# 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

## Preliminary version: Do neither cite nor distribute

**Abstract**

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

*JEL classification:* C91, D81

**February 2009**

---

* We thank Graham Loomes, Peter Wakker, ??? and seminar participants in Exeter, Barcelona, 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

Nowadays there exists abundant evidence that expected utility (EU) theory fails to provide an accurate description of peoples´ choice behavior under risk. One main problem is that often choices typically 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 Sugden, 2004, Schmidt, 2004, or Abdellaoui, 2009 for recent surveys). Most of these theories rely on independence conditions which are weakened variants of the independence axiom of EU. Experimental investigations of these weakened independence conditions revealed violation rates which are similar to those reported for the independence axiom of EU (Wakker, Erev, and Weber, 1994; Wu, 1994; Birnbaum and Chavez, 1997; Birnbaum, 2005, 2008).

Many studies have additionally shown that people typically make errors when choosing between risky lotteries (see e.g. Camerer, 1989; Starmer and Sugden, 1989; Harless and Camerer, 1994; Hey and Orme, 1994) which means that in repeated choice problems they choose one option in the first repetition but the other option in the second one. Such errors imply that choices involve a stochastic component. To take into account a stochastic component is also necessary for econometric evaluations of the performance of the alternatives to EU. Nowadays one of the most intensively discussed questions in decision theory is how to model this stochastic component (recent papers among many others are e.g. Gul & Pesendorfer, 2006; Blavatskyy, 2007, 2008; Conte, Hey, and Moffat, 2007; Hey, Morone, and Schmidt, 2007; Wilcox, 2008a, b; Harrisson and Rutström, 2008; etc.)

An interesting question in this context is whether the empirical performance of EU improves if we model the stochastic component properly. For instance already in 1995 John Hey 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" (Hey, 1995, p. ???; see also Buschena and Zilberman, 2000). If this is really the case then some of the reported violations of EU should be at least partly caused by errors instead of being intrinsic violations. There exist several recent studies showing 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) for violations of transitivity, and Schmidt and Hey (2004), Butler and Loomes (2007) for preference reversals. In the present paper we will analyze whether reported violations of the independence axiom of EU and violations of weaker independence conditions may be caused by errors. In the next section we show that even EU with a Fechner error term may generate the systematic pattern of violations of the independence axiom observed in experimental research. A previous study by Schmidt and Neugebauer (2007) provides some evidence in favor of this model. The study proposes that most errors can be excluded if only choices are considered where subjects chose the same option three times in row, since it is rather improbable that a subject makes the same error three times in row. 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 of this issue. We also perform a repeated choice experiment and fit an error model which is neutral with respect to violations of any independence condition. This model allows us to discriminate precisely which part of violations can be attributed to errors and which part should be considered as "real" violation. Note that such an analysis is not possible with EU plus a Fechner error term since this model presupposes that true preferences can be represented by EU and, thus, satisfy the independence axiom.

A further systematic deviation from EU and in fact also from most of the alternatives to EU is given by splitting effects resulting from violations of coalescing. A splitting effect occurs if splitting an event with a given consequence into two separate events systematically influences choice behavior. There exists robust evidence that splitting an event with a good (bad) consequence increases (decreases) the attractiveness of a lottery in comparison to other lotteries with a good (bad) consequence increases (decreases) the attractiveness of a lottery in comparison to other lotteries (Starmer and Sugden 1993; Birnbaum and Navarette, 1998 Humphrey 1995, 2001).[1] While Birnbaum and Navarette employed splitting effects in order to generate substantial violations of first-order stochastic dominance, the papers of Humphrey show that splitting effects have contributed to previously observed violations of transitivity. It may well be the case that splitting effects may also contribute to violations of independence conditions. We will analyze this question while controlling for errors at the same time.

[1] For similar evidence of splitting effects in other contexts than choice under uncertainty see Weber, Eisenführ and von Winterfeldt (1998), Bateman et al. (1997), and Neugebauer, Schmidt, and Starmer (2008).

This paper is organized as follows. The next section presents our theoretical model and discusses in general the issue of testing independence in the presence of errors. Section 3 is devoted to our experimental while section 4 reports the results. Some concluding observations are discussed in section 5.

## 2 Errors and Violations of Independence

In this section we will discuss the possible role of errors for generating systematic violations of independence and introduce our error model. Consider a simple variant of the common ratio effect taken from Birnbaum (2001).

