*, 59–83.

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

Butler, D. J., & Loomes, G. C. (2007). Imprecision as an account of the

preference reversal phenomenon. *American Economic Review, 97*, 277–297.

Butler, D., & Loomes, G. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization, 80*, 511–522.

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

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

Conlisk, J. (1989). Three variants on the Allais example. *The American Economic Review, 79*, 392–407.

Conte, A., Hey, J. D., & Moffatt, P. G. (2011). Mixture models of choice under risk. *Journal of Econometrics, 162*, 79–88.

Cox, J. C., Sadiraj, V., & Schmidt, U. (2015). Paradoxes and mechanisms for choice under risk. *Experimental Economics, 18*, 215–250.

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

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

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

Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica, 74*, 121–146.

Harless, D., & Camerer, C. F. (1994). The predictive utility of generalized

expected utility theories. *Econometrica, 62*, 1251–1289.

Harrison, G. W., & Rutström, E. (2009). Expected utility and prospect theory: One wedding and a decent funeral. *Experimental Economics, 12*, 133–158.

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

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

Hey, J. D., Morone, A., & Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty, 39*, 213–235.

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.

Humphrey, S. J. (2006). Does learning diminish violations of independence, coalescing, and monotonicity? *Theory and Decision, 61*, 93–128.

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

Kilka, M., & 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*, 301–323.

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

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

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty, 4*, 29–59.

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

McDonald, J. (2009). *Handbook of Biological Statistics*. Baltimore: Sparky house publishing (2nd edition).

Özdemir, T., & Eyduran, E. (2005). Comparison of chi-Square and likelihood ratio chi-Square tests: Power of test. *Journal of Applied Sciences Research, 1*, 242–244.

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.

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

Schmidt, U. (2004). Alternatives to expected utility: Formal theories. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory*, *Extensions* (Vol. II, pp. 757–838). Dordrecht: Kluwer Academic Publishers.

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

Schmidt, U., & Neugebauer, T. (2007). Testing expected utility in the presence of errors. *Economic Journal, 117*, 470–485.

Schmidt, U., & Seidl, C. (2014). Reconsidering the common ratio effect: The roles of compound independence, reduction, and coalescing. Kiel working paper no. 1930.

Sonsino, D., Ben-Zion, U., & Mador, G. (2002). The complexity effects on choice with uncertainty – Experimental evidence. *The Economic Journal, 112*, 936–965.

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

Starmer, C., & Sugden, R. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypothesis. *Ann. Oper. Res., 19*, 79–102.

Starmer, C., & 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 S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory*, *Extensions* (Vol. II, pp. 685–755). Dordrecht: Kluwer Academic Publishers.

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

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

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty, 2*, 235–264.

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

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

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

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Research in Experimental Economics, 12*, 197–292.

Wilcox, N. T. (2011). ‘stochastically more risk averse’: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics, 162*, 89–104.

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

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

---

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