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 for a utility function $u$ with the normalization $u(0) = 0, u(23) > (<) 0.5u(46)$ implies $0.02u(23) > (<) 0.01u(46)$. 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. This particular result is also called "certainty effect" in reference to the fact that people more often choose the "sure thing" in the second choice. Can this result have occurred by random errors? Note that in principle systematic deviations from independence can also be explained by EU plus a Fechner error term. In this case a subject chooses $R$ over $S$ if $EU(R) - EU(S) + \varepsilon > 0$ where $\varepsilon$ 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 may much more easily influence the choice between $R$ and $S$ than the choice between $R'$ and $S'$ implying that we observe much more 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 an alternative error model which we explain next.

Suppose we assume 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 well 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, is it possible that $p_{RS'} = p_{SR'} = 0$ given the data in Table 1? The answer is "yes," even though the observed response rates are significantly different.

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. This probability is the sum of four terms, each representing the probability of having one of the "true" patterns and having the appropriate pattern of errors and correct responses to produce each observed data pattern. For example, the person who truly has the *RR′* pattern could produce the *RS′* pattern by correcting 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.

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. There are four "true" probabilities, which sum to 1, and two "error" probabilities. Thus, we have three degrees of freedom in the data and five parameters to estimate. That means that we have many solutions possible. Two solutions that fit the data perfectly are shown in the table below:

| **Parameter** | Model 1: EU fits | 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

Thus, we can "save" EU in this case by allowing that people might have errors in their responses. So given the error model and our data in Table 1 it is not possible to conclude that true preferences cannot be represented by EU. In order to reach a firmer conclusion we need a way to estimate parameters that do not assume 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. We will do this 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 *RR* pattern is given as follows:

(2) $P(RR) = p(1 - e)(1 - e) + (1 - p)e^2,$

where $p$ is the true probability of preferring $R$ and $e$ is the error rate on this choice.

By adding replications to a 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 parameters need be estimated). The general model (which allows all four probabilities to be non-zero) is now over-determined, with 10 degrees of freedom. The EU theory is a special case of this model in which two of the true probabilities are fixed to zero; therefore, the difference in chi-squares provides a chi-square test with two degrees of freedom. In sum, without replications, two theories are perfectly compatible with these data, one of which assumed EU is true. However, with replications we can estimate the error terms and determine the accuracy of EU model.

This study will include experiments in which there are up to four replications. With two choices and four replications, there are 256 possible response patterns ($4^4 = 256$). Because many of the patterns will be observed with zero frequency, we use the G-statistic 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. 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 deviations from EU are significant or whether they are just caused by errors in the response of subjects.

## 3 Experimental Design

The experiment was conducted at the University of Kiel with 54 subjects, 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 exceeds the usual wage of students (about 8 Euro per hour) considerably.

Lotteries were presented as in Figure 2 and subjects had to circle their choice. Prizes were always ordered form 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$ when drawing a ticket from 1 to 50, 30$ for a ticket between 51 and 80, and 40$ for a ticket between 81 and 100. All this was explained in the instructions which were give to the students in printed from and read out aloud. At the end of instructions, subjects had to answer four transparent dominance questions which were controlled by the experimenter before proceeding.

| | | | | | |
|---|---|---|---|---|---|
| A: | 50% to win | 20 $ | or B: | 33% to win | 10 $ |
| | 30% to win | 30 $ | | 34% to win | 15 $ |
| | 20% to win | 40 $ | | 33% to win | 18 $. |

**Figure 2:** Presentation of lotteries

Lotteries in the booklets were presented in a pseudo-random order. The ordering of lotteries was different in each booklet and no choice problem was followed by another testing the same independence property. Only after finishing one booklet a subject received the next one. The ordering of handing out booklets was randomly determined for each subject. 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 stimuli involved 11 tests of independence conditions, nine of which being investigated in both, coalesced and split variants. All these 20 tests were replicated four times with counterbalanced left-right positioning. Additionally, in order to test the attentiveness of subjects, each booklet involved two transparent stochastic dominance questions, one based on outcome monotonicity and one on event monotonicity.

Our tests of independence conditions and the involved lottery pairs 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 which 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 variants of the lottery pairs. For the tests of independence conditions in split variants we used the canonical split form of these pairs. In the canonical split form of a pairwise choice, both lotteries are split so that there are equal probabilities on corresponding ranked branches and the number of branches is equal in both gambles and minimal. A presentation of the lottery pairs employed in the split tests can be found in the appendix. Note that each pairwise choice problem presented in Table 3 has a unique canonical split form.

The first six tests in Table 3 are four common consequence effects (CCE1-4) and two common ratio effects (CRE1 and 2). CCEs and CREs are the most common design for testing the independence axiom of EU, also the paradoxes of Allais are special variants of a CCE and a CRE. CCEs can be formally described by $S = (x, p_1; s_2, p_2; s_3, p_3)$, $R = (x, q_1; r_2, q_2; r_3, q_3)$, $S' = (x, p_1 - \alpha; s_2, p_2; s_3, p_3; x', \alpha)$, and $R' = (x, q_1 - \alpha; r_2, q_2; r_3, q_3; x', \alpha)$, i.e. $S'$ and $R'$ are constructed from $S$ and $R$ by shifting probability mass $(\alpha)$ from the common consequence $x$ to a different common consequence $x'$. Consequently, an EU maximizer will prefer $S$ over $R$ if and only if she will prefer $S'$ over $R'$. Note that in Table 3 the first lottery of a choice problem always characterizes the lotteries $S$ and $R$ and the second one the lotteries $S'$ and $R'$. For CCE1 we have for instance $x = 0, p_1 = 0.8, p_2 = 0.2, s_2 = 19, p_3 = 0$ for $S$, $q_1 = 0.90, q_2 = 0.10, r_2 = 44, q_3 = 0$ for $R$ and $S'$ and $R'$ are constructed by setting $\alpha = 0.4$ and $x' = 44$. The lotteries in the four CCEs of our experiment are taken from Starmer (1992) who observed high violation rates for these lotteries. The typical pattern of violations in CCE1-4 is that people prefer $R$ over $S$ but $S'$ over $R'$. The same is true for the two CREs (CRE1 and 2) presented in Table 3. A CRE can be formally described by $S = (x, 1 - \beta(1 - p_1); s_2, \beta p_2)$, $R = (x, 1 - \beta(1 - q_1); r_2, \beta q_2)$, $S' = (x, p_1; s_2, p_2)$, and $R' = (x, q_1; r_2, q_2)$, i.e. $S$ and $R$ are constructed from $S'$ and $R'$ by multiplying all probabilities by $\beta$ and assigning the remaining probability $1 - \beta$ to the common consequence $x$. EU implies again that people choose either the risky or the safe lottery in both choice problems. In CRE1 (taken from Birnbaum, 2001) and CRE2 (taken from Starmer and Sugden, 1989), however, substantial violations of EU have been observed with many people choosing $R$ and $S'$.

| **Problem** | **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 | |
| CI | 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 were employed to derive alternative theories. We focus on axioms which are implied by rank-dependent utility (Quiggin, 1981, 1982; Luce 1991, 2000; Luce and Fishburn 1991), cumulative prospect theory (Starmer and Sugden, 1989; Tversky and Kahneman, 1992; Wakker and Tversky, 1993), and configural weight models (Birnbaum and McIntosh, 1996). A central axiom in this context is tail independence (TI) which was first proposed by Green and Jullien (1988) using the term ordinal independence. Formally, TI demands that $S = (x_1, p_1;\ \ldots;\ x_i, p_i;\ x_{i+1}, p_{i+1};\ \ldots;\ x_n, p_n) \quad R = (x_1, p_1;\ \ldots;\ x_i, p_i;\ x_{i+1}, q_{i+1};\ \ldots;\ x_n, q_n)$ if and only if $S' = (x_1, q_1;\ \ldots;\ x_i, q_i;\ x_{i+1}, p_{i+1};\ \ldots;\ x_n, p_n) \quad R' = (x_1, q_1;\ \ldots;\ x_i, q_i;\ x_{i+1}, q_{i+1};\ \ldots;\ x_n, q_n)$ where $x_1 \geq x_2 \geq \ldots \geq x_n$. TI demands that 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 chance if this tail is replaced by a different common tail. Note that in the definition above the upper tail is the common tail and thus the condition is called upper tail independence (UTI). 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 a very general property which is implied by many models including all variants of rank-dependent utility (RDU) as well as cumulative prospect theory (CPT). Therefore, rejecting TI would provide serious evidence against all these models. In his experiments, Wu (1994) observed violation rates of UTI of up to 50%. Similar evidence has been reported by Birnbaum (2001) and Wakker, Erev, and Weber (1994) where the latter paper tests comonotonic independence, the analogue to TI in choice under uncertainty. Our study tries to find out whether the reported violations of TI may be due to splitting effects and/or errors. 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 cumulative independence (CI), which demands that decision weights depend only on cumulative probabilities. Formally, CI demands that S = $(s_1, p_1;\ s_2, p_2;\ \alpha, p_3) \quad R = (r_1, p_1;\ \gamma, p_2;\ \alpha, p_3)$ if and only if $S' = (s_1, p_1 + p_2; \gamma, p_3)$ $R' = (r_1, p_1; \gamma, p_2 + p_3)$, where $\alpha > \gamma > s_2 > s_1 > r_1$. Substantial violations of CI have been reported by Birnbaum and Navarette (1998) and Birnbaum, Patton, and Lott (1999). Our lottery pairs are taken from the latter paper which observed violation rates of 40.1% for these pairs, where most of the violating subjects preferred $R$ and $S'$.

The final property we test is distribution independence (DI). Whereas configural weight models and original prospect theory imply that DI holds, it should be violated according to RDU and CPT, at least if the weighting function is inverse-S shaped as commonly suggested 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 = $(s_1, \beta; s_2, \beta; \alpha, 1 - 2\beta)$ $R = (r_1, \beta; r_2, \beta; \alpha, 1 - 2\beta)$ if and only if $S' = (s_1, \delta; s_2, \delta; \alpha, 1 - 2\delta)$ $R' = (r_1, \delta; ; r_2, \delta; \alpha, 1 - 2\delta)$ where $\alpha$ is either the highest or the lowest outcome in both lotteries. If $\alpha$ is the highest outcome, the condition is called upper distribution independence (UDI), otherwise lower distribution independence (LDI). The lotteries used in our tests of UDI and LDI are taken from Birnbaum (2005). 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 will first comment on our results with respect to first-order stochastic dominance. In each booklet there were two transparent dominance questions. With four booklets we have altogether eight tests of dominance. Out of our *54* subjects five violated dominance once and one subject twice. We conclude from this result that our subjects were sufficiently attentive and motivated. This conclusion is supported by the fact that all our estimated error rates (see Table 4) are between *2%* and *22%* (mean *11.4%*) which is quite in line with estimations in related studies.

Table 4 gives an overview over our results on our tests of the single independence conditions. For all conditions listed in the first column we report the estimated probabilities (or relative frequencies) of the four possible response patterns to be reflecting true preferences in columns two to five. An index *S* in the first column indicates that this test of the respective independence condition was performed by presenting all lotteries in their canonical split form. In columns six and seven we report the error rates in the choices between *S* and *R (e)* as well as between *S′ and R′ (e′)*. The final column presents the statistics of a chi-square test between the fit of a general model – i.e. 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. a model corresponding to the special case in which $p_{RS'} = p_{SR'} = 0$ must hold. 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 a significance level of *5% (1%)*.

| Property | $p_{SS}$ | $p_{SR}$ | $p_{RS}$ | $p_{RR}$ | $e_1$ | $e_2$ | Test |
|---|---|---|---|---|---|---|---|
| CCE1 | 0.44 | 0.02 | 0.30 | 0.24 | 0.15 | 0.11 | **20.36**** |
| CCE1$_S$ | 0.52 | 0.20 | 0.00 | 0.28 | 0.13 | 0.16 | **12.77**** |
| CCE2 | 0.02 | 0.00 | 0.10 | 0.88 | 0.02 | 0.08 | **15.33**** |
| CCE2$_S$ | 0.09 | 0.03 | 0.05 | 0.84 | 0.07 | 0.07 | **7.61*** |
| CCE3 | 0.25 | 0.21 | 0.16 | 0.39 | 0.16 | 0.12 | **18.69**** |
| CCE3$_S$ | 0.52 | 0.24 | 0.00 | 0.25 | 0.13 | 0.16 | **12.63**** |
| CCE4 | 0.67 | 0.01 | 0.29 | 0.02 | 0.14 | 0.09 | **21.96**** |
| CCE4$_S$ | 0.80 | 0.01 | 0.02 | 0.17 | 0.14 | 0.12 | 0.82 |
| CRE1 | 0.25 | 0.00 | 0.64 | 0.11 | 0.11 | 0.07 | **44.64**** |
| CRE1$_S$ | 0.44 | 0.00 | 0.46 | 0.10 | 0.15 | 0.05 | **27.21**** |
| CRE2 | 0.57 | 0.00 | 0.20 | 0.23 | 0.14 | 0.11 | **18.00**** |
| CRE2$_S$ | 0.84 | 0.02 | 0.01 | 0.12 | 0.17 | 0.12 | 0.45 |
| UTI | 0.06 | 0.01 | 0.52 | 0.40 | 0.13 | 0.18 | **18.76**** |
| UTI$_S$ | 0.17 | 0.01 | 0.00 | 0.82 | 0.14 | 0.18 | 0.05 |
| LTI | 0.04 | 0.00 | 0.14 | 0.82 | 0.05 | 0.15 | 3.96 |
| LTI$_S$ | 0.04 | 0.00 | 0.01 | 0.95 | 0.06 | 0.08 | 0.24 |
| CI | 0.14 | 0.08 | 0.13 | 0.66 | 0.13 | 0.22 | 3.88 |
| CI$_S$ | 0.13 | 0.09 | 0.02 | 0.76 | 0.13 | 0.09 | 5.07 |
| LDI | 0.94 | 0.00 | 0.00 | 0.06 | 0.02 | 0.05 | 0.00 |
| UDI | 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

We will first analyze the tests of the independence axiom of EU. This axiom demands $p_{RS'} = p_{SR'} = 0$ whereas empirical research on CCEs and CREs has reported systematical violations, most of them in the direction of the *RS′* pattern. Using as in previous research only coalesced lotteries, we can confirm this result in our analysis: In all our six tests (CCE1-4, CRE1-2), the independence axiom is rejected in favor of the more general model. 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 probabilities never exceed *3%*). The only exception is CCE3 where the estimated probability of pattern *SR′* slightly exceeds that of *RS′*. Consequently, we can conclude as our first result that the typical evidence on violations of the independence axiom of EU can be also observed in a framework controlling for errors which means that these violations can be regarded as true violations.

A quite different picture arises when the same test are performed with lotteries presented in split form. From our six tests, two ($CCE4_S$ and $CRE2_S$) are insignificant (i.e. EU cannot be rejected) and two ($CCE1_S$ and $CCE3_S$) are significant but precisely in the opposite direction as observed in previous tests and in our analysis relying on coalesced lotteries. From the two remaining tests, only one (CRE1) clearly supports 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, both for CCEs and CREs. Presenting lotteries in their canonical split forms does not at all generate the clear pattern of violations reported in previous report and also found in our analysis relying on coalesced lotteries. So the question arises whether the previous evidence should be indeed regarded as evidence against the independence axiom or whether an interpretation in terms of violations of coalescing seems to be more appropriate.

We can now comment on tests of the weaker independence conditions. For UTI we can observe a substantial and systematic violation as 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) and Wakker, Erev, and Weber (1994). 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 must hold. It is, however, astonishing that this clear evidence of violations of TI entirely disappears if we present lotteries 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. It seems that CPT and other rank-dependent models may be only a descriptively valid representation of preferences if lotteries are always presented in split forms.

Our new version of TI, LTI, does not generate significant violations. The same is true for CI where also splitting does not have a visible impact. Comparing this result to the high violation rates of CI observed in previous papers (Birnbaum and Navarette, 1998; Birnbaum, Patton, and Lott 1999) with lotteries identical to those in our coalesced test, may lead to the conclusion that errors may have contributed to the previous results. This interpretation is supported by the fact that the estimated error rates in our coalesced test of CI are the highest of all our tests.

Our tests of DI confirm previous results by Birnbaum and Chavez (1997) and Birnbaum (2005). Significant violations of DI are also not observed in our tests which provide support for configural weight models but challenges 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:** Splitting effects

Since our results show a substantial influence of splitting effects on testing independence properties we also provide a direct analysis of splitting effects in Table 5. This table compares choice in a given lottery pair in coalesced form with choices in the same pair in split form. If no splitting effects occur, choice should be identical in both problems. This means that a subject either chooses the risky lottery in both problems or the safe lottery in both problems. In contrast, Table 5 shows that many people choose differently in the coalesced and split problems even if we control for errors ($e$ ($e'$) is the estimated in error rate of the choice problem stated first (second) in the first column of the table). The last column shows again the statistics of a chi-square test comparing the fit of a model which satisfies coalescing and thus implies $p_{RS'} = p_{SR'} = 0$, and the fit of a general model that allows for splitting effects and thus for non-zero probabilities of all four possible response patterns. It turns out that in nine out of 14 analyses the null $p_{RS'} = p_{SR'} = 0$ of has to be rejected in favor of the general model allowing for splitting effects.

## Appendix

| Problem | No. | $p_1$<br>$s_1$ | $p_2$<br>$s_2$ | $p_3$<br>$s_3$ | $p_4$<br>$s_4$ | $q_1$<br>$r_1$ | $q_2$<br>$r_2$ | $q_3$<br>$r_3$ | $q_4$<br>$r_4$ |
|---|---|---|---|---|---|---|---|---|---|
| $CCE1_S$ | 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 | | |
| $CRE2_S$ | 22 | *0,80* | *0,06* | *0,14* | | *0,80* | *0,06* | *0,14* | |
| | | 0 | 28 | 28 | | 0 | 0 | 45 | |
| | 21 | *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 |
| $CI_S$ | 